RESEARCH ARTICLE OPEN ACCESS

# Survey of VM Load Balancing Algorithm in Cloud Environment

Er. Pooja [1], Er. Vivek Thapar [2]
Research Scholar [1], Assistant Professor [2]
Department of Information Technology
Department of Computer Science and Engineering
Guru Nanak Dev Engineering College
Punjab - India

**ABSTRACT**
Cloud computing is a general term for distributed the user request called load over the empty virtual machines. The distributed system faces the imbalance load over the virtual machines which degrade the performance of the cloud. Load balancing helps to improve the performance of cloud environment by dividing the workload over virtual machines in efficient manner using existing algorithms. The Virtual machine load balancing algorithm plays a very important role in load balancing of cloud. There are various categories in which we can divide the virtual machine different load balancing algorithm. We are going to discuss these various classifications virtual machine load balancing algorithm in our paper. We divide the paper into various sections and discuss various aspects. This paper presents the study of existing static or dynamic load balancing algorithm. This analysis can further help in the design improved algorithm.

***Keywords:*** - Cloud Computing; Load balancing; Static load balancing; Dynamic load balancing; Genetic Algorithm.

## I. INTRODUCTION

Cloud Computing is technology which provides "IT resources" and online storage over the internet to the users. Internet plays the most important role in a cloud. In Cloud Computing users can access resources all the time through internet without worrying about the hardware needed. In todays the business surroundings competes to finding ways to cut cost and maximize profit. Cloud computing comes up as one of the apparent technology. A new image of computing, cloud computing has appear to change the old ways of computing. This technology allows the information technology world to use computer resource effectively and more efficiently.

Cloud computing provide business opportunities to executives who can invest in acquiring servers as well as data storage computers to provide cloud services. The cloud computing is a virtual pool of computing resources such as software, platform, infrastructures, applications, storage and information provides to users through internet. Cloud Computing is a paradigm for extensive distributed computing that makes use of existing technologies such as virtualization, service-orientation, and grid computing. The resources and services can be accessed by scaling up or down as per consumer demands. Cloud Computing providers typically charge consumers on a pay-per-use model.

Load balancing is the one of the challenging task in cloud environment. In load balancing to balance the load equally to each virtual machine and to enhancing the overall performance of system or performance it use different techniques. Some problem during load balancing occurred due to bandwidth usage, network congestion etc. There are some existing algorithms which are able to solve these types of problems. The load balancing algorithm is dividing into two types they are static or dynamic algorithm. This paper divided into two sections. In section II, we are going to discuss the various existing static algorithms. These different available algorithm does not able to change the according to the user requirements. Static algorithms provide the homogeneous environment to the users. In the static algorithm it depends on the previous knowledge of the system. In section III, we are going to discuss the various dynamic algorithms. It provides the heterogeneous environment to the users. Dynamic algorithm gives better performance than the static algorithm and it based on the current system.

## II. STATIC LOAD BALANCING ALGORITHM

The static load balancing algorithm used in homogeneous environment and it does not able to changing user

requirements as well as load. Static load balancing methods are cooperative multitasking that means once the load is allocated to the virtual machine it cannot be transferred to another virtual machine. The static load balancing method helps to reduce the execution time to complete a single task. The static algorithm while distributing the load over the various nodes it does not take into the previous state and behavior of the node. The various issues to be addressed by static algorithms are resource utilization, response time, power consumption and scalability. The drawback of this approach load fluctuation while distributed the load impact on the overall performance of the system. The various static algorithms are given as below:

***Round Robin Algorithm:*** Round Robin [15] based algorithm which introduces the concept of load balancing. In cloud environment there are many of the existing algorithms are available to maintain the load balancing. Round robin is one of the algorithm uses to enhance the performance of load distributed to the nodes. In [5] load balancing policies each time a new user request for virtual machine to execute the task they don't save the previous state allocation of the virtual machine. To resolve this problem developing an efficient virtual machine load balancing algorithm using round robin approach. In [9] distributed the request according to the priority of the datacenter for better performance and overcome the problems of the round robin approach. The principle behind Round robin algorithm technique in which time is divided and interval is allotted to each job. Each node completes its job in a given interval of time. In this algorithm job is completed in a circular executive and handling all processes without any priority. It maintains a queue of incoming requests and allocates time scheduling manner to virtual machines in a datacenter. Thus in a specific time quantum each request is allowed to be executed i.e. request is still incomplete, it has to wait for the next round. When request is complete it allow other process to take charge to another request for complete the job.

***Randomized Algorithm:*** The randomized algorithm is static in nature. In [15] introduces the various load balancing algorithm one of them is the randomized algorithm. In this algorithm the user request randomly select the node to perform the task. The node is selected on random selection, without having any information about the present or earlier load on the node. This algorithm gives the best performance for special purpose applications and well suited when the nodes has equal load. Each node in the machine maintains its own load record. The major issue occur sometime on a single node is overloaded while other nodes is under loaded. When one node goes overloaded it effect on the performance of the algorithm.

***Min-Min Load Balancing Algorithm:*** The min-min load balancing algorithm used for those tasks which have minimum completion time. In [20] min-min algorithm load balancing proposed by two strategies. In first strategy the task having minimum execution time scheduled first. Therefore, small task chosen first by min-min to performed the fast execution of the resources. In second strategy the task are rescheduled to use the heavy loaded resources. The mechanism of this algorithm is to find for the minimum completion time for the entire task. The number of small task is more than as compared to the number of larger task. The algorithm search for the minimum execution time task assigning those tasks to the corresponding machine that produces the minimum completion time and achieves better performance. The same procedure is repeated until all tasks are scheduled. The limitation of this algorithm is to produce a high computational power cause by the use of resources to perform the smaller task.

***Max-Min Load Balancing Algorithm:*** In [18] max-min algorithm maintain the task status includes the updated time, execution time and completion time to improve the performance. Therefore, it also maintains the virtual machine status table to create a resource model and task model. The Max-min load balancing algorithm is similar to the min-min algorithm. In earlier algorithm min-min find out the minimum execution time among the entire task therefore, max-min algorithm find out the maximum execution time among the entire task. The task is scheduled according to the maximum time on the corresponding machine expected to perform well.

TABLE 1

SURVEY OF STATIC LOAD BALANCING TECHNIQUES

| Algorithm | Proposed by: research focus/contribution/features | Compared algorithm | Performance | Future work |
|---|---|---|---|---|
| Round Robin | Rakesh Kumar Mishra et al.: Distributed the request according to the priority of the datacenter for better performance. | Not given | Response Time, Resource Utilization and Cost | Not Given |
| Randomized | Soumya Ray et al. : Identify qualitative components for simulation in cloud environment. | Static and dynamic algorithms | Optimal Resource Utilization, Maximize throughput, Minimum Response Time, and Avoid Overload | Not Given |
| Min-min | T. Kokilavani et al. : Load Balanced Min-Min (LBMM) scheduling algorithm used in grid computing proposed by using two strategies. In first strategy the task having minimum execution time scheduled first. In second strategy the task are rescheduled to use the heavy loaded resources. | Min-min algorithm | It reduces the makespan and increases the resource utilization | The proposed algorithm applying on actual grid environment and considering the cost factor. |
| Max-min | Ying Chi Mao et al. : Introduces elastic cloud max-min which maintains the virtual machine status table to create a resource model and task model. | RR and Max-min | It reduces the respond time of task | Not Given |

## III. DYNAMIC LOAD BALANCING ALGORITHM

The dynamic load balancing algorithms based on the heterogeneous environment. Each virtual machine node is able to change the load requirements. These algorithms are able to transfer of task to the remote machine. While transferring the load to the different nodes if one of the node fails it will not halt the system but it little effect on the performance of system. Dynamic algorithms are used for the real time communication which can increase the traffic in the system. It composed of two policies: transfer and location. The transfer policies are those in which a task transfers to one node to another. Location policy use to execute a transferred task assigns a remote node. The various dynamic algorithms are given as below:

***Ant colony optimization:*** The objective of this algorithm [12] is trying to minimize the makespan and maintain the load across all the nodes with give the better results than first come and first serve. The ant colony optimization load balancing technique ability to find the optimal path to find the solution for optimization problems. The ants search for new food that all are work together by using the existing food as a source to come back to the nest. In [16] modify the existing ant colony technique and that has been applied from the perspective of grid network systems or cloud with the aim of load balancing. The task depends on the type of first node that was encountered whether it was under loaded or overloaded.

***Honeybee Foraging Algorithm*:** The honeybee foraging algorithm [21] is task for balance the load across virtual

machines effectively. The foraging behavior of honey bees mapping the task to those virtual machines which are idle or under loaded. In [14] investigates the three distributed proposed solution one of them is the honeybee foraging algorithm. In the experimental work of this paper applied for the m virtual server types with n representative of bees gives the better performance. The main concept behind of honey bee algorithm indicates that it consists of forager. Forger is the type of those bees with job called reapers and bees without job called scout bee. The group or patch of scout bees moves around in search of food and gives the information to the reapers bees. When scout bees come back and they perform the waggle dance regarding direction and distance information, all are move to collect the food. The servers are the group of virtual machines and each virtual machine has a queue process. After processing a request it calculates the profit that the bees show in their waggle dance. This computation of profit on each virtual machine causes additional overhead.

***Biased Random Sampling*:** Biased random sampling algorithm [22] introduces mechanism to improve the selection process of virtual machine by selecting the least loaded virtual machine from the neighbor list and uses cost based load computation. The neighbor is chosen randomly when walk from one node starts. In [23] generating regular resource allocation for distributed load balancing. According to the Biased Random Sampling technique update resource availability and assign new jobs. The load balancing achieved no need to monitor resource availability of the nodes a statistical mechanical model use in this paper for load balancing based on complex networks.

***Active Clustering*:** Active Clustering [15] based algorithm which introduces the concept of load balancing. In cloud environment there are many of the existing algorithms are available to maintain the load balancing. Active clustering is one of the algorithm uses to enhance the performance of load distributed to the nodes. The origin behind the active clustering similar nodes is grouped together called matchmaker node.

***Central Queue*:** Central queue algorithm based on the dynamic load balancing policy this policy [24] proposed a central job dispatcher for distributed system. This policy operation is to motivate of single queue multiserver system. The three aspects studied for this system is average job response time, information exchanges due to overhead and heterogeneous load. The central queue algorithm based on the principle of dynamically distribution of request as a cyclic FIFO queue. Each activity inserted into the queue received and removed by the queue manager. The queue manager receives the request for an activity it removes the first activity from the queue and sends it to the requester. Whenever, queue manager remove the activity at that time a new activity is assigned to it.

***Token Routing*:** In [15] a few existing algorithm can maintain the load balancing which provide better strategies for efficient job scheduling. The token routing is one of them use in order to gain the maximum profit in cloud computing. The main objective of this algorithm is to minimizing the system cost using tokens. There is no communication overhead is generated the token build their own knowledge based which derived from the previous received tokens.

***Genetic Algorithm:*** A new scheduling improved genetic algorithm [2] proposed by merge two scheduling algorithms min-min and max-min. However, job takes the minimum time to complete the user's task and it also able to schedule efficiently a number of jobs on number of machines. The cloud computing system adapted this technique for better scheduling and completes the user's task in minimum time. In [3] presents a genetic algorithm improved for scheduling task in grid environment, which are able to increase the efficiency process met with the feasible results. A proposed strategy [4] used for the simulation a MATLAB toolkit to allocation of incoming requests to the server. This proposed strategy is called as improvised genetic algorithm. An individual node at the time of load balancing considers the cost value as a fitness function. An improved Load balancing aware genetic algorithm [17] used with two methods min-min and min-max to solve the problem of task scheduling in cloud environment. On the other hand the time load balance model helps to find the best solution among the sets of solutions. A genetic algorithm is a type of finding the optimal solution to a problem. The algorithm generates a population of possible solutions to the problem and lets them emerge over multiple generations to find better solutions. The advantage of this genetic algorithm is that it can handle a vast search space applicable to composite objective function and can ward off being trapping into local optimal solution. It reduce the problem of balancing load with simplifying the GA in the

section of cloud computing. The genetic algorithm have some techniques to for better results using techniques inspired by natural evolution, such as crossover, mutation and evaluation.

1. **Crossover:** The main aim is to obtain the better quality chosen the two-point crossover which feed the next generation by selecting individuals from the parental generation and interchanging their genes, new individuals are obtained.

2. **Mutation:** After the crossover, each of the individuals of the chromosomes will be mutated. A mutation is a change that occurs swapping. To generate individual some value of gene can swap with another value of gene.

3. **Evaluation:** Evaluation deals with execution time and cost. Evaluation to create single population from combination of evolutionary effects of these species.

TABLE 2

SURVEY OF DYNAMIC LOAD BALANCING TECHNIQUES

| **Algorithm** | **Proposed by: research focus/contribution/features** | **Compared algorithm** | **Performance metrics/ improvement** | **Future Work** |
|---|---|---|---|---|
| Ant colony optimization | Nishant Kumar, et al. : Modify the existing ant colony technique and that has been applied from the perspective of grid network systems or cloud | Not Given | Detections of overloaded and under loaded nodes | Not Given |
| Honeybee Foraging | Martin Randles, et al. : Investigates the possible solutions proposed for load balancing | Honeybee Foraging, Biased Random Sampling and Active Clustering | Fault tolerance, high availability and scalability | Study the interplay and trade-offs in large-scale complex systems |
| Biased Random Sampling | M. Randles, et al. : It propose that gives a distributed load-balancing scheme by generating almost regular resource allocation networks | Active Clustering | Effective, scalable and reliable load-balancing scheme | Combining various distributed solutions for load-balancing on these large-scale, complex computational structures. |
| Active Clustering | Soumya Ray, et al. : Identify qualitative components for simulation in cloud environment. | Static and dynamic algorithms | Minimum Response Time, and Avoid Overload | Not Given |
| Central Queue | H. C. Lin, et al. : Proposed a central job dispatcher for distributed system. | Not Given | Communication delay is reduced to zero | Work on three aspects average job response time, overhead due to information exchanges and sensitivity to heterogeneous load |
| Token | Soumya Ray, et al. : Identify | Static and dynamic | Minimum | Not Given |

| | | | | |
|---|---|---|---|---|
| Routing | qualitative components for simulation in cloud environment. | algorithms | Response Time, and Avoid Overload | |
| Genetic Algorithm | Zhi-Hui Zhan, et al. : An improved Load balancing aware genetic algorithm used with two methods min-min and min-max to solve the problem of task scheduling. | Load balance aware genetic algorithm (LAGA) and Min-max Max-min genetic algorithm | Optimized makespan and Time Load Balancing (TLB) | Try to model the problem by other scheme for real world application |

## IV. CONCLUSION AND FUTURE WORK

This paper targets on the various existing static and dynamic load balancing algorithm. These algorithms are helps to solve the problems conducted while distributed the load across the virtual machines. In the above discussion the static and dynamic algorithms are capability to performing the accurate distributed system across all the nodes. This analysis of various algorithm helps in designing the new enhanced load balancing technique or algorithm. Future work will be a comparative study on existing algorithms in cloud analyst for developing improved load balancing algorithm and find out the best load balancing algorithm in cloud computing environment.

## REFERENCES


[1] Harmandeep Singh Brar, Vivek Thapar, Kunal Kishor "*A survey of Load Balancing Algorithms in Cloud Computing*" IJCST, Vol.2, Issue 3, May-june 2014.

[2] Chenhong Zhao, Shanshan Zhang, Qingfeng Liu, Jian Xie, Jicheng Hu, "*Independent task scheduling based on Genetic Algorithm in Cloud Computing*", Journal of IEEE, 2009.

[3] Yin, Hao, Huilin Wu, and Jiliu Zhou. "*An improved genetic algorithm with limited iteration for grid scheduling.*" Sixth International Conference on. IEEE, 2007.

[4] Joshi, Garima and S. K. Verma "*Load Balancing Approach in Cloud Computing using Improvised Genetic Algorithm: A Soft Computing Approach*" International Journal of Computer Applications, Volume 122 – No.9, July 2015.

[5] Nusrat Pasha, Amit Agarwal, Ravi Rastogi "*Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment*" IJARCSSE, Vol.4, Issue 5, pp. 2277 128X, May 2014.

[6] Kabir, Md. Shahjahan, Kh. Mohaimenul Kabir and Dr. Rabiul Islam "*Process of Load Balancing In Cloud Computing Using Genetic Algorithm*" Electrical & Computer Engineering: An International Journal (ECIJ) Volume 4, Number 2, June 2015.

[7] Malhotra, Manisha, and Aarti Singh. "*Adaptive Framework for Load Balancing to Improve the Performance of Cloud Environment.*" Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on. IEEE, 2015.

[8] RajKumar Buyya. "*Cloud Analyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments*" pp. 433-659, June 2009.

[9] Rakesh Kumar Mishra, Sandeep Kumar, Sreenu Naik B "*Priority Based Round-Robin Service Broker Algorithm for Cloud Analyst*" Journal of IEEE.

[10] Radi, Mohammed. "*Weighted Round Robin Policy for Service Brokers in a Cloud Environment*" Journal of IEEE.

[11] Yatendra Sahu and R.K. Pateriya. "*Cloud computing overview with load balancing techniques.*" International Journal of Computer Applications (2013).

[12] Kun Li, et al., "*Cloud Task scheduling based on Load Balancing Ant Colony Optimization*" Journal of IEEE (2011).

[13] Tangang, Ranzhi Zhan, Shibo and Xindi "*Comparatively Analysis and Simulation of Load Balancing Scheduling Algorithm based on Cloud Resource*" Journal of Springer (2014).

[14] Martin Randles, David Lamb and A. Taleb-Bendiab "*A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing*" Journal of IEEE (2010).

[15] Soumya Ray and Ajanta De Sarkar "*Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment*" International Journal on Cloud Computing: Services and Architecture (IJCCSA),Vol.2, No.5, October 2012.

[16] Kumar Nishant, et al. "*Load Balancing of Nodes in Cloud Using Ant Colony Optimization*" Journal of IEEE (2012).

[17] Zhi-Hui Zhan, et al. "*Load Balancing Aware Genetic Algorithm for Task Scheduling in cloud Computing*" Journal of Springer (2014).

[18] Ying Chi Mao, Xi Chen and Xiaofang Li "*Max-Min Task Scheduling Algorithm for Load Balancing in Cloud Computing*" Journal of Springer (2014).

[19] Huankai Chen, et al. "*User-priority guided Min-Min scheduling algorithm for load balancing in cloud computing*." Journal of IEEE, 2013.

[20] T. Kokilavani and D.I. George Amalarethinam "*Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing*" International Journal of Computer Applications, Volume 20– No.2, April 2011.

[21] Y. S. Sheeja and Jayalekshmi "*Cost effective load balancing based on honey bee behaviour in cloud environment*" Journal of IEEE, Dec. 2014.

[22] Ariharan V and S. S. Manakattu "*Neighbour Aware Random Sampling (NARS) algorithm for load balancing in Cloud computing*" Journal of IEEE, March 2015.

[23] M. Randles, et al., "*Biased random walks on resource network graphs for load balancing*" Journal of springer, Nov 2009.

[24] H. C. Lin and C. S. Raghavendra "*A dynamic load-balancing policy with a central job dispatcher (LBC)*" Journal of IEEE.

[25] Deepak Kumar Patel, et al., "*Survey of load balancing techniques for Grid*" Journal of science direct.