Vandana Publications

Volume-4, Issue-4, August-2014, ISSN No.: 2250-0758

International Journal of Engineering and Management Research

Available at: www.ijemr.net

Page Number: 105-109

# Priority Based Dynamic Resource Allocation for Green Cloud Environment: A Survey

Rituraj Dixit[1], Prashant Buyan[2], Vijay Kumar[3]

[1,2]Research Scholar, CSE, HRIT Ghaziabad/UPTU, Lucknow, Uttar Pradesh, INDIA

[3]Computer Science, SSVIT, Bareilly, Uttar Pradesh, INDIA

**ABSTRACT**

**Cloud computing allows his subscribers to balance their resource usage based on requirements. Many of the touted gains in the cloud model come from resource multiplexing throughout virtualization technique. We offer a scheme that manipulate virtualization technique to assign physical machine resources vigorously based on application demands and maintain green computing by optimizing the number of physical machine in use. We initiate the notion of "skewness" to compute the unevenness in the multi-dimensional resource utilization of a physical machine. By reducing skewness, we can mingle dissimilar types of workloads smartly and get better the overall utilization of physical machine resources. We develop a set of heuristics that prevent overload in the system effectively while saving energy used. Trace driven simulation and experiment results demonstrate that our algorithm achieves good performance.**

***Keywords:*** VMM(Virtual Machine Monitor), APM(Actively used Physical Machine), LNM(Local Node Manager), IDV(Intelligent Desktop Virtualization), SLA(Service Level Agreement), EC(Elastic Computing)

## I. INTRODUCTION

Generally, Cloud computing can be illustrated as an equivalent for distributed computing over a network, with the capability to run a program or application on many distributed computers at the same time. It specially refers to a computing machinery or group of computing machineries commonly referred as a server connected through a communication network such as the Internet. Network-based services, which show to be provided by real server hardware and are in fact served up by virtual machinery simulated by software running on one or more real machines, is often called cloud computing. Such virtual servers do not physically survive and can hence be moved around and scaled up or down on the fly without affecting the end user.

One of the elementary aspects of virtualization technologies engaged in cloud environments is resource consolidation and managing. Using hypervisors inside a bunch environment allows for a number of standalone physical machines to be consolidated to a virtualized situation, thereby need less Physical Machine (PM) physical resources than ever before. The perception of Cloud computing has not only adjusted the field of distributed systems but also fundamentally changed how businesses consume computing today. While Cloud computing offers many advanced features, it still has some shortcomings such as the moderately high working cost for both public and private Clouds. Better resource allocation is depends on maneuver over utilization of physical machines and virtual machines.

Large Cloud deployments entail vast number of physical machines and megawatts of power. Therefore, there is a need to create professional Cloud computing system that utilizes the strengths of the Cloud while minimizing its power footstep. Virtual machine monitors (VMMs) like Xen provide a mechanism for grouping virtual machines (VMs) to physical resources [1].

Green computing is distinct as the number of servers used should be diminished as long as they can still convince the needs of all VMs. Idle PMs can be turned off to save energy. Simply make better performance over cloud surroundings by optimizing the number of servers in use. The area of Green computing is also becoming increasingly essential in a world with partial energy

resources and an ever-rising request for more computational power.

Users with the Amazon EC2 service [2] for instance, without knowledge of where their VM instances run. It is up to the cloud provider to make certain the underlying physical machines (PMs) have acceptable resources to meet their needs. VM live immigration technology makes it probably to change the mapping between VMs and PMs while applications are running. A strategy issue remains as how to decide the mapping adaptively so that the resource demands of VMs are met while the number of PMs used is minimized. This is challenging when the resource needs of VMs are heterogeneous due to the varied set of applications they run and vary with time as the workloads grow and shrink [3].

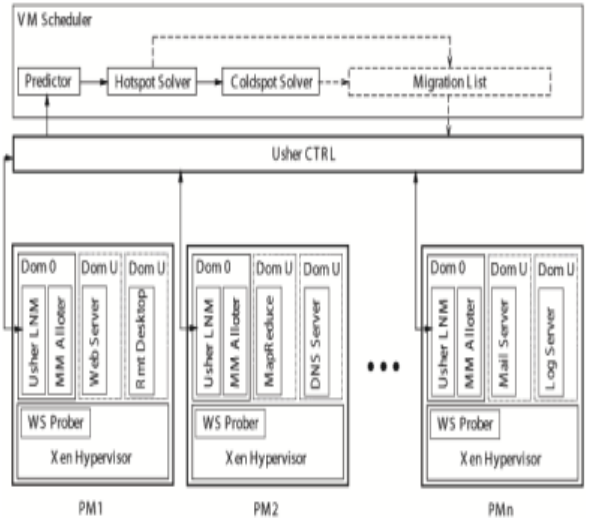


Figure 1: System Architecture

Successful separating of a machine to support the simultaneous execution of multiple operating systems causes several challenges.

- Virtual machines must be inaccessible from one another: it is not satisfactory for the execution of one to adversely distress the performance of another.
- It is obligatory to sustain a variety of different operating systems to accommodate the heterogeneity of admired applications.
- The performance transparency launched by virtualization should be diminutive.

The spine of any data centre is its data network. Without this, no machine is able to communicate with any other machine, or the outside world. As data centers are compactly crammed with servers. Para-virtualization allows many OS occurrences to run simultaneously on a single physical machine with soaring performance, providing enhanced use of physical resources and isolating individual OS instances.

## II. PRIOR APPROACH

Christopher Clark [4] proposed the intend options for drifting operating systems running services with live constraints, focusing on data center and bunch or cluster environment. They have also initiated and analyzed the notion of writable working set, and presented the blueprint, implementation and assessment of soaring performance OS migration built on crest of the Xen VMM. They have investigated an additional promote allowed by virtualization: that of live OS migration. Immigrating an entire OS and all of its applications as one unit permitted us to avoid many of the difficulties faced by process-level migration approaches.

In particular the narrow interface between a virtualized OS and the virtual machine monitor(VMM) makes it simple to avoid the problem of residual dependencies[5] in which the original host machine must linger available and network-accessible in order to over haul certain system calls or even memory accesses on behalf of migrated processes. Michael Nelson [6] proposed a devise and implementation of a system that utilizes virtual machine technology to provide fast, transparent application migration.

Carl A. Waldspurgeret [7] proposed a number of novel ESX Server mechanisms and policies for administrating memory. One of their proposed methods was ballooning approach, which has recovered the pages considered smallest amount valuable by the operating system functioning in a virtual machine. Secondly, an idle memory tax approach has attained proficient memory utilization while maintaining performance separation guarantees. Finally, Content-based page sharing and hot I/O page remapping approaches have exploited the transparent page remapping to eliminate redundancy and reduced copying overheads. All the four approaches were combined to efficiently maintains virtual machine workloads that over commit memory.

Norman Bobroffet introduced a forceful server migration, consolidation algorithm and proposed Dynamic Allotment of Virtual Machines for Managing SLA Violations [8].The above designated algorithm has been performed significant enhancement over fixed server consolidation in reducing the amount of obligatory capacity and the rate of service level agreement violations. The greedy resource allocation algorithm has been modified in [9] by Jeffrey S. Chaseet.al. Their proposed system is based on a monetary approach to managing shared server resources, in which services “bid” for resources as a function of distributed performance. The system often monitors load and plans resource allotments by estimating the value of their effects on service performance.

Tathagata Daset.al proposed a system [10] to save desktop energy by virtualizing the user’s desktop

computing situation as a virtual machine (VM) and then migrated between the user's physical desktop machine and a VM server, depending on whether the desktop computing situation was actively used or idle. Thus, the user's desktop environment was "always on", maintaining its network present fully even when the user's physical desktop device was switched off and thereby saved energy.YuvrajAgarwalet.al explained the structural design and implementation of Sleep server [11], a system that enabled hosts to changeover to such low-power sleep states while still maintaining their application's supposed network presence using an on stipulate proxy server.

NiltonBilaet.al proposed a fractional VM Migration technique [12] that transparently migrated only the working set of an idle VM. Partial VM migration arche type can deliver 85% to 104% of the energy savings of full VM resettlement, while using less than 10% as much network resources, and providing migration latencies that are two to three orders of degree smaller. Super computers are adequately powerful to use virtualization to presence the delusion of many smaller virtual machines (VMs), each running a separate operating system example. In [3], the authors presented a VMM called Xen hypervisor (116 VMM), which has allowed manifold commodity operating systems to share traditional hardware in a secure and resource managed manner, but without suffering either performance or functionality. This was done by providing an idealized virtual machine thought to which operating systems such as Linux, BSD (Berkeley Socket Distribution) and Windows XP, can be ported with minimal effort. The structural design of the system is presented in Figure 1. Each PM runs the Xen hypervisor (VMM) which wires a privileged domain 0 and one or more domain U. Virtualization of computing resources has been an increasingly regular practice in recent years, specifically in data centre environments. This has helped in the rise of cloud computing, where data centre machinists can over-subscribe their physical servers through the use of virtual machines in order to enlarge the return on investment for their infrastructure.

Overload evasion and Green computing parameters are playing chief of better resource allocation over green cloud background. There is an intrinsic trade-off between the two goals in the face of shifting resource needs of VMs. For overload evasion, we should keep the utilization of PMs low to reduce the possibility of overload in case the resource needs of VMs increase later. For green computing, we should keep the utilization of PMs reasonably high to make efficient use of their energy [13].

***Disadvantages of Existing System***

- Existing Priority based resource allocation failed to concentrate on both phase overload avoidance and green computing.
- There is an inherent trade-off between the two goals in the face of changing resource needs of VMs. For overload avoidance, we should keep the utilization of PMs low to reduce the possibility of overload in case the resource needs of VMs increase later.
- Existing priority based algorithm failed to helps cloud admin.
- Existing economy based service depends on many parameters like cost of resource, time needed to access, task type, number of processors needed to run the job or task etc. these external parameters not sufficient to achieve better RA (Resource Allocation**).**

## III. OUR APPROACH

Cloud computing is a model which enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.

A cloud environment consists of multiple customers requesting for resources in a dynamic environment with possible constraints. In the existing economy based models of cloud computing, allocating the resource efficiently is a challenging job. So new approach that should be allocates resource with minimum wastage and provides maximum profit. Existing Priority based resource allocation failed to concentrate on both phase overload avoidance and green computing.

We made minor changes over resource allocation phase in cloud medium. Our supplement stuff always focuses resource allocation under cost based aspect alone. Resource allocation is used to assign the available resources based on premium membership. Even though keep an eye over "dynamic resource allocation and hot spot mitigation". It's a new kind of Novelty approach for Green – Cloud Computing Environment. Cost based resource allocation doesn't require the detailed level of analysis and value measurement necessary to employ a value-based resource allocation strategy. Our novelty approach of the cost based resource allocation algorithms over cloud computing never affect green environment.

***Advantages of Proposed System***

- Our proposed technique flexible with dynamic cloud environment, like multiple customers requesting for resources in a dynamic environment with possible constraints.
- Priority algorithm is used for a better resource allocation of jobs in the cloud environment used for the simulation of different models or jobs in an efficient way.
- Develop a set of heuristics that prevent overload in the system effectively while saving energy used.
- In the existing economy based models of cloud computing, allocating the resource efficiently is a challenging job.
- Our new approach that allocates resource with minimum wastage and provides maximum profit.
- The developed resource allocation algorithm is based on different parameters like time, cost, No of processor request etc.

- In cloud network multiple users are requesting same resource, here hot spot mitigation improve the machine level efficiency.

***List of Modules in Proposed System***

Module 1: Measure 'SKEWNESS'
Module 2: Detect Hot and cold spots
Module 3: Hot spot improvement
Module 4: Green computing
Module 5: Optimized Scheduling algorithm

***Algorithm to Measure Skewness***

- We introduce the concept of skewness to quantify the unevenness in the utilization of multiple resources on a server.
- By minimizing the skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources.
- n - Number of Resources
  $r_i$ - Utilization of $i^{th}$ resource
  $\bar{r}$ -Average Utilization of all resources for server p.

$$\text{Skewness (p)} = \sqrt{\sum_{i=1}^{n}\left(1-\frac{r_i}{\bar{r}}\right)^2}$$

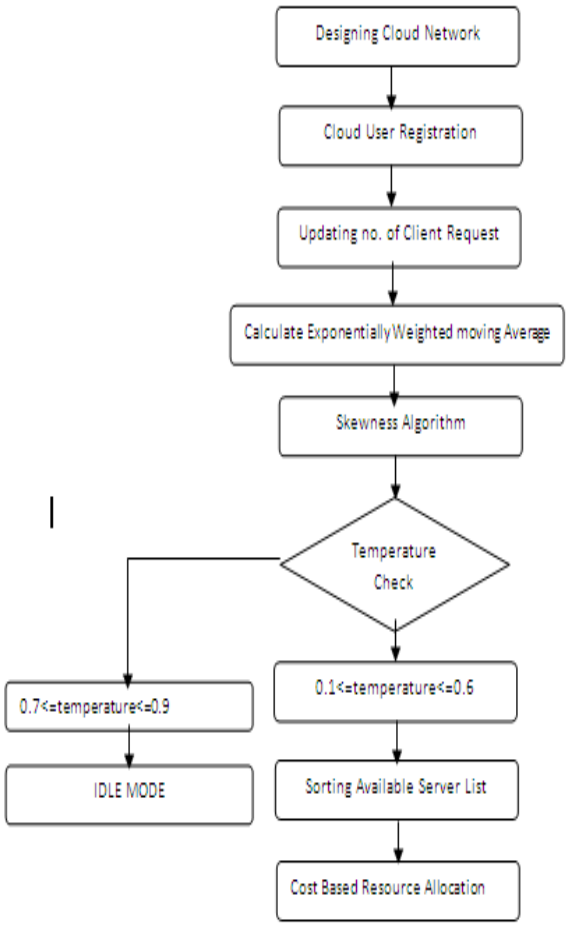


Figure 2: Flow Chart of Proposed System

## IV. CONCLUSION

Priority Based Dynamic Resource Allocation in Green Cloud Computing is a novel approach which allocate resources to the jobs depending upon their priority which depend upon premium membership i.e. cost based membership and priority also depend upon cost. Green Computing is also being achieved in this work as the temperature of the machines goes above the limit that machine goes in Hot Spot Zone and that machine will not take part in further cycle. In all this novel approach take care of skewness, resource utilization and green computing.

## 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," in Proc. of the ACM Symposium on Operating Systems Principles (SOSP'03), Oct. 2003

[2]"Amazon elastic compute cloud (Amazon EC2), http://aws.amazon.com/ec2/."

[3] 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," in Proc. of the ACM Symposium on Operating Systems Principles (SOSP'03), Oct. 2003

[4]C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach,I. Pratt, and A. Warfield, "Live migration of virtual machines," in Proc. of the Symposium on Networked Systems Design and Implementation(NSDI'05), May 2005.

[5] D. Milojicic, F .Douglis, Y. Paindaveine, R. Wheeler, and S. Zhou. Process migration. ACM Computing Surveys, 32(3):241-299, 2000.

[6] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in Proc. of the USENIX Annual Technical Conference, 2005.

[7] C. A. Waldspurger, "Memory resource management in VMware ESX server," in Proc. of the symposium on Operating systems design and implementation (OSDI'02), Aug. 2002

[8] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic placement of virtual machines for managing sla violations," in Proc. of the IFIP/IEEE International Symposium on Integrated Network Management (IM'07),2007.

[9] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle, "Managing energy and server resources in hosting centers," in Proc. Of the ACM Symposium on Operating System Principles (SOSP'01), Oct.2001.

[10] T. Das, P. Padala, V. N. Padmanabhan, R. Ramjee, and K. G. Shin, "Litegreen: saving energy in networked desktops using virtualization," in Proc. of the USENIX Annual Technical Conference, 2010.

[11] Y. Agarwal, S. Savage, and R. Gupta, "Sleep server: a software-only approach for reducing the energy consumption of personal computers within enterprise

environments," in Proc. of the USENIX Annual Technical Conference, 2010.
[12] N. Bila, E. d. Lara, K. Joshi, H. A. Lagar-Cavilla, M. Hiltunen, and M. Satyanarayanan, "Jettison: Efficient idle desktop consolidation with partial VM migration," in Proc. of the ACM European conference on Computer Systems (EuroSys'12), 2012.
[13] Zhen Xiao, Weijia Song, and Qi Chen, "Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, 2013.