Energy Efficiency in Cloud Data Centers Using Load Balancing

Ankita Sharma *, Upinder Pal Singh **

** Research Scholar, CGC, Landran, Chandigarh*

*** Assistant Professor, CGC, Landran, Chandigarh*

ABSTRACT

Cloud computing is an expanding area in research and industry today, which involves virtualization, distributed computing, internet, and software and web services. This paper presents an approach for scheduling algorithms that can maintain the load balancing. In this research work we have developed power optimization algorithm which over comes the limitations of the previous algorithms[Round Robin, Equally Spread Current Execution Algorithm, Throttled Load Balancing which are used for the over load management of the data leading to positive consequences in terms of overall power consumption of the data centre thus helping in green computing. As due to undue overload of traffic and then overhead due to mitigation and migration of the virtual machines to balance out the operations there is always an impact on the power consumption, if there is more overload, there is bound to be more power consumption, and if balancing works well, there is bound to be an optimized trade-off for energy consumption. Results have shown that overall impact of power consumption is reduced by using the proposed algorithm.

Keywords- Cloud computing, Green computing, Load balancing, Round robin, equally spread current execution, Throttled, Virtual machine.

I. INTRODUCTION

Cloud Computing : Cloud computing can be defined as “a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers”.[4]

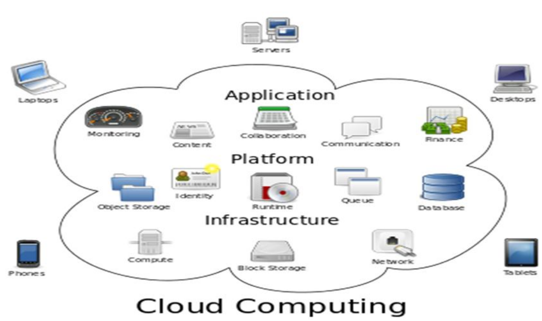

Fig 1: Cloud Computing

Green computing in clouds:

Green Computing, or Green IT, is the practice of implementing policies and procedures that improve the efficiency of computing resources in such a way as to reduce the energy consumption and environmental impact of their utilization. As High Performance Computing (HPC) is becoming popular in commercial and consumer IT applications, it needs the ability to gain rapid and scalable access to high end computing capabilities. This computing infrastructure is provided by cloud computing by making use of datacenters. It helps the HPC users in an on-demand and payable access to their applications and data, anywhere from a cloud. [6]

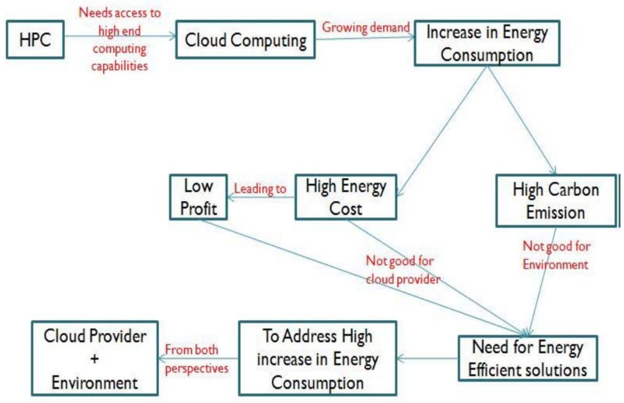

Fig 2: Green Computing in Clouds

Need of load balancing in cloud computing: Load balancing in clouds is a mechanism that distributes the excess dynamic local workload evenly across all the nodes. It is used to achieve a high user satisfaction and resource utilization ratio making sure that no single node is overwhelmed, hence improving the overall performance of the system. Proper load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc. [6]

Apart from the above-mentioned factors, load balancing is also required to achieve Green computing in clouds which can be done with the help of the following two factors:

- ***Reducing Energy Consumption*** - Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a cloud, hence reducing the amount of energy consumed.

- ***Reducing Carbon Emission*** - Energy consumption and carbon emission go hand in hand. The more the energy consumed, higher is the carbon footprint. As the energy consumption is reduced with the help of Load balancing, so is the carbon emission helping in achieving Green computing.

II. LITERATURE REVIEW

1) Anton Beloglazov and Rajkumar Buyya: Anton and Buyya have proposed Markov chain model to overcome limitations of heuristic or statistics based approaches to identify and mitigate overloading problems in datacenters. They have proposed a novel approach that for any known stationary workload and a given state configuration optimally solves the problem of host overload detection by maximizing the mean inter-migration time under the specified QoS goal based on a Markov chain model. They heuristically adapt the algorithm to handle unknown non-stationary workloads using the Multisize Sliding Window workload estimation technique. Through simulations with real-world workload traces from more than a thousand PlanetLab VMs, it was shown that the proposed approach outperforms the best benchmark algorithm and provides approximately 88% of the performance of the optimal offline algorithm.

2) Daniel J. Abadi: In his study it discuss the limitations and opportunities of deploying data management issues on cloud computing platforms. He anticipated that large scale data analysis tasks, decision support systems, and application specific data marts are more likely to take advantage of cloud computing platforms than operational, transactional database systems. A review of the currently available open source and commercial database options that can be used to perform such analysis tasks, and concluded that none of these options, as presently architected, match the requisite features. He concluded that a hybrid solution that combines the fault tolerance, heterogeneous cluster, and ease of use out-of-the-box capabilities of MapReduce with the efficiency, performance, and tool plugability of shared-nothing parallel database systems could have a significant impact on the cloud database market.

3) Qi Zhang, Lu Cheng and Raouf Boutaba: Zhang et al in their study presented a comprehensive review of literature on cloud computing. They concluded that even though cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand. However, despite the fact that cloud computing offers huge opportunities to the IT industry, the development of cloud computing technology is currently at its infancy, with many issues still to be addressed.

4) Bhaskar Prasad Rimal, Eunmi Choi and Ian Lumb: In this study researchers have first developed a comprehensive taxonomy for describing several cloud computing architectures. Then they have send this taxonomy to survey these cloud computing services. The researchers have concluded that there is a need for further progress on several issues such as standards, portability, interoperability, mappings to business architecture, security and privacy, multi – supplier and hybrid sourcing, management and governance plus business analytics for clouds.

5) Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao and Shun-Sheng Wang: As known cloud computing refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner utilizing the computing resources (service nodes) on the network to facilitate the execution of complicated tasks that require large-scale computation. Thus, the selection of nodes for executing a task in the cloud computing must be properly considered to exploit the effectiveness of the resources. Wang et al, in this study, propose a two-phase scheduling algorithm under a three-level cloud computing. The proposed scheduling algorithm combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms that can utilize more better executing efficiency and maintain the load balancing of system.

III. PROBLEM FORMULATION

From high power consumption problems arises:

- Insufficient or malfunctioning cooling system can lead to overheating of the resources reducing system's reliability and devices lifetime.
- High power consumption by infrastructure leads to substantial carbon dioxide emissions contributing to green house effect.

The primary purpose of the cloud system is that its client can utilize the resources to have economic benefits. A resource allocation management process is required to avoid underutilization or overutilization of the resources which may affect the services of the cloud. Some of the jobs may be rejected due to the overcrowding for the virtual machines by the current jobs in the cloud system. Resource allocation and Efficient scheduling is a precarious characteristic of cloud computing based on which the performance of the system is evaluated. The examined characteristics have an impact on cost optimization, which can be obtained by improved overall response time and data processing time with the help of improved algorithm.

IV. EXISTING WORK

Existing Load Balancing Algorithms for Cloud Computing: Distribute workload of multiple network links to achieve maximum throughput, minimize response time and to avoid overloading. We use three algorithms to distribute the load. And check the performance, time and cost.

- ***Round Robin Algorithm (RR):*** It is the simplest algorithm that uses the concept of time quantum or slices. Here the time is divided into multiple slices and each node is given a particular time quantum or time interval and in this quantum the node will perform its operations.
- **Equally Spread Current Execution Algorithm (ESCE):** In spread spectrum technique load balancer makes effort to preserve equal load to all the virtual machines connected with the data centre. Load balancer maintains an index table of Virtual machines as well as number of requests currently assigned to the Virtual Machine (VM). If the request comes from the data centre to allocate the new VM, it scans the index table for least loaded VM.

- **Throttled Load Balancing Algorithm (TLB):** In this algorithm the load balancer maintains an index table of virtual machines as well as their states (Available or Busy). The client/server first makes a request to data centre to find a suitable virtual machine (VM) to perform the recommended job. The data centre queries the load balancer for allocation of the VM. The load balancer scans the index table from top until the first available VM is found or the index table is scanned fully.

V. PROPOSED WORK

Earlier for implementing green computing in cloud data centers some policies or techniques has been introduced. But now we will use :

- Round robin algorithm
- Equally Spread Current Execution Algorithm (ESCE)
- Throttled Load Balancing Algorithm (TLB)

I will also create an algorithm which would reduce the overall energy consumption in data center and it would be also able to confirm SLA'S.

Our algorithm will be expandable to both public and private cloud .Our algorithm will have five components which includes:

- Power monitoring system
- Power optimizing algorithm
- Power consumption calculator
- Reconfigurable temporal policy of workload management
- Feedback loop

We will be comparing the above three load balancing algorithms with the newly created one in order to show that how the energy consumption is reduced.

VI. EXPERIMENTAL SETUP

Simulation: Simulation is a technique where a program models the behavior of the system (CPU, network etc.) by calculating the interaction between its different entities using mathematical formulas, or actually capturing and playing back observations from a production system. The available Simulation tools in Cloud Computing today are: simjava, gridsim and CloudSim. [5]

CloudSim using Java 1.7: CloudSim is a framework developed by the GRIDS laboratory of University of Melbourne which enables seamless modeling, simulation and experimenting on designing Cloud computing infrastructures. CloudSim is a self-contained platform which can be used to model data centers, host, service brokers, scheduling and allocation policies of a large scaled Cloud platform. This CloudSim framework is built on top of GridSim framework which is also developed by the GRIDS laboratory. Hence, the researcher has used CloudSim to model datacenters, hosts, VMs for experimenting in simulated cloud environment. [5]

VII. OBJECTIVES OF THE STUDY

1. Develop simulated monto carlo based cloud based simulated model.
2. Develop and build load balancing [Round Robin, Equally Spread Current Execution Algorithm, Throttled Load Balancing] Algorithm and observe power consumption.
3. Based on the observations and limitations of the above algorithms develop an improved algorithm.
4. Evaluate all the algorithms.

VIII. METHODOLOGY

The complete flow chart of the methodology is depicted as below:

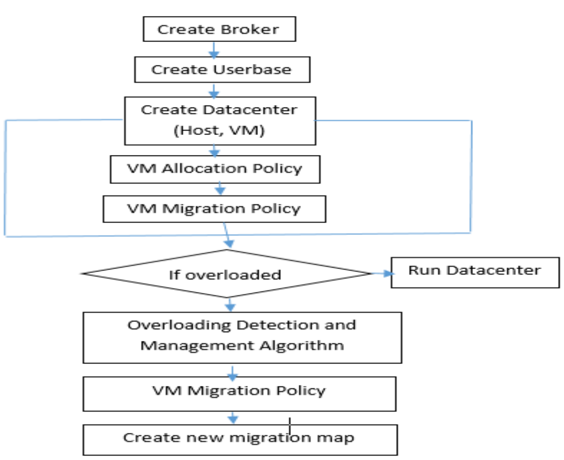

Fig 3: Flowchart

IX RESULTS

In this we shall discuss the result of the experimental results done on the cloud simulator with the existing and proposed optimization methods. However, here is a table showing the parameters involved in evaluating the performance of the algorithms.

TABLE I

• **Total Number Of Experiments**	**15**
• **Total Number Of Performance Parameters Evaluated**	**10**

TABLE II

Parameters used in the simulation include the following based on which performance of algorithm has been taken in terms of energy consumption

Performance Parameter	**Definition**	**How it helps in evaluation**
Number of Host	Host executes actions related to management of virtual machines (e.g., creation and destruction). A host has a defined policy for provisioning memory and bw, as well as an allocation policy for Pe's to virtual machines. A host is associated to a datacenter. It can	Number of Host determines how big is the datacenter and how many machines can be used in case they fail or become overloaded /over subscribed

	host virtual machines.	
Number of Virtual Machines	It runs inside a Host, sharing host List with other VMs. It processes loudest. This processing happens according to a policy, defined by the Cloudlet Scheduler. Each VM has an owner, which can submit cloudlets to the VM to be executed.	Number of VMs determines how much of the workload it can process and it follows a policy.
Number of VM migrations	It is the total Number of total Migrations a Power Optimization algorithm have done in the experiment.	It determines how many VM migration were done by the VM allocation and migration algorithm.
VM selection mean, **VM selection stDev** **(Execution time)**	Time in seconds taken in selecting VM of out allocated Vms when work is to be executed in the event of normal work schedule or in case or detection of overload or	It is the descriptive statistics test which is the quantitatively describing the main features of a collection of data (VM selection time)

	adversity.	
VM reallocation mean **(Execution time)** **VM reallocation stDev** **(Execution time)**	Time in seconds taken by the Power Optimization /reallocation algorithm.	It is the descriptive statistics test which is the quantitatively describing the main features of a collection of data (VM selection time)
Total mean time (Execution time)	This is the total mean of time taken in seconds	It is the descriptive statistics test which is the quantitatively describing the main features of a collection of data (VM selection time)
Total stDev time (Execution time)	This is the total standard deviation of time taken in seconds.	It is the descriptive statistics test which is the quantitatively describing the main features of a collection of data (VM selection time).
Parameter safety	It is the threshold value which determines how much of work is safe to process.	Utilization of CPU

Graphs:

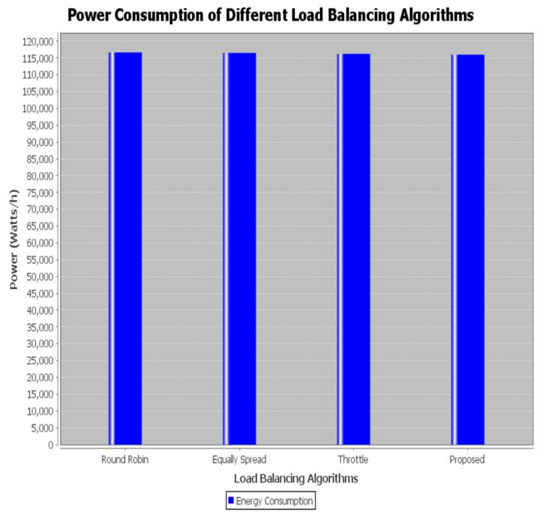

Fig 4: Energy Consumption

It is apparent from the above bar graphs that the power optimization consumes less as compared to the other algorithms , the power is calculated in terms of watts/h , and all the calculations were based on the multiple experiments with constant parameters with respect to the number of hosts, Number of Virtual machine , the ratio between the host and VM , as well as other factors which included the use of same VM allocation and migration policy with particular threshold as safety parameter [load value after which it is overloaded with the workload /cloudlet] to it reach maximum level of workload processing. It seems that proposed algorithm is able to reduce the power consumption due to the fact in this less number of migrations are occurring due to proper load balancing and due to the fact the CPU utilization is co –factored with bandwidth. If there is more consumption of internode bandwidth or more consumption with external network of the datacenter it leads to rationalization of workload by consolidation with underutilized machines which automatically reduces the load and hence the voltage usage and consequently the energy consumption.

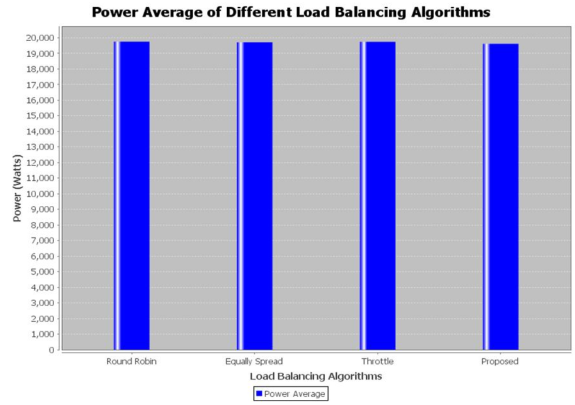

Fig 5: Power Average

This bar graph shows the power or the energy consumption in terms of the mean or faction of time for which the whole set of host and VM have consumed the energy/power over period of time it has worked since it has been allocated work or cloudlet if formed , and it can be seen from the above graph that this is also taking similar shape and magnate propositional to the first energy bar graph , the proposed power optimization algorithm is taking less energy as compared to the round robin , equally spread and throttle , all it seems the reduction in energy consumption is about 1 % less than compared to others .

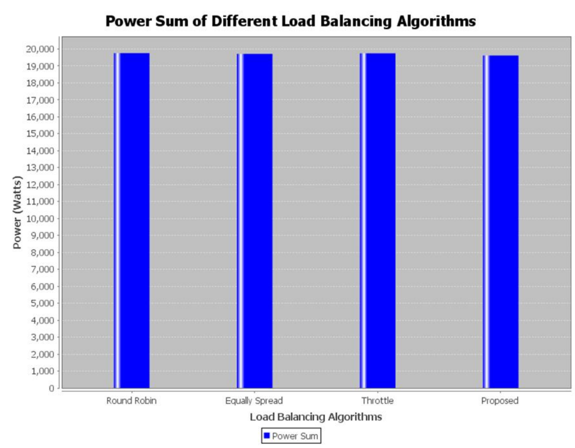

Fig 6: Power Sum

It can be seen from the above graph that the power consumption when sum up for the whole datacenter is less in case of the new proposed algorithm due to the fact that the better consumption pattern of data usage and processing is clear from two factors which include the bandwidth (datacenter internal and external bandwidth link) . This measure is calculated on the basis of adding up of the energy consumption of power aware datacenter having host and other devices.

X CONCLUSION

In this research work we have developed an algorithm that helps in optimization of the energy consumption along with the operation of load balancing in data centre which is power aware , and it has been found that when the proposed algorithm in tandem with other policies that help to run the data center is able to reduce energy consumption due to the fact , whenever , uneven balance of workload in terms of either under utilization or over utilization our algorithm works to consolidate the work load processing by introducing data load controller which also takes into account the inter –node bandwidth also , as some servers or host machines may be having VM that are working in some precedence or in some particular workflow , or simply there is dependency between them to process the workflow steps , thus need some level of optimization . Thus our algorithms takes one step further in green computing.

IX. FUTURE SCOPE

Energy- efficient computing cannot be achieved without the integration between computer science, electrical engineering, mechanical engineering, and environmental science. Designing data centers for the developing regions require a vertically integrated efforts to drive key energy-efficient technologies in computing (cloud computing), electronics (low power CPUs and systems), and building systems (spot rack cooling, higher ambient temperatures, and natural convention cooling). Collectively, these technologies address very significant near-term and long-term energy challenges and environmental issues. The Green Cloud Computing in which cloud computing using low power CPUs servers, and renewal energy and most important, which is closer to the end user. The future must have an approach for a low energy use data centers using cloud computing that are powered with renewable energy options optimized by the highly predictive algorithms that are based on intelligent network of agents which keep track of demand and supply of both sources of power and consumer of power from various distributed regions.

X REFERENCES

[1] Anton Beloglazov and Rajkumar Buyya, *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, July 2013, pages 1366 -1379.

[2] Bhaskar Prasad Rimal, Eunmi Choi and Ian Lumb, *A Taxonomy and Survey of Cloud Computing Systems*, Fifth International Joint Conference on INC, IMS and IDC 2009.

[3] Daniel J. Abadi, *Data Management in the Cloud: Limitations and Opportunities*, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 2009.

[4]Available:http://www.studymode.com/essays/Cloud-Computing-47687250.html

[5] Jasmin James and Dr. Bhupendra Verma, *Efficient Vm Load Balancing Algorithm for a Cloud Computing Environment*, Jasmin James et al. / International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 4 No. 09 Sep 2012.

[6] Nidhi Jain Kansal and Inderveer 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. [7] Qi Zhang, ·Lu Cheng and ·Raouf Boutaba, *Cloud computing: state-of-the-art and research challenges*, J Internet Serv Appl (2010) 1: 7–18.

[8] Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao and Shun-Sheng Wang, *Towards a Load Balancing in a Three-level Cloud Computing Network*, 978-1-4244-5539-3/10 ©2010 IEEE.