



# QOS Differentiation of Various Cloud Computing Load Balancing Techniques

Abhinav Hans  
Gndu Rc Jalandhar

Navdeep Singh  
Gndu Rc Jalandhar

Kapil Kumar  
Gndu Rc Jalandhar

Mohit Birdi  
Gndu Rc Jalandhar

## ABSTRACT

With an increase in the demands the Cloud computing has become one of the ongoing scalable approach to fulfill the cloud based necessities. The biggest advantage of the cloud computing is the ability to overcome the infrastructural challenges those are earlier faced by other technologies. Since the technology is new, therefore the development of the whole structure is not so efficient. It does have a lot of issues on which various scientists and others are working on. Scheduling, load balancing, fault tolerance, etc. are various challenges faced by cloud computing. For this purpose various techniques and algorithms have been proposed. In this paper discussion will be done on the issue of load balancing in cloud computing and different types of load balancing techniques have been reviewed and a comparative analysis have been made among all the existing techniques.

## General Terms

Cloud computing, Load balancing, Algorithms.

## Keywords

Genetic algorithm, Stochastic Hill Climbing, Token Routing, Ant colony optimization algorithm, Central queuing, Generalized Priority Algorithm, SJF, Honey Bee Foraging, Connection mechanism.

## 1. INTRODUCTION

Cloud computing is one the emerging technology, which provides the service of software for lease. Cloud computing allows the user to make demands and provide the services accordingly to the users. People now a day's get the services according to their demands therefore they do not have to pay unnecessary for the things they don't have to use. People now a day's get the services according to their demands therefore they do not have to pay unnecessary for the things they don't have to use. Cloud computing is efficient and scalable, but maintaining the stability of processing so many jobs in the cloud computing environment is a very complex problem with load balancing receiving much attention from researchers [1]. Load balancing is one the challenge faced by cloud computing nowadays and the reason for this challenge is the increase in the users demand. At the infrastructural level, it is impossible to maintain the one or more free service to fulfill the demand at the same time. Since without any load balancing techniques the system will face the traffic and ultimately gets crashed. Application of load balancing and redundant mirrored databases in cluster techniques, across multiple availability zones, reduces the chance of outages that could simultaneously affect the services in cloud systems [2]. In this paper our main aim is to provide an evaluation and comparative study of existing load balancing.

We will discuss the various load balancing comparison parameters and then give a brief introduction to all the load balancing techniques existed. And the end of the paper

comparison among all techniques is done.

## 2. MEASUREMENT PARAMETERS OF LOAD BALANCING TECHNIQUES

An efficient load balancing techniques is the one that avoids the traffic and balances the load equally therefore there must be some parameters to evaluate the load balancing techniques to get better resource distribution for the user demands. In this paper various load balancing measurement parameters have been used to evaluate the load balancing techniques which are discussed below:

### 2.1 Throughput

It is the amount of outputs that a computer can perform on a given inputs in the given amount of time.

### 2.2 Response time

It is the amount of time used to put the user query in concern for execution.

### 2.3 Fault tolerance

It is the ability of the load balancing algorithm that allows to keep working properly in some failure condition of the system.

### 2.4 Scalability

It is the ability to expand itself according to required conditions.

### 2.5 Performance

It is the overall check of the algorithms working. It comprises the completion of the given task against present known standards like accuracy, cost and speed.

### 2.6 Resource utilization

It is used to keep a check on the utilization of various resources.

## 3. LOAD BALANCING ALGORITHMS

Various load balancing techniques are discussed below:

### 3.1 Genetic Algorithm

In [7] author proposed a novel algorithm for load balancing using genetic approach. Author tries to eliminate the challenge of the inappropriate distribution of the execution time, which uses to create the traffic on the server. The Author had used a simulation environment to perform the execution and results of their own proposed algorithm. Also author compared its simulation result with the existing techniques like First come first server (FCFS) and Round Robin (RR).

### 3.2 Stochastic Hill Climbing

In [11] author proposed a soft computing load balancing approach for load balancing. Author used the local



optimization approach stochastic hill climbing for allocation of incoming jobs to the virtual servers or servers. Author analyzed the performance of the algorithm with the help of the cloudAnalyst simulator, which is a cloudSim visual modeler. The performance is analyzed both quantitatively and qualitatively and made a comparison with Round Robin (RR) and First Come First Serve (FCFS). It is one of the forms of the hill climbing algorithm which is used for the load balancing. Just like hill climbing, which chooses the steepest uphill move stochastic hill climbing chooses randomly from the uphill moves with effective probability.

### **3.3 Token Routing**

In [13] author tries to propose an algorithm that downplays the system cost by displacing tokens around the system. Since in the scalable cloud surroundings, less information about the dispersion of workload is there due to the problem of communication bottleneck which leads to uneven dispersion of the workload among the systems. Therefore the heuristic approach of token based load balancing is put into consideration to remove the drawback of the token routing algorithm. This algorithm provides the fast and efficient routing decision. Since this algorithm proves to be helpful as the agent does not need to have an idea of the complete knowledge of their global state and neighbor's working load. So in order to make their decision regarding the passing of the token, they actually build their own knowledge base which is actually derived from the previously received tokens hence leads to no communication overhead.

### **3.4 Ant colony optimization algorithm**

In [4] author had proposed a load balancing algorithm based on trail laying principle. Ants when go out for search of food leave a trail of the path for the other ants so that they do not have to find the path. The ants will just follow the trailed path and will find food. Similarly author used this principle and maintains a pheromone table of the path that the agents use from source to destination and the routing preferences. Path is updated at each new location with the distance value which allows the other mobile agents to choose the shortest path. Author had used this algorithm for load balancing of the demands on the cloud by finding the server's free capacity and treating all the servers as the virtual server. To solve the problem of the load balancing author had chosen the behavior of the ants in search of the food. Since ants had a very intelligent way for finding the food by the method of shortest distance author took that into the consideration. Ants use the principle of trail laying by dropping the pheromones on the ground by stopping on some points on the way through their gesture which is secreted by pheromone gland. That puts a trail for the ants to come back to their colony after they find food.

### **3.5 Central queuing**

In [14] author proposes an algorithm that works on the principle of dynamic distribution. According to this algorithm a queue manager is present which inserts each new coming activity to the queue and keeps a record with it. Whenever there is a request for an activity to the queue manager, it removes the ongoing first activity from the queue and sends it to the one who made the request and if no activity is present in the ready state in the queue then the request is buffered until a new activity is available. Whenever there are certain unanswered requests present in the queue and a new activity comes then the first such request is removed from the queue

and new activity is assigned to it. Local load manager sends a request for the new activity to the central load manager. Whenever a processor load falls under the threshold and the central manager replies to the request if gear up activity is found other than queues the request until new activity gets in.

### **3.6 Generalized Priority Algorithm**

In [12] the author proposed a load balancing algorithm known as generalized priority algorithm. Author had used the concept of the priority system which he uses further to provide the priority to the tasks and balances the load accordingly. The tasks are prioritized according to the sizes of the tasks such that the task with highest size gets the highest priority. Also the virtual servers are prioritized according to their MIPS value such that the virtual server with the highest MIPS value gets highest priority. Hence the load balancing is done accordingly. Also author made a brief comparison with First Come First Serve (FCFS) and Round Robin (RR) scheduling algorithm. Since author had used the simulation environment therefore he tries to compare its algorithm results with the existing algorithms to make a difference on the bases of the average waiting time, response time and execution time.

### **3.7 Shortest Job Scheduling Algorithm**

In [10] the author tries to load the incoming traffic on their server with the help of shortest job first algorithm. In this paper author simulated various number of jobs with different execution time and analyzed the results for the model. Also author made a clear conclusion that the shortest job first algorithm decreases the average waiting time for the jobs. Since the shortest job first had the advantage of choosing the jobs which have least execution time therefore it proves itself to be the powerful load balancing approach. Shortest job first had an advantage of reduced waiting time for the processes which make it a powerful approach.

### **3.8 Honey Bee Foraging**

In [9] author had proposed an algorithm honey bee foraging algorithm. Honey bee with the dancing behavior inspired the author to use this idea for algorithm. When honey bees go in search of food, they perform the special dance called waggle dance on finding the food to tell their remaining members that they had found the food. The type of the dance tells the quality and the quantity of the food they found and also the dance tells the exact distance of food from beehive. Author used this idea and grouped the servers under the virtual server with their own virtual queue. Each server processing the demand from its queue first calculates the profit which is analogous to the quality that bees show in waggle dance. In load balancing this profit or the waggle dance is equal to the amount of time required to fulfill the request or the resources used to fulfill the request.

### **3.9 Connection Mechanism**

In [15] author proposes an algorithm based on the least connection mechanism which itself is a part of dynamic scheduling algorithms. The algorithm runs dynamically i.e. on the run time and tries to count the number of connections for each server dynamically to estimate the load. Author used the concept of load balancer in his proposal which records the connection number of each server. Whenever a new connection is allocated, the total number of existing connections increases which increases the load and decreases when the connection is terminated.

#### 4. CONCLUSION AND FUTURE WORK

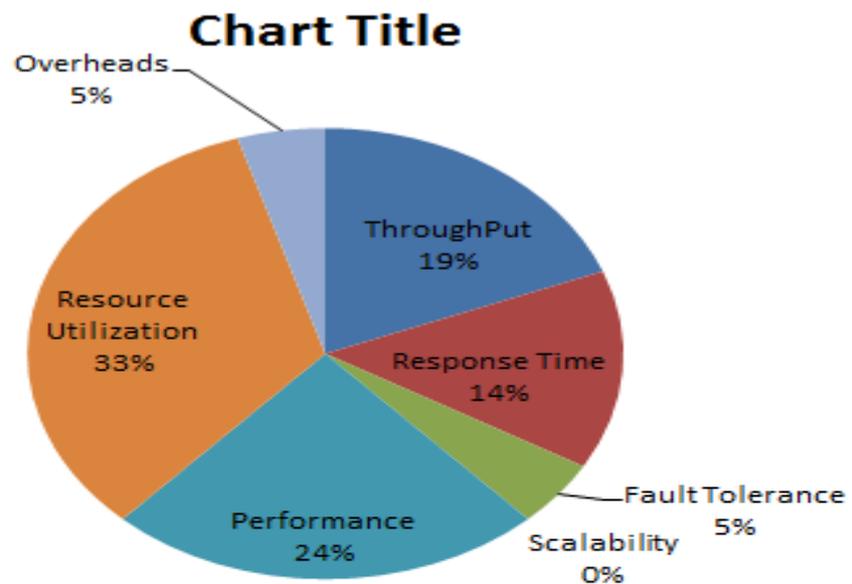
Cloud computing is most usable now a day's therefore the load balancing is becoming a big challenge to overcome. There are various techniques that are proposed by the various researchers to overcome the challenge of load balancing. In this paper a survey has been made and compared various existing load balancing techniques. A comparative analysis is made on different algorithms of load balancing using certain

parameters to evaluate the effectiveness of the algorithms. Since all the techniques covered in this paper are not fully effective as evaluated, therefore for the future purpose one can develop a new load balancing technique with comparison to these all techniques, making it more effective. There are many more parameters which can affect the efficiency of the load balancing techniques which can be included in further study for evaluating the new load balancing techniques and try to cover all the features which are not present in the algorithms covered in this paper.

**Table 1: Comparison Of Existing Load Balancing Techniques Based On Measurement Parameters Discussed In Section 2**

Metrics	Throughput	Response Time	Fault Tolerance	Scalability	Performance	Resource Utilization	Overhead
Genetic algorithm	x	x	x	x	✓	✓	x
Stochastic hill climbing	✓	✓	x	x	✓	✓	x
Token ring	x	✓ ✓	x	x	x	x	✓
Ant colony Optimization	x	x	x	x	✓	✓	x
Central queuing	✓	x	✓	x	✓	x	x
Generalized Priority	✓	x	x	x	x	✓	x
Shortest job Scheduling	x	x	x	x	x	✓	x

<b>Honey bee foraging</b>	x	x	x	x	x	✓	x
<b>Connection mechanism</b>	✓	✓	x	x	✓	✓	x



**ALGORITHMS WITH PARAMETRIC VALUES**

- 1.Genetic Algorithm: [Color indicators]
- 2.Stochastic hill climbing: [Color indicators]
- 3.Token ring: [Color indicators]
- 4.Ant Colony optimization: [Color indicators]
- 5.Central Queing: [Color indicators]
- 6.Generalized Priority: [Color indicators]
- 7.Shortest Job Scheduling: [Color indicators]
- 8.Honey Bee Foraging: [Color indicators]
- 9.Connection MEchanism: [Color indicators]

Figure 1:Graphical Representaion Of Of Existing Load Balancing Techniques Based On Measurement Parameters

**5. REFERENCES**

[1] Gaochao xu, junjie pang, and xiaodong fu” A load balancing model based on cloud partitioning for the public cloud” iee transactions on cloud computing year 2013.

[2] Zenon chaczko,venkatesh mahadevan,shahrazad aslanzadeh and christopher mcdermid” availability and load balancing in cloud computing” 2011 international conference on computer and software modeling.

[3] Soumya ray and ajanta de sarkar” execution analysis of load balancing algorithms in cloud computing environment” international journal on cloud computing: services and architecture.

[4] Ranjan kumar and G sahuo” load balancing using ant colony in cloud computing” international journal of information technology convergence and services (ijits).

[5] Nidhi jain kansal\* and indrveer chana” existing load balancing techniques in cloud computing: a systematic



- re-view” journal of information systems and communication.
- [6] Zenon chaczko, venkatesh mahadevan 2, shahzad aslanzadeh and christopher mcdermid” availability and load balancing in cloud computing” 2011 international conference on computer and software modeling.
- [7] Kousik dasguptaa, brototi mandalb, paramartha duttac, jyotsna kumar mondald,santanu dame “a genetic algorithm (ga) based load balancing strategy for cloud computing” international conference on computational intelligence: modeling techniques and applications (cimta) 2013.
- [8] Yi lua, qiaomin xiea, gabriel klotb, alan gellerb, james r. Larusb, albert greenbergc” join-idle-queue: a novel load balancing algorithm for dynamically scalable web services” iee conference on foundations of computer science.
- [9] Ram prasad padhy, p goutam prasad rao “load balancing in cloud computing systems” thesis from national institute of technology, rourkela-769 008, orissa, india may, 2011.
- [10] Poonam devi mr. Trilok gaba”implementation of cloud computing by using short job scheduling” international journal of advanced research in computer science and software engineering.
- [11] Brototi Mondala, kousik dasguptaa, paramartha duttab”load balancing in cloud computing using stochastic hill climbing-a soft computing approach”science direct.
- [12] Dr. Amit agarwal, saloni jain” efficient optimal algorithm of task scheduling in cloud computing environment” international journal of computer trends and technology (ijctt).
- [13] Zenon Chaczko, Venkatesh Mahadevan, Shahzad Aslanzadeh, Christopher Mcdermid (2011)“Availability and Load Balancing in Cloud Computing” International Conference on Computer and Software Modeling IPCSIT vol.14 IACSIT Press,Singapore 2011
- [14] M.Beltran, A.Guzman and J.L. Bosque(2011), “Dealing withheteroginity in clusters” in proceeding of the Fifth International Symposium on Parallel and Distributed Computing, ISPDC.
- [15] P.Warstein, H.Situ and Z.Huang(2010), “Load balancing in a cluster computer” In proceeding of the seventh International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE.