International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

The Improved Throttled Algorithm of Load Balancing on Cloud Computing

Sweeti Sahu[1], Prof. Saurabh Sharma[2], Prof. Vishal Paranjape[3], Prof. Vikash Verma[4]
[1,2,3,4]*Global Nature Care Sangathan Group of Institutions, Jabalpur(M.P), India*

***Abstract*— In a cloud computing context, load balancing is crucial. The performance of cloud computing is enhanced by a better load balancing algorithm, which offers a more effective service. This paper proposes a new dynamic time quantum priority-based load balancing method that works well in cloud computing environments. In a cloud computing environment, round robin is a very simple algorithm, but its drawbacks include long waiting times, slow reaction times, and overhead for context transitions. In this research, we present a novel dynamic round robin algorithm that uses an execution time slice to calculate the dynamic time quantum (TQ) for each service request in each round. The Priority Component (PC) of the process has been calculated in this case using the priority. Our experimental findings show that lowering average waiting times, overall context shifts, and average turnaround times results in greater performance.**

***Keywords*—cloud computing, load balancing, round robin algorithm, turnaround time, waiting time, context switch.**

I. INTRODUCTION

Organizational resources can access a variety of cloud services, including servers, storage, and applications, through cloud computing, an internet-based computing service. In the IT sector, it is significant. By offering the flexibility of resources that were divided among the developers, it has demonstrated how to construct software. In addition to being able to serve several users concurrently, cloud computing also allows for resource reallocation based on user demand. Users have the option to view their data from anywhere in the world thanks to this service. A financial business model that demonstrates multiple approaches to managing IT resources is provided by cloud computing [1]. The most crucial role in a cloud computing system is load balancing. The process of allocating resources so that no one resource is overused and that all are being used to their full potential is known as load balancing [2]. Each resource operates more quickly, reducing the reaction time. Any resource should not be under or overloaded, according to the load balancer [3]. It is difficult to maintain all constraints in a cloud computing environment, including security, dependability, and throughput.

Virtual Machine (V.M.) is one of a select few components that need additional attention [4]. A virtual machine (V.M.) is a software platform that replaces a physical machine and allows for the execution of services just like a real machine. Although there is no direct link to any real hardware built, it operates quite effectively and functions like a real machine.

Datacenters (DCC) store, manage, share, circulate, and protect an organization's data. Therefore, users have received infrastructure level services. It is simply the platform that houses the IT components. The brain, or central component, of a cloud is the platform where the most vital services run. Here, the server has been seen as a component of the data centre. The load balancer controls and distributes load to several virtual machines such that nodes are overly or underloaded. Since any one of the nodes failing could result in data loss or being unavailable, load balancing must be done correctly [8][1].

The two types of load balancing algorithms depend on the implementation strategy. [5]. The system is currently in a state where the Static Load Balancing Algorithm has no bearing on the choices made by the load balancer. The load balancer decides which virtual machine the service will be allotted to and run on. While the dynamic load balancing algorithm is based on the machine's present state. The load balancer determines each virtual machine's current load before allocating service to the most appropriate and suitable virtual machine. The load balancer must attempt to satisfy the following properties in any load balancing method, including the maximum number of context switches, throughput, CPU utilisation, minimal turnaround time, response time, and waiting time.

Following is how the remainder of the paper is organised. A brief synopsis of related work is provided in Section II. The dynamic time quantum priority-based technique that has been suggested and is suitable for load balancing in a cloud computing environment is discussed in Section III. Results from the experiment in Section IV are displayed. The conclusion is discussed in section V.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

II. Related Work

The load balancer uses the concept of time slices in basic round robin algorithm. In this method, the execution time is distributed into multiple time slices and each node gets a particular time interval like the concept of time scheduling. Within given time quantum each service request is served by the processor. After completion of each time slice, the next user can request and it comes for processing. If the client's service request completes its execution within assigned time quantum or time slice then user would not wait for next round otherwise user has to wait for next round. The basic and primary advantage of this algorithm is that the load balancer allocates the equal time for every service request which ensure fairness.

Been defined by sum of maximum burst time and minimum burst time which is divided by 2. In FairRR[7] algorithm they have used the bust time(BT) to calculate the dynamic time quantum(TQ) for every process which is in ready queue. First, they choose minimum bust time from all the processes and then allocate dynamic TQ to each process where TQ is min BT.

In Priority Based Round Robin algorithm [6] and Intelligent Time Slice (ITS) has been estimated and allocated the different processing time to each service according to their assigning priority basis. All above methods focus on different measuring parameters of this algorithm. The following parameters are

A. Burst Time (BT)

BT is actually time that is required to complete execution of particular service request.

B. Time Quantum (TQ)

TQ is the given time period for a service to allow to access a VM.

III. Implementation

Cloudsim Software

- CloudSim is a simulation tool that allows Cloud developers to test performance of their provisioning policies in a repeatable and controllable environment free of cost. It helps
- to tune the bottlenecks before real-world deployment. It is a simulator; hence it doesn't run any actual software. It can be defined as "running a model of an environment in a model of hardware" and technology specific details are abstracted.
- CloudSim is It is basically a Library for Simulation of Cloud Computing Scenarios. It has some features such as it support for modeling and simulation of large scale Cloud Computing infrastructure, including data centers on a single physical computing node. It provides basic classes for describing data centers, virtual machines applications, users, computational resources, and Policies.
- It also supports evaluation of Green IT policies. User can use it as building blocks for simulated Cloud environment and can add new policies for scheduling, load balancing, and new scenarios. It is flexible enough to be used as a library that allows you to write a desired scenario by writing a Java program
- The simulation and analysis of the performance of the three load balancing algorithms are performed using the "Cloud Analyst" tool [5]. It allows the user to run multiple simulations with small parameter changes, and also allows you to customize the location of the users who create the application and the location of the data centers [6]. Let's indicate the terminology of the emulator (Fig. 4):
    - Region: in Cloud Analyst, the world is divided into 6 regions that coincide with the 6 major continents in the world;
    - User Base: User Base is considered as a single unit, and is used to generate traffic;
    - Data Processing Center: brokerage services determine which center should accept and process the request that comes from each user database;
    - VmLoadBalancer: it is responsible for distributing the load to the available data center. VmLoadBalancer distributes the load in the data center based on the load balancing policy.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

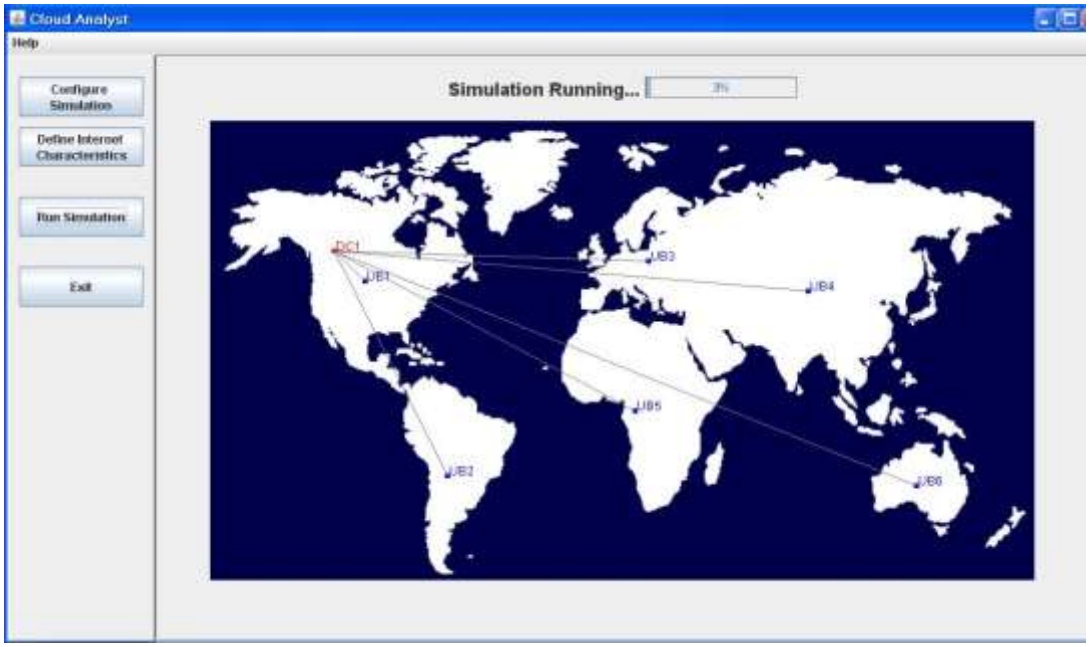

Fig. 4. Cloud Analyst Simulator.

In the modeling process, CloudSim 4.0 software was used.

Data Center Architecture:

The data center is home to the computational power, storage, and applications necessary to support an enterprise business. The data center infrastructure is central to the IT architecture, from which all content is sourced or passes through.

Proper planning of the data center infrastructure design is critical, and performance, resiliency, and scalability need to be carefully considered.

Region Division:

The whole earth is divided into five regions to match with the classification of cloud analyzer. Region divisions and the countries within these regions are as follows:[5]

S.No.	Region	Countries
1	R0	USA
2	R1	Countries of North America
3	R2	Countries of European union
4	R3	Countries of Asia like china, India
5	R4	Country from Africa
6	R5	Australia

Here the simulation and performance analysis will be done by using the cloud analyst tool, as depicted in Figure 5.1.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

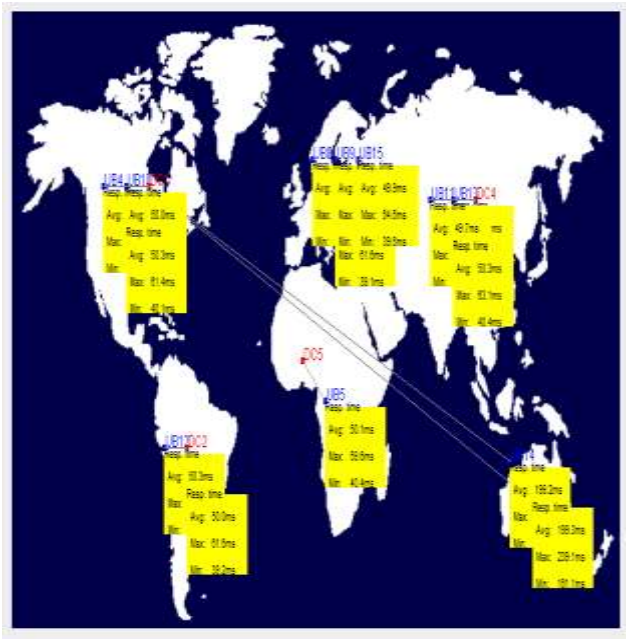

Fig. 5.2: The simulation and performance analysis diagram by using the cloud analyst tool

IV. DATA CENTER NETWORK DESIGN

The Design and simulation for performance analysis will be done by using the OPNET simulation software, as shown in Figure 5.2 bellow.

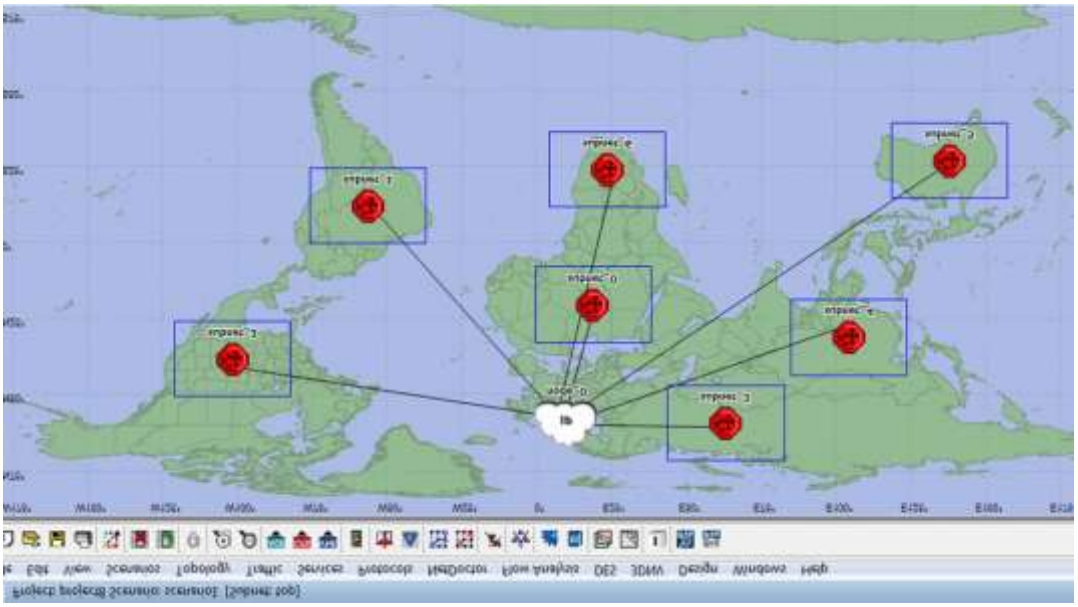

Fig. 5.3: The simulation, design and analysis by using

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

OPNET

Fig. 5.4 illustrate the design of data-centers subnets, it show that each subnet include storage devices, server, routers and workstations. There are five subnets each of them are about data-center and they connected together by wire line.

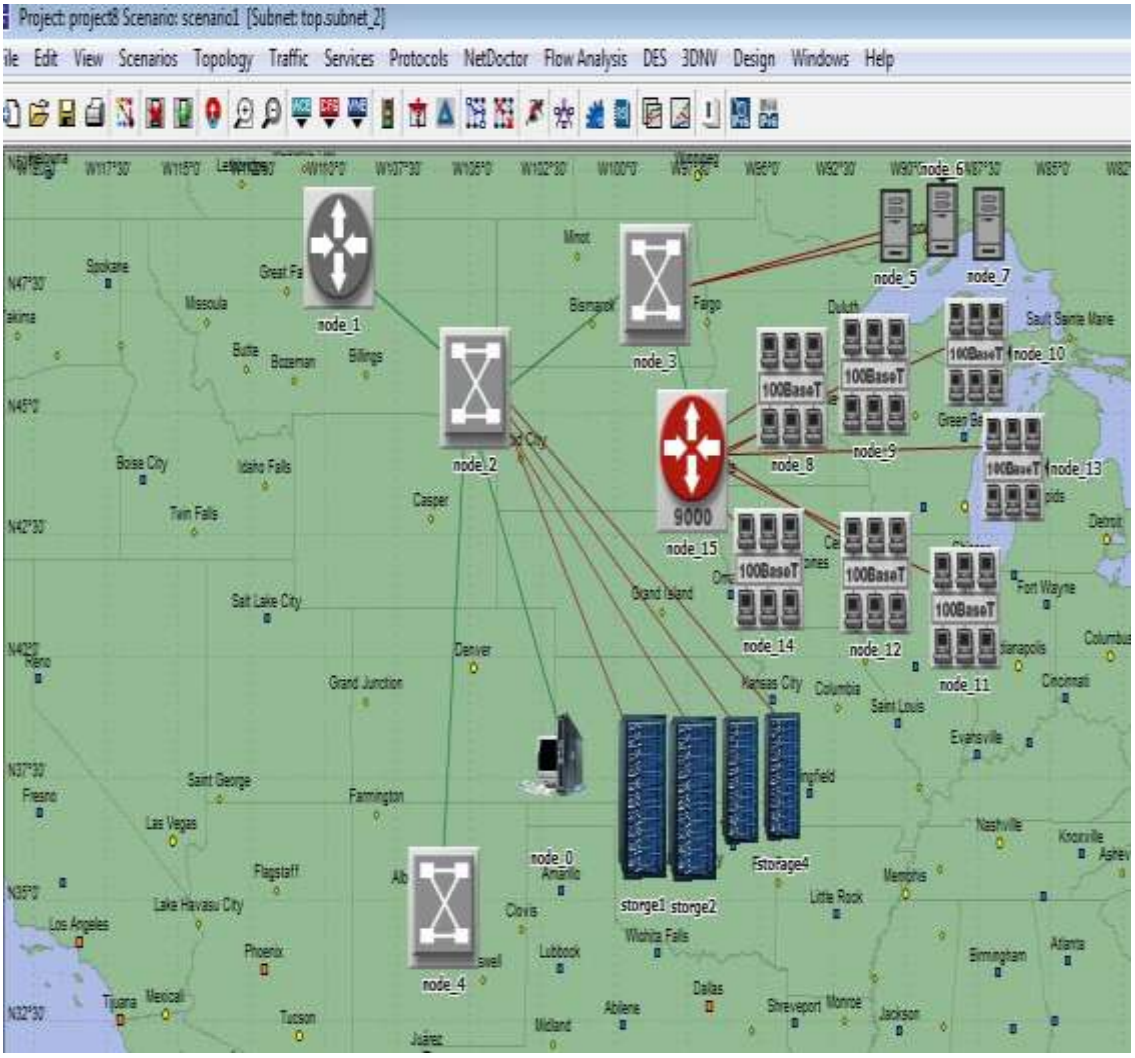

Fig. 5.4: The simulation and design of datacenter (subnet)

Each region contains one data-center by using OPNET Data Center Configuration Parameter: Total 05 data center will be considered for the simulation environment. Architecture of each data centre is given in table (2)

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

TABLE 3.1
Initial conditions

VM	Core units	Hours
VM1	2	3
VM2	1	1
VM3	6	6
VM4	3	2
VM5	3	5
VM6	2	2

Here we define VM, VM core units and time. Total capacity of physical machine is 9 unit. Show the matrix, it is initial stage, If VM complete before 3.5 hours than VM is short life job otherwise it is long life job.

In second stage incoming VM1 needs 3 hours time for running the job. here VM1 runs before 3.5 hours so VM1 is short life job.

TABLE 3.2
Stage:2 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1(3 hours)	
Long life		
Remaining unit	7	

Therefore VM1 store in 1st physical machine and need 2 core unit. So remaining capacity of physical machine is 7 unit.

TABLE 3.3
Stage: 3 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1(2 hours) VM2(1 hours)	
Long life		
Remaining unit	6	

In third stage new VM2 is arrive which need 1 hour for run and 1 unit capacity. So our remaining capacity is 6 unit for 1st physical machine.

In table VI, new VM3 is arrive. At the same time VM1 completed 1hour work and VM2 is finish.VM3 require 6 hours and 6 unit capacity. So our remaining unit is 1.

In the fifth stage, remaining hour of VM1 and VM3 is 1 and 5. New VM4 enter with 2 hours and 3 unit. But physical machine 1 is contain only one unit capacity so VM4 place in physical machine 2.

In sixth stage, VM5 is enter with 5 hours and 1 unit capacity. VM5 first check 1st physical machine. PM1 contain remaining capacity of 1 unit. So VM5 place in 1st physical machine. So remaining unit is 0 for PM1.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

TABLE 3.4
Stage: 4 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1(1 hours) VM2 is Finish	
Long life	VM3(6 hours)	
Remaining unit	1	

TABLE 3.5
Stage: 5 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1(1 hours)	VM4(2 hours)
Long life	VM3(5 hours)	
Remaining unit	1	6

TABLE 3.6
Stage: 6 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1 is finish	VM4(2 hours)
Long life	VM3(4 hours) VM5(5 hours)	
Remaining unit	0	6

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

TABLE 3.7
Stage: 7 of example

	PM1 (Total unit= 9)	PM2 (Total unit= 9)
Short life	VM1 is finish	VM4(2 hours) VM6(2 hours)
Long life	VM3(4 hours) VM5(5 hours)	
Remaining unit	0	4

In next stage, PM1 is full so newly coming VM place in PM2. Show the final condition of example in table IX.

V. RESULT

System Performance

In first graph, we define the placement process of VM. In existing work all short life and long life VM placement is done on same physical machine and in our proposed work, we arrange all short life VMs on same container and long life VM on same container.

So our resource utilization is maximizing, show in graph.

In second graph, we define utilization of resource or PM. According to the Optimization function if the division of VM(mips) and PM(mips) is near to the 1 than fragmentation is less, so utilization of PM is more. And the value is the nearest to the 0 than the fragment is more, so utilization is less. Base on that optimized value of 1 and 0 we create graph.

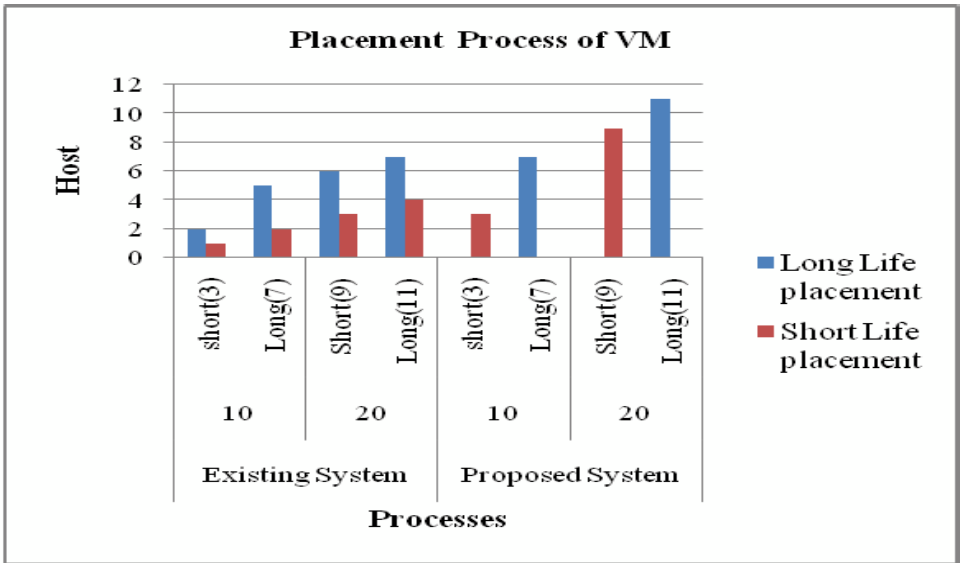

Fig. 5.0 Placement process of VM

In table 5.1, we define existing placement process. Here all short life and long life VM place in same PM.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

TABLE 5.1
Existing placement process

Existing system				
	10		20	
	Short(3)	Long(7)	Short(9)	Long(11)
Long life placement	2	5	6	7
Short life placement	1	2	3	4

Table 5.2 defined our proposed strategy. Where we create short life and long life container.

TABLE 5.2.
Proposed placement process

Proposed system				
	10		20	
	Short(3)	Long(7)	Short(9)	Long(11)
Long life placement	0	7	0	11
Short life placement	3	0	9	0

Table 5.3 shows the placement of short life job according to load in existing system and proposed system.

TABLE 5.3.
Placement of short life job

Placement of short life VM		
	Existing system	Proposed system
0 to 2	7	7
2 to 4	3	3
4 to 6	8	6
6 to 8	7	0
8 to 10	5	0

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 08, August 2022)

In fig. 5.2 we define utilization of resource or PM. According to the Optimization function.

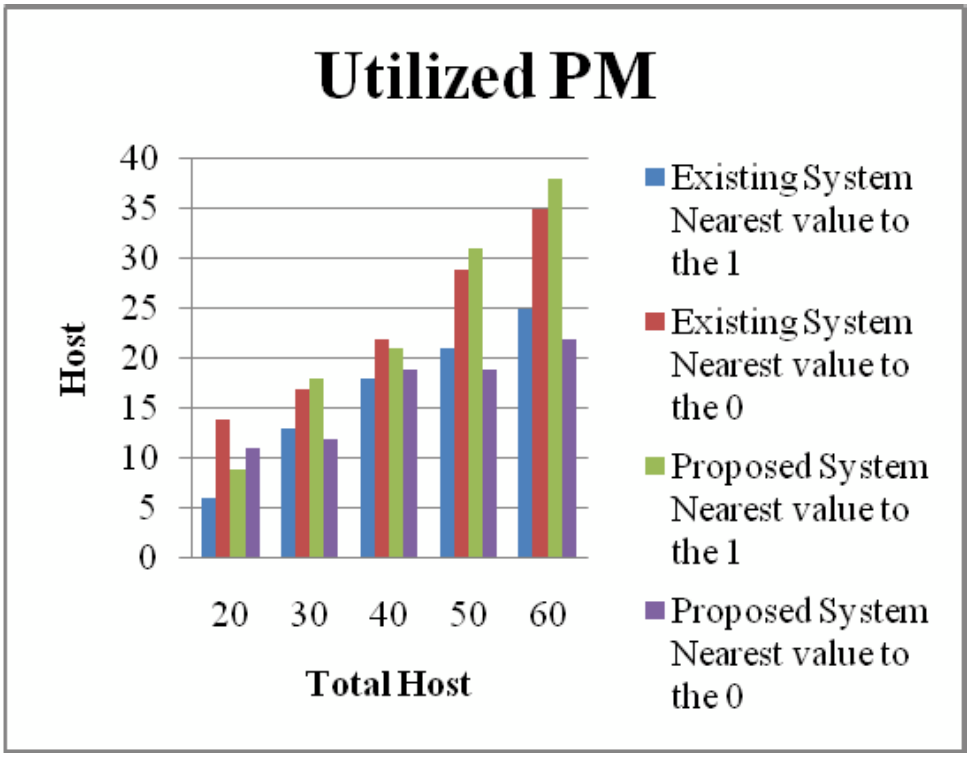

Fig. 5.2 Utilization of PM

Existing System			Proposed System		
	Nearest value to the 1	Nearest value to the 0	Nearest value to the 1	Nearest value to the 0	
20	6	14	9	11	
30	13	17	18	12	
40	18	22	21	19	
50	21	29	31	19	
60	25	35	38	22	

VI. Conclusion And Future Work

After doing rigorous survey on various issues in resource utilization, we found that load balancing, VM placement and fragmentation are the greatest issue in cloud computing. So we propose dynamic priority based spill over technique and add the concept of short life/long life container for solving the fragmentation issue. Our architecture maximize the resource utilization also we minimize fragmentation issue with using short life/long life container at physical machine in our architecture.

REFERENCES

[1] Amandeep, ‘‘Analysis of Load Balancing Techniques in Cloud Computing’’, International Journal of Computers & Technology, vol. 4, no.2, pp.2277- 3061, March-April, 2013.(references)

[2] Angona Sarker,Ali Newaz Bahar,SM Shamim, "A Review on Mobile Cloud Computing", International Journal of Computer Applications (0975 – 8887)

[3] Aarti Singha, Dimple Junejab, Manisha Malhotraa,"Autonomous Agent Based Load Balancing Algorithm in Cloud Computing"- International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015)

[4] B.Subramani, ‘‘A New Approach For Load Balancing In Cloud Computing’’, IEEE ,vol.2, pp. 1636-16405, May 2013.

[5] Buyya, Rajkumar. "Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility". In Proceedings of the 2009 9th IEEE/ACM.

[6] Geethu Gopinath P , Shriram K Vasudevan,"An in- depth analysis and study of Load balancing techniques in the cloud computing environment"-. 2nd International Symposium on Big Data and Cloud Computing (ISBCC’15).

[7] Ruhi Gupta. "Review on Existing Load Balancing Techniques of Cloud Computing." International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 2, February 2014.

[8] Jinhua Hu," A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", 3rd International Symposium on Parallel Architectures, Algorithms and Programming 978-0-7695- 4312-3/10 © 2010 IEEE (published)

[9] Jayant Baliga, Robert W. A. Ayre, Kerry Hinton, and Rodney S.1Tucker, Fellow IEEE, “Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport”.

[10] S. Kapoor, and C. Dabas, Cluster based load balancing in cloud computing, Proc. Eighth International Conference in Contemporary Computing (IC3), 2015, 76-81.

[11] Klaithem Al Nuaimi," A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", 2012 IEEE Second Symposium on Network Cloud Computing and Applications 978-0-7695-4943-9/12

[12] Mohit and Jitender Kumar, International Journal of Advanced Research in Computer Science and Software Engineering ,"A Survey of Existing Load Balancing Algorithms in Cloud Computing".

[13] S.Mohinder, R,Ramesh, D.Powar, ‘‘Analysis of Load Balancers in Cloud Computing’’, International Academy of Science, Engineering & Technology, vol.2, May 2013.

[14] Poulami Dalapati1, G. Sahoo,"Green Solution for Cloud Computing with Load Balancing and Power Consumption Management"- International Journal of Emerging Technology and Advanced Engineering,Vol3:2013

[15] Saurabh Kumar Garg and Rajkumar Buyya, “ Green Cloud computing and Environmental Sustainability”.(references)

[16] Sidhu A, S.Kinger, “Analysis of Load Balancing techniques in Cloud Computing”, Council for innovative research international Journal of Computer & Technology, vol.4, March-April 2013.

[17] Sumalatha M.R, C. Selvakumar, T. Priya, R. T. Azariah, and P. M. Manohar, "CLBC-Cost effective load balanced resource allocation for partitioned cloud system", Proc. International Conference on Recent Trends in Information Technology (ICRTIT), 2014, 1-5.

[18] The Amazon Elastic Compute Cloud (Amazon EC2), http://aws.amazon.com/ec2/

[19] Yilin Lu," A Hybrid Dynamic Load Balancing Approach for Cloud Storage", 2012 International Conference on Industrial Control and Electronics Engineering 978-0-7695-4792-3/12 © 2012 IEEE.

[20] Yuvapriya Ponnusamy, S Sasikumar,"Application of Green Cloud Computing for Efficient Resource Energy Management in Data Centres", International Journal of Computer Science and Information Technologies,Vol3:2012.