

A Survey of Soft Computing Techniques based Load Balancing in Cloud Computing

Mayur S Pilavare
Research Scholar
Department of Computer Science
Parul Institute of Technology, Vadodara, Gujarat

Amish Desai
Assistant Professor
Department Of Computer Science
Parul Institute of Technology, Vadodara, Gujarat

ABSTRACT

Cloud computing has become a new style of computing over the internet. There are so much advantages to use cloud over the internet as well as some critical issues are also there those must be resolved in order to improve the performance of the cloud. Cloud computing faces several issues over the internet those are like load management, task scheduling, fault tolerance, several security issues are there. Among them the load balancing is one of the main issue to be resolved in the order to improve the efficiency of the cloud. The load balancing describes that the dynamic workload of the system is distributed among various nodes in such a manner that none of the node is overloaded or underutilized. The load may be CPU load, memory capacity, delay or network load. There are various load balancing algorithms are designed to equally divide the load among the resources and maximize the utilization while minimizing the total task execution time. The new concept was discussed earlier that was using soft computing techniques to perform enhanced load balancing in cloud computing. Various load balancing algorithms are discussed here and comparative study is done on the basis of their performance.

General Terms

Cloud computing, Load balancing algorithms, Virtual Machines.

Keywords

Cloud computing, soft computing, load balancing strategies.

1. INTRODUCTION

Cloud computing has become an emerging phenomenon now a days. Cloud computing has been used world wise as per the users requirements. Different types of users use the cloud computing system as per their work like business use, educational purpose or personal uses to use the software or hardware functionalities that are provided by external third parties. Cloud computing is the delivery of services over the internet. Cloud services include web mails, file storage, online business applications, and social networking sites etc. the cloud computing provides services to user from anywhere that network connection must be available. It provides a shared pool of resources, including networks, computing processing power, specialized corporate and applications and data storage spaces. The cloud computing provides on demand self services that means customers can request and manage their own resources. Cloud computing mainly consists of three service models that mean the mainly three kinds of services are provided to the users those are as below.

A. SAAS(Software-as-a-Service)

The software as a service model consists of pre-made application that contains an operating system, hardware and networks.

B. PAAS(Platform-as-a-service)

In the platform as a service model the customer develops or installs its own software and applications and uses the operating system, hardware and networks.

C. IAAS(Infrastructure-as-a-service)

In infrastructure as a service provides hardware and network only to the customer and it develops its own operating systems, software and applications.

The cloud computing mainly gets deployed on four deployment servers by which the users can access the services provided by it. Four cloud deployment servers are explained as below:

A. Public Cloud

It is offered over the internet and free to access to any one those are like social networking sites, emails, and online storage devices.

B. Private Cloud

This kind of clouds is provided by particular organizations and is accessed by authenticated users only.

C. Community Cloud

This kind of service is shared among selected enterprises and is managed by particular owner or cloud service provider.

D. Hybrid Cloud

This is combination of two clouds and combination of different methods of resource pooling.

2. LOAD BALANCING IN CLOUD COMPUTING ENVIRONMENT

In cloud computing where the client requests are most important to be executed as per requirements and requests arrive at variable time periods. So it is most important to balance the load over the servers. The load is the amount of the computational work that a system performs those are CPU load, network delay load and amount of memory used. Simply load balancing is a strategy or technique that ensures that none of the load in system remains overwhelmed or underutilized or idle while others are executing. Load balancing approach mainly focuses on maximum throughput, optimizing resource usage, avoid overloading and minimize response time. Load Balancing can be static or dynamic.

A. **Static Load Balancing:** In this approach the algorithm must know all the system resources before getting executed and load is distributed equally over all the nodes.

B. **Dynamic Load Balancing:** In this approach the load balancing depends on the current state of the

system. If any of the nodes are over loaded then the load is shifted to another node.

3. EXISTING LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING USING SOFT COMPUTING APPROACH

In recent time various soft computing approaches are used for load balancing in cloud computing for balancing the load over the various resources. Such that each of the resources are completely equally utilized and any of the resource didn't lie idle. Various load balancing techniques those uses the soft computing approach are discussed below.

In 2013, Kousik Dasgupta et al. presented the Genetic Algorithm based Load Balancing[1]. In this paper, A genetic algorithm based load balancing is used in cloud infrastructure to minimize the make span of a given task sets. As the size of cloud rises the service providers have to handle more amount of requests to keep the performance better they use some of load balancing algorithms like Min-Min, Round Robin, Stochastic Hill Climbing (SHC), First Come First Serve (FCFS). But here they have used the genetic algorithm that provides the reduced response time compared to the algorithms listed above. The genetic algorithm uses the natural selection strategy so by that the performance is increased and effective load balancing is done. As the sizes of data sets increases and the number of virtual machine increases the genetic algorithm performs sufficiently better n reduces the response time.

In 2012, Ratan Mishra et al. presented the Ant Colony Optimization technique that uses soft computing approach for load balancing [2]. In this paper they have suggested the ant colony optimization. In this paper as describes the ants are behaviorally unsophisticated insects. Ants are having limited memory and the individual behavior is the main random component of them. Ants are used to execute various complicated tasks with higher reliability and consistency. The ant colony optimization(ACO) concentrates on finding the shortest path depending on spaces the ants lay pheromones trails while travelling from nest to food or vice versa. Same like this various methods like Honeybee Foraging system is also the same the bees indicates to others about the food quality and quantity. In ant colony algorithm each pheromone table represents routing preferences for each destinations and each column represents probability of choosing a neighbor as next hop. This algorithm has drawback that can be improved by making the clusters in the cloud.

In 2013, Zhao Li et al. presented the study of link load based load balancing approach[4] they have discussed about the network link load balancing is an NP- Complete problem so its infeasible to use traditional Genetic Algorithm. Using Genetic Algorithm it's beneficial to find global optimal solution much faster. Genetic Algorithm uses the concept of Darwin's theory of natural evaluation. In link load balancing network load is taken in account by link bandwidth link utilization limitations, the delays and other factors. Here they have enhanced the traditional Genetic Algorithm that used the Roulette Wheel selection to the improves genetic algorithm by using the optimal retention and roulette selection mechanism. To balance the traffic over the network improved genetic algorithm is developed to solve the network link load problem. By improved genetic algorithm they succeed to meet the network bandwidth latency, link utilization, conducting global optimization to find best path to meet the conditions.

In 2013, Broto Mondal et al. presented the novel approach of using the stochastic hill climbing algorithm for load balancing [5]. In this paper the soft computing based load balancing approach has been used. A local optimization approach Stochastic Hill Climbing is used for allocation of incoming jobs to servers or to the virtual machines to get maximum optimization of available resources. A stochastic Hill Climbing and Load optimization algorithm is a loop that continuously moves in one direction of increasing value which is uphill and stops when it reaches a peak where no neighbor has a higher value.

In 2012, O.M.Eldeki et al. suggested an improved max min approach for load balancing[6]. In this paper, the author has suggested the modified max-min algorithm that is based on the expected time and takes it on selection basis of procedure. The max-min algorithm allocated higher task priority to the resources while they execute the lower priority tasks parallel. The total make span is calculated by execution of long task. They have done minimizing the waiting time of short jobs by assigning large tasks to be executed by slower resources and parallel execute small tasks concurrently on fastest resources to finish large number of tasks during finalizing at least one large task on slower resources. The increased opportunity of concurrent resources is increased. This algorithm calculates expected completion time of submitted tasks on each resource. After that the task having overall maximum execution time is assigned to a resource having minimum completion time, so all calculated times are updated and the process is repeated till all tasks are executed.

In 2014, Stuti Dave et al. suggested to use the round robin concept for load balancing at virtual machine level in cloud environment [7]. In this paper, the author specifies an improved round robin algorithm that id Fair Round Robin Algorithm that provides simplest solution to load balancing in cloud computing. It maintains queue of incoming requests and allocates to Virtual Machine in Time Scheduling manner. Each request is allocated a limited time spectrum and if task is completed in that time then it's ok but if task is not completed then it has to wait for next round and request gets completed and allows other processes to Virtual Machine to get allocated. The Priority Based Round Robin (PBDRR) is part of round robin algorithm that is used for calculating the intelligent time and which will allocate different time Quantum and each process gets priority. The author suggested the FairRR algorithm that provides simplicity intact. Complex algorithm increases the executing nodes load and decreases its performance. The FairRR provides maximum scheduling when burst time of incoming request load is having great variance.

Shridhar G. Domanal et al. implemented the modified throttled algorithm to apply load balancing in cloud environment [8]. In this paper they have introduced the updated or we can say modified throttled algorithm for load balancing in cloud computing. This algorithm focuses on assigning the incoming jobs to already available Virtual Machines with using proper intelligence. This algorithm maintains the index table of Virtual Machines as well as the current state of the VM. In modified throttled algorithm the attempt is made to achieve improved response time and effective and efficient usage of existing. On the incoming results request of client available VM is selected depending on availability and requests and id of virtual machine is submitted to data center otherwise -1 is returned. On next incoming request the next indexed VM is selected and depending on the state of VM same procedure is requested.

The sole difference between modified and simple throttled algorithm is in simple index table is parsed from first index each time.

In 2013, Seyed Mohssen Ghafari et al. suggested the BEE-MMT (Artificial bee colony algorithm Minimal Migration Time) technique [9] that is used on the basis of energy consumption or CPU utilization performance degradation due to migrations (PDM) and number of Virtual Machine migrations. This algorithm uses artificial bee colony algorithm (ABC) to fine the overloaded hosts and applies the MMT algorithm to shift one or more Virtual Machines from those overloaded hosts to decrease load of them. At the same time it searches the under loaded nodes and if possible transfer all Virtual Machines allocated to these hosts and transfers them to sleep mode.

Huankai Chen et al. suggested the user priority guides min-min scheduling algorithm to implement the load balancing in cloud [10]. In this paper, the author has introduced the improved load balancing algorithm on basis of Min-Min algorithm to reduce the make span and increase the resource utilization (LBIMM). In traditional Min Min algorithm it does not consider the work load so some resources may always get busy and some nodes lays still. In improved algorithm the user priority is taken into account and called as PA-LBIMM. It divides the task into two groups that are with higher priority and lower priority. It will first run higher priority tasks and then schedule lower priority tasks as scheduled.

4. DISCUSSION AND COMPARATIVE STUDY

The algorithms those are discussed in above section those are previously implemented for load balancing in cloud computing. Some of them are modified or enhanced versions of previously defined or discovered algorithms. Some of algorithms are still lacking of improvements are to be done on them. So here we are discussing the comparative study of them and various environments are defined and compared with each other [11].

5. CONCLUSION

In this paper we have discussed various soft computing techniques those are used for load balancing in cloud computing and are still to be improved and unexplored. In this paper we have also discussed the comparison of various algorithms based on their execution environment. One of the main issue of the cloud computing is load balancing because overloading of the system may lead towards the poor performance of the system. So such more efficient work is still to be done for load balancing in cloud. Our paper focuses on different load balancing algorithms which use soft computing approach in cloud computing environment. Further this can be simulated with cloudsim and the graphical representation and results can be explored.

Table 1: Comparative study of various algorithms

Algorithms	Static Environment	Dynamic Environment	Centralized Environment	Distributed Environment	Hierarchical Environment
Round Robin	Yes	No	Yes	No	No
Ant Colony	No	Yes	No	Yes	No
Genetic Algorithm	No	Yes	Yes	No	No
Min-Min	Yes	No	Yes	No	No
Max-Min	Yes	No	Yes	No	No
Stochastic Hill Climbing	No	Yes	No	No	Yes
OLB	Yes	No	Yes	No	No
Active Clustering	No	Yes	No	Yes	No

6. REFERENCES

- [1] Kousik Dasgupta, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mondal, Santanu Dam, "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing" in Proc. of Elsevier, Procedia Technology 2013
- [2] Ratan Mishra and Anant Jaiswal, "Ant colony Optimization: A Solution of Load balancing in Cloud", in International Journal of Web & Semantic Technology (IJWest), Vol.3, No.2, pp. 33-50, 2012.
- [3] Tushar Desai, Jignesh Prajapati "A Survey Of Various Load Balancing Techniques And Challenges In Cloud Computing" INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 11, NOVEMBER 2013
- [4] Zhao Li, Dong Yu-Min, Haung Chen-Yang "Study of link load balancing based on improved Genetic Algorithm" IEEE 2013 COMPUTER SOCIETY
- [5] Brototi Mandal, Kousil Dasgupta, Paramartha Dutta " Load Balancing In Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach" Science Direct C3IT-2012

- [6] O.M.Elxeiki, M.Z.Reshad, M.A.Elsoud “Improved Max-Min Algorithm in Cloud Computing” *IJCA (0975-8887)* Volume 50- No. 12 July 2012.
- [7] Stuti Dave, Prashant Mehta “Utilizing Round Robin Concept for Load Balancing Algorithm at Virtual Machine Level in Cloud Computing” *IJAC (0975-8887)* Volume 94-No.4, May 2014.
- [8] Shridhar G. Domanal, G. Ram Mohana Reddy “Load Balancing in Cloud Computing Using Modified Throttled Algorithm”.
- [9] Seyed Mohssen Ghafari, Mahdi Fazeli, Ahmad Patooghy, Leila Rikhtechi “BEE-MMT: A Load Balancing Method for Power Consumption Management in Cloud Computing” 978-1-4799-0192-0/13/\$31.00 ©2013 IEEE.
- [10] Huankai Chen, Frank Wang, Dr. Na Helian, Gbola Akanmu “User – Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing”.
- [11] Alok singh, Vikas Kumar Tiwari, Dr.Bhupesh Gour “A Survey On Load Balancing In Cloud Computing Using Soft Computing Techniques” *IJARCCCE* Vol. 3, Issue 9, September 2014
- [12] [www.google.com/ Cloud computing - Wikipedia, the free encyclopedia.html](http://www.google.com/Cloud%20computing%20-%20Wikipedia,%20the%20free%20encyclopedia.html)