Adaptive Load Balancing Algorithm for Efficient Resource Utilization in Cloud Computing

Ms. Avani Patel[1], Mr. Krunal Panchal[2]

[1] P.G. Student, [2]Assistant Professor

[1, 2]Department of Computer Engineering

[1,2]L. J. Institute of Engineering and Technology, Ahmedabad, Gujarat, India

Abstract

Cloud Computing is an emerging area in the field of the Information Technology (IT). Load balancing is one of the important issues in cloud computing environments. It is required to distribute the dynamic workload across multiple nodes to ensure that no single node is overloaded .It helps to achieve a high user satisfaction and resource utilization ratio by balancing the dynamic workload among nodes. In this paper we propose an adaptive load balancing algorithm for efficient resource utilization in cloud computing. The proposed algorithm is dynamic in nature and considers the dynamic threshold value for load balancing thus it increases the resource utilization as compared to existing algorithm.

Keywords**—***Cloud Computing, Load Balancing, Virtual Machines, Workload*

I. INTRODUCTION

The definition of "cloud computing" provided by USA National Institute of Standards and Technology (NIST) is as follows: "Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1]."

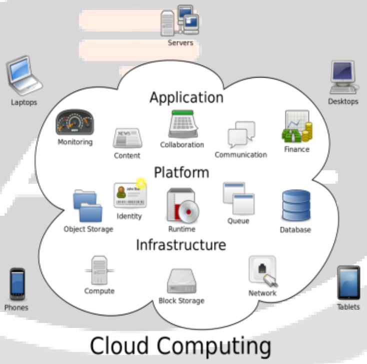

Fig.1 Cloud Computing [9]

The cloud means the applications and services that are offered from data center to all over the world. These applications and services are offered over the internet. The services provide by cloud computing are infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS) that are made available as pay-as-you-go model to clients[2].Cloud Computing Deployment Model refers to the location and management of the infrastructure cloud services. The Deployment Model of cloud computing are Private Cloud,

Community cloud, Public cloud and Hybrid cloud. Cloud Computing contain some essential characteristics that are rapid elasticity, on-demand self-service, resource pooling, broad network access, and measured service.

Load balancing is one of the major issues in cloud computing [3] it is a mechanism which distributes the dynamic local workload evenly across all the nodes in the whole cloud. It will avoid the situation where some nodes are heavily loaded while others are idle or doing little work. It helps to achieve a high user satisfaction and resource utilization ratio. Load balancing is the process of improving the performance of the system by shifting of workload among the processors. Workload of a machine means the total processing time it requires to execute all the tasks assigned to the machine [4]. The load balancer determines which web server should serve the request. The load balancer uses various scheduling algorithm to determine which server should handle and forwards the request on to the selected server [8].

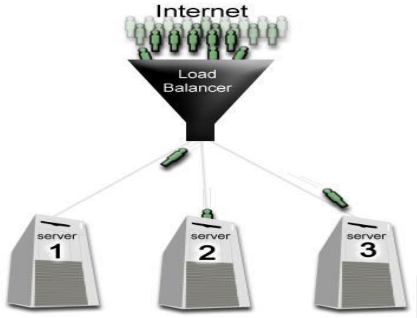

Fig.2 Load Balancer in Cloud Computing [8]

Load balancing is also needed for achieving Green computing in clouds [6]. The factors responsible for it are:

• Limited Energy Consumption:

Load balancing can reduce the amount of energy consumption by avoiding over hearting of nodes or virtual machines due to excessive workload [6].

• Reducing Carbon Emission:

Energy consumption and carbon emission are the two sides of the same coin. Load balancing helps in reducing energy consumption which will automatically reduce carbon emission and thus achieve Green Computing [6].

The goals of load balancing are:

- To improve the performance substantially
- Built a fault tolerant system using backup plans.
- To maintain the system stability
- To accommodate future modification in the system
- Increase resource utilization

II. RELATED WOK

A. Modified Throttled Algorithm

Like Throttled Algorithm it maintains an index table containing list of virtual machines and their states. The first VM is selected in same way as in Throttled. When a request arrives it searches the table and if a match is found on the basis of size and availability of the machine, then the request is accepted otherwise -1 is returned and the request is queued When the next request arrives, the VM at index next to already assigned VM is chosen depending on the state of VM and the usual steps are followed, unlikely of the Throttled algorithm, where the index table is parsed from the first index every time the Data Center queries Load Balancer for allocation of VM[16].

B. Central Load Balancer

Central Load Balancer (CLB) maintains a table containing the state of each VM along with their priority. The priority is calculated based on the CPU speed and capacity of memory. The VMassignment policy is similar to that of Throttled except that in this algorithm the VM with highest priority will get the first preference. If it is busy then the VM with next highest priority is checked and the process continues until a VM is found or the whole table is searched. The algorithm efficiently balances load in a heterogeneous environment but it suffers from bottleneck as all the requests will come to central load balancer. Moreover the algorithm is based on the priority of VMs which is calculated in a static way and is not updated during job allocation [10].

C. VM-Assign Load Balancing Algorithm

It maintains information about each VM and the number of requests currently allocated to which VM. When a request to allocate a new VM arrives, it identifies the least loaded VM. If there are more than one, the first identified is selected. Load Balancer returns the VM id to the Data Center Controller [10]. Then if next request comes it checks the VM table, if the VM is available and it is not used in the previous assignment then, it is assigned and id of VM is returned to Data Center, else we find the next least loaded VM and it continues. This algorithm will utilize all the VMs completely and this algorithm will not use the VM if it is already allocated in the last round [10].

D. Cloud Light Weight

A new load balancing method, named Cloud Light Weight (CLW), which not only balances the Virtual Machines'(VM) workload in cloud computing datacenters, but it also assures QoS for users. It reduces both the number of VM migration processes and the migration time during applications execution. It can also prevent SLA violation by balancing the system load based on the tasks requirements [13].

CLW is a dynamic load balancing approach, no assumption is made about the system workload and the decision making is based on the current system's status.

E. Agent Based Dynamic Load Balancing Algorithm

Agent Based Dynamic Load Balancing (ABDLB) is an approach in which mobile agent plays very important role, which is a software entity and usually defined as an independent software program that runs on behalf of a network administrator. It has ability to learn. This scheme makes system independent form constant monitoring of the servers for the purpose of load balancing, which the requirement of existing dynamic load balancing algorithms in cloud is computing. It greatly reduces the communication cost of servers, accelerates the rate of load balancing which indirectly improves the Throughput and Response Time of the cloud [14].

At the shared pool of servers, agent complete one cycle in two walks [14]:

- First walk it moves from initiation server to last Server and gathers information from all servers, for making appropriate decision for load balancing and
- In second walk it balances the server’s load on the basis of average load of the cloud.

F. Ant Colony Optimization (ACO) Algorithm

This technique of load balancing is based on Ant Colony Optimization (ACO) concept. ACO is inspired from the ant colonies that work together in foraging behavior. The ants work together in search of new sources of food and simultaneously use the existing food sources to shift the food back to the nest. The ants leave a pheromone trail upon moving from one node to another. By following the pheromone trails, the ant subsequently came to the food sources. The intensity of the pheromone can vary on various factors like the quality of food sources, distance of the food, etc. The ants use these pheromone trails to select the next node. A Data Center server is known as node in the proposed system [15].

Pheromone Updation

The ant will use two types of pheromone for its movement. They are:

- Foraging Pheromone (FP): While moving from underloaded node to overloaded node, Ant will update FP.
- Trailing Pheromone (TP): While moving from overloaded node to underloaded node, Ant will update TP.

The ACO Algorithm gives optimal resource utilization. The Performance of the system is enhanced with high availability of resources, thereby increasing the throughput [15].

III.PROPOSED LOAD BALANCING ALGORITHM

The proposed architecture for load balancing is as follows:

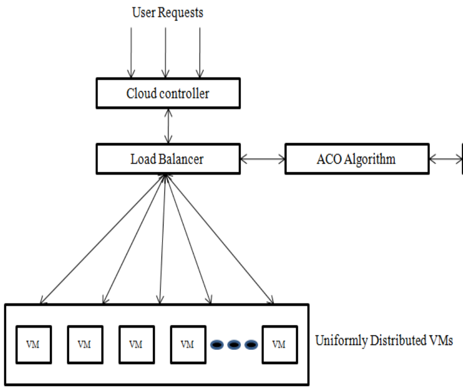

Fig.3 Proposed work model

Algorithm: Adaptive Load Balancing algorithm
Input: No of incoming jobs x1, x2……..xn ; Available VM y1, y2………yn
Output: All incoming jobs x1, x2…..xn are allocated one by one to the VM y1, y2………yn.

The proposed load balancing algorithm will balance the load among virtual machines. It will distribute the load based on dynamic threshold value. It maintains Global Instance queue and thus when user request arrives at the cloud controller it will parse this queue and assign the VM ate first index to the user request.

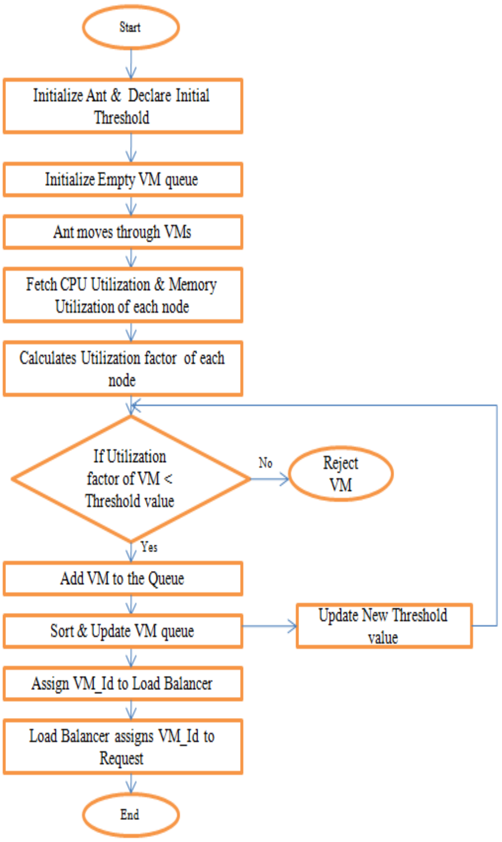
Start
Initialize Ant & Declare Initial Threshold
Initialize Empty VM queue
Ant moves through VMs
Fetch CPU Utilization & Memory Utilization of each node
Calculates Utilization factor of each node
If Utilization factor of VM < Threshold value
No
Reject VM
Yes
Add VM to the Queue
Sort & Update VM queue
Update New Threshold value
Assign VM_Id to Load Balancer
Load Balancer assigns VM_Id to Request
End

Fig.4 Proposed Flow Diagram

In the flow of the algorithm ant will periodically checks the health of all the nodes. Ant will fetch the CPU Utilization and Memory Utilization of each VM. And then it will calculate Utilization factor of these parameters for each VM. This utilization factor of VM will be compared with Threshold value. And if the condition satisfies than that VM will be added to the queue otherwise it will be rejected and thus Instance queue will be updated. New threshold value will be also calculated by considering the VMs in the queue.

Thus, when a user request arrives at cloud controller it queries load balancer. And the Load balancer will parse the queue and assign VM_ID to the cloud controller. Cloud controller sends request to VM_ID and notifies Load Balancer about resource allocation.

IV. IMPLEMENTATION AND RESULTS

The proposed algorithm is implemented using J2EE IDE and we have used the most central and well-known service of AWS known as Amazon Elastic Compute Cloud, also known as "EC2".Instances are launched and configured using EC2 service of Amazon Web Services. For experimentation we have observed 9 VMs over 400 user requests. Requests are distributed uniformly on VMs using dynamic threshold value.

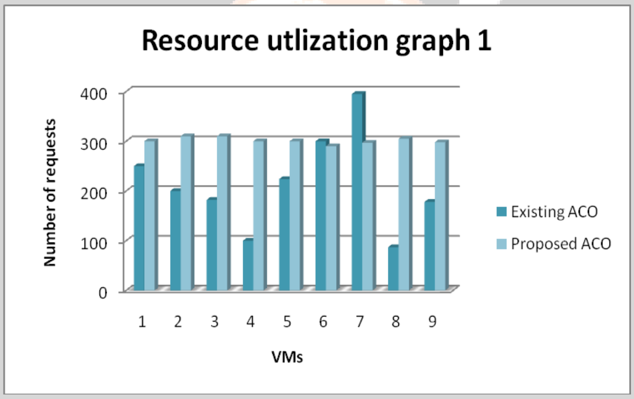

Fig.5 shows the graph for resource utilization with existing and proposed ACO technique.

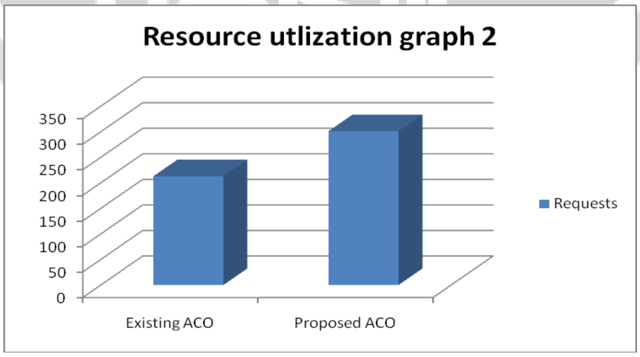

Fig.6 shows the graph for resource utilization with existing and proposed ACO technique. It shows that more number of requests are handled using proposed technique as compared to the existing technique. Thus, it has more resource utilization o f VMs as compared to existing technique.

V. CONCLUSION & FUTURE WORK

Load balancing is one of the main challenges in cloud computing since large numbers of requests are submitted to the data center. Existing load balancing techniques that have been studied does not consider current resource utilization for load distribution. Therefore, there is need to develop a load balancing technique that can increase the resource utilization. Thus, a load balancing technique based on load balancing technique based on Ant Colony Optimization is proposed for efficient resource utilization. Future work is to implement this technique with consideration of Response time of VM for efficient resource utilization in cloud computing.

VI. REFERENCES

[1] Lee Badger,TimGrance, Robert Patt-Corner,JeftVoas,"Cloud Computing Synopsis and Recommendations", National Institute of Standards and Technology, Special Publication 800-146, , May 2012.

[2] Martin Randles, David Lamb, A. Taleb-Bendiab, “A ComparativeStudy into Distributed Load Balancing Algorithms for Cloud Computing”,2010 IEEE 24th International Conference on Advanced Information

[3] Networking and Applications Workshops.B. P. Rima, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems", Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Korea, August 2009, pages 44-5l.

[4] Sandeep Sharma, Sarabjeet Singh, Meenaksshi Sharma, "Performance Analysis of Load Balancing Algorithms", World Academy of Science, Engineering and Technology, 2008.

[5] Hisao Kameda, EL-Zoghdy Said Fathyy and Inhwan Ryuz Jie Lix, "A Performance Comparison of Dynamic vs Static Load Balancing Policies in a Mainframe, Personal Computer Network Model",Proceedings Of The 39th IEEE Conference on Decision &Control,2000.

[6] Nidhi Jain Kansal, Inderveer Chana, "Cloud Load Balancing Techniques: A Step Towards Green Computing", IJCSI, Vol. 9, Issue1,January2012.

[7] Ghutke, B.; Shrawankar, U., “Pros and Cons of Load Balancing Algorithms for Cloud Computing” Information Systems and Computer Networks (ISCON), 2014 International Conference on, 2014, Pages: 123 - 127, DOI: 10.1109/ICISCON.2014.6965231

[8] Tejinder Sharma, Vijay Kumar Banga, “Efficient and Enhanced Algorithm in Cloud Computing,” International Journal of Soft Computing and Engineering (IJSCE), Vol. 3, Issue 1, pp. 2231-2307,2013.

[9] https://en.wikipedia.org/wiki/Cloud_computing visited on 1/5/16 at 4.45PM.

[10] Shaw,S.B.; SinghA.K.,“A survey on scheduling and load balancing techniques in cloud computing environment”Computer and Communication Technology (ICCCT), 2014 International Conferenceon 2014,Pages:8-795, DOI: 10.1109/ICCCT.2014.7001474

[11] Gulshan Soni , Mala Kalra “A Novel Approach for Load Balancing in Cloud Data Center” In Advance Computing Conference (IACC), 2014 IEEE International, 2014, DOI:10.1109/IAdCC.2014.6779427.

[12] Shridhar G.Damanal and G. Ram Mahana Reddy “Optimal Load Balancing in Cloud Computing By Efficient Utilization of Virtual Machines” In Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on , 2014 , DOI: 10.1109/COMSNETS.2014.6734930.

[13] Mohammadreza Mesbahi , Amir Masoud Rahmani , Anthony Theodore Chronopoulos “ Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing ” In International Conference on Data Science & Engineering (ICDSE) , 2014, DOI:10.1109/ICDSE.2014.6974610.

[14] Grover, J., Katiyar, S. “Agent based dynamic load balancing in Cloud Computing” Human Computer Interactions (ICHCI), 2013 International Conference on, 2013, DOI:10.1109/ICHCI-IEEE.2013.6887799

[15] "A Technique Based on Ant Colony Optimization for Load Balancing inCloud Data Center” Information Technology (ICIT), 2014 International Conference on , 2014 ,Pages: 12 - 17, DOI: 10.1109/ICIT.2014.54

[16] Shridhar G. Domanal, G. Ram Mohana Reddy, “Load Balancing in Cloud Computing Using Modified Throttled Algorithm”, IEEE, International conference. CCEM 2013.