International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 06, June 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Measuring the performance of load balancing algorithms for efficient Virtual Machine allocation in Cloud Computing

Iliyasu Adamu[1], Kabir I Musa[2], Abdussalam Y. Gital[3]

Department of Management Science, Abubakar TafawaBalewa University (ATBU), Bauchi - Nigeria

iliyasumails@gmail.com,

imkabir@atbu.edu.ng,

asgital@gmail.com

Abstract

Load balancing algorithm is introduced to distribute incoming request to individual device or available servers that makes up a cloud network. The techniques have optimized the response time among various clusters and avoiding one server overloading with a lot instruction while others are left idle. There are various forms of load balancing algorithms that ensure load on various servers are balanced equally, to ensure no single server is over-loaded or under-loaded by end-user's request. However these algorithms perform differently under different parameters. In the same time, there is absent of recent research works that review the performance of both static and dynamic load balancing algorithms such as RR, throttled and ESCE under different parameters for efficient VM allocation. This paper focus on reviewing the performance of various loads balancing algorithm within Virtual Machine policy for efficient VM allocation under different parameters using CloudAnalysis. The performance of throttled algorithms has a minimal response time and datacenter processing time when compare with RR and ESCE under different parameters.

Keywords:Cloud computing, datacenter processing time, load balancing, load balancing algorithms, response time.

I. Introduction

Our societies recently are being reshaped by the advancement in Information Technology (IT) field. The advancement in Cloud Computing technology is one of those fields due to its features and current technological advancement. Rather than storing, managing and processing of an organizational database in-house, you can rent such services and access them remotely at a cheaper rate[1].Cloud computing (CC) is an evolving computing paradigm that enables outsourcing of all organizational IT needs such as data storage, computation and application software's, through large Internet. The shift towards this service-oriented computing paradigm is driven primarily by ease of administration and management process evolving from software upgrade and fixing bugs [2]. It is the service delivery nature of cloud computing that makes it such a disruptive force in every field of IT industry.The cloud computing capabilities are rented on-demand bases and no hardware or software assets

need to purchase outright by the organization ([9], [13]).[7] Opined that cloud computing has revolutionizing the way and manner in which software are built and delivered over the networks.

Load balancing is a vital issue in cloud computing that discuss the skilled and scalable distribution of network workload. The number of users requires and the need for service is increasing day by day. As such load balancing techniques is required to balance the network load on various computer system, servers and network. The network load that need balancing include network, servers, computer system, memory capacity and CPU load or delay [11]. The introduction of load balancing algorithms in cloud computing willprevent any single cloud computing server from breaking down due to high load or overloaded instructions/command. The algorithm is introduced to distribute incoming request to individual device or available server that makes up a cloud network. The techniques have optimized the response time among various clusters and avoiding one server overloading with a lot instruction while other servers are left idle.

There are various forms of load balancing algorithms that ensure load on various servers are balanced equally, to ensure no single server is over-loaded or under-loaded by end-user's request.These algorithms perform differently under different workload [5].Several load balancing algorithms had been proposed by a number of different researchers which focus on different key elements such as cloud service processing time, response time and costs. However these algorithms comparisons neglect performance of traditional load balancing under many criteria to burst network workload. In the same time, research works that compares both static and dynamic load balancing under different parameters are quite few. Motivated by this researchable problem, this to overcome the challenges in selecting some various load balancing algorithm to use and to meet up the good quality of service delivery in cloud computing user. This paper achieve that by comparing the performance of round robin (RR), throttled and equally spread current execution load balancing algorithm under different parameters using cloud simulator (CloudSim) version 3.0.1.

II. Load balancing

Load balancing is simple the process of network load distribution among various computer system, network or servers. This is achieved by identification and utilization of both over loaded and under-loaded servers [13]. [1]Define load balancing as the splitting of network workloads and resources to numerous servers so that, no single network server is overloaded or under-loaded. It`s achieved by splitting resources among numerous computer system, network and servers.

The load balancing in cloud computing is one of the recent most useful and yet challenging research area for distributing workload among various virtual machine (VM) at the data centers [5]. The load balancing algorithm needs to perform a heterogeneous type of cloud service resources on cloud data center along with its current state while allocating individual user task to various cloud service resources [9].According to [12] load balancer is a device that performs the application or network traffic distribution across bunch of network servers and clusters. This is achieved by enhancing network load balancing responsiveness and raises network appearances. A load balancer lies between the network servers and cloud consumer client/customer. According to [4] load balancing is most critical aspects of network task scheduling in cloud computing by attributing work

load between various servers. Such that network resource usage and utilization can be efficiently maximized and the best network throughput can be achieved.

III. The need for Load Balancing

Load balancing is required to migrates the network load or task between various datacenters on a network to increase the data processing speed for the network data centers [10].Load balancing in cloud computing has been an essential part needed for process distribution of workload among various servers known as clusters over the years. It's needed to prevents any single cloud computing server from breaking down due to high load or overloaded instructions/command. Load balancing algorithm is required to distribute incoming request to individual device or available server that makes up a cloud network. The techniques have optimized the response time among various clusters and avoiding one server overloading with a lot instruction while other servers are left idle.

IV. Classification of Load Balancing

According to [6] and [8] load balancing algorithms are classify in two main types namely; static and dynamic load balancing algorithms with some other algorithms within them, as shows in figure 1.

1. **The Static load balancing algorithms: -** as the name implies, load is statistically distributed among various network servers.These algorithms allocate network workload statistically among various processor before execution and it requires a prior knowledge state of the entire system [16].
2. **The dynamic load balancing algorithms: -** as the name implies, load is dynamically distributed among various network servers. In these types of algorithms, network workload is distributed dynamically among processors during program execution of algorithms. These techniques are complex when compared to static but it has more fault tolerant and overall better network performance [16].

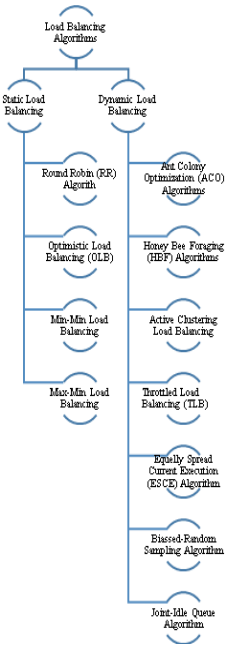


Figure 1: Classification of load balancing algorithms.

Source: [6] and [8]

V. Types of Load Balancing Algorithms

For the purpose of this research work, analysis are carryout on the three (3) widely used existing load balancing algorithms, that falls under both static and dynamic load balancing algorithms types. These algorithms are as follows:

1. **Round Robin (RR) load balancing algorithms: -**it's a type of static algorithms that randomly select load balancing across network devices. As the name implies, its uses a round robin fashion in selecting and allocating workload load on network device. RR is one of the most efficient and most popular static loads balancing algorithms type. Its essential role is played by network data centers in handling load balancing process within cloud computing service. Its uses time division known as time slice for time sharing of load balancing. In RR when data centers controllers received service request from the user, it passes it to the algorithms for load balancing [14].
2. **Equally Spread Current Execution (ESCE) load balancing algorithms: -**it's a type of dynamic algorithms that spread task equally across the network devices. As the name implies, ESCE spread current execution or task equally across the available network device by distributing equal load between every server in the data center. The load process are assigned priority at the initiating state of the algorithms to check the network device capacity and sizes for transferring task to the server that can handle it efficiently ant faster. These algorithms allocate workload based on network device capacity and size of the matching virtual machine on the network [14].
3. **Throttled Load Balancing (TLB) algorithms: -**it's a type of dynamic load balancing algorithm that is based on the principles of finding a suitable virtual machine using record table that keep track of all virtual machine on the network for load balancing. Its maintain records that keep track of all the virtual machine (VM) in the form of a periodic table and their status (either busy or idle) to enable fast and easy network load balancing. Whenever a client request is initiated, the network load balancer looks into the throttled table for choosing the virtual machine suitable for the type of initiated request and allocates the task to the available best suited virtual machine on the network. After allocating the stated task to the selected virtual machine, the table details are updated with the latest virtual machine status. In case of absents of available virtual machine to handle users request, the request is kept on queue status, till when virtual machine is available to handle the task[8].

Performance metrics for efficient VM allocation in Load Balancing

Load balancing among different cloud computing resources is a key issue that needs to make sure there is an optimal usage of network resources by allocating various task to available network devices. There are a number of predefined metrics measures used to compare various load balancing algorithms and check their efficiency as shows in table 1. These metrics determined the performance of load balancing algorithms under different parameter such as overall response time and make span tasks. These metrics must be chosen properly to meet up the demand for load balancing for network versatility, low network overhead, under/over loading one

server over the other and proficient network resource allocation[8], [15]. For the purpose of this research work, seven (7) load balancing performance metrics is considered as shows below:

i. **Throughput**: - it's a type of load balancing metrics used to calculate the number of completed task in a given period of time. It determines the number of successfully completed task per unit of time.

ii. **Response time**: - it's a type of load balancing metrics used to calculate the overall time taken to react to user request. the lower the response time the better the network performance.

iii. **Network overhead**: - it's a type of load balancing metrics used to calculate total network device overhead. For effective network performance, the overall network overhead needs to be minimal.

iv. **Network scalability**: - it's a type of load balancing metrics used to calculate the scaling capability of network device. For effective performance network scalability need to be improve for efficiency.

v. **Network performance**: - it's a type of load balancing metrics used to calculate the overall network capability and performance over a given period of time. It's used to determine the system efficiency during load balancing. It can be improved and enhanced by reducing response time by increasing network throughput.

vi. **Resource utilization**: - it's a type of load balancing metrics used to calculate the overall usage of network device. For effective network performance resource utilization most be optimal by load balancing algorithms.

vii. **Fault tolerance**: - its types of load balancing metrics used to compute and handle network device tolerance when fault arises. For effective load balancing algorithm,a network device must handle fault tolerant.

VI. Comparative analysis of the performance metrics for VMAllocation in load balancing

Table 1:Comparative analysis of the performance metrics for VM allocation in load balancing

Load balancing algorithms		Performance Metrics						
		Network Throughput	**Response Time**	**Cost Overhead**	**Network Scalability**	**Network Performance**	**Resource Utilization**	**Fault Tolerance**
Static	**RR**	High	High	High	High	High	High	High
	OLB	High	High	High	Medium	Medium	High	Low
	Min-Min	Medium	High	High	Low	Medium	High	Low
	Max-Min	Medium	High	High	Medium	High	High	Medium
Dynamic	**ACO**	High	High	High	Low	Low	High	Low
	HBF	High	High	Low	Low	Medium	Low	Low
	Active Clustering	Medium	High	High	Low	Low	High	Medium
	Throttled	Medium	High	Low	High	High	High	High
	ESCE	Low	Medium	Low	Medium	Medium	High	Low
	Biased-Random	Low	Low	High	Low	High	High	Medium

Joint-Idle	Medium	High	High	High	High	Low	Low

Source: [8], [13]

Table 1 shows the comparative analysis of the performance evaluation of some selected load balancing algorithms on different parameters for efficient VM allocation. The performance evaluation of each algorithm is classified under different parameters such as high, medium and low to represent the impact of the metrics on each algorithm. In static load balancing algorithms, Round Robin (RR) performs better than the other static algorithms. While in dynamic load balancing algorithm, throttled (TLB) performs better than the other dynamic algorithms.

VII. Testing parameters

A set of testing parameters are considered when simulating the task scheduling algorithms. These parameters play an important role in testing the performance of load balancing algorithm under different number of virtual machine and data centers. These parameters include the following;

1. Average response time (t): - as the name impliesitsuse to calculate the overall total average time taking to received cloud service response. It calculate the overall time interval taking between sending of cloud service request and time taking to receive the corresponding feedback/response[3]. ART is denote usingthe following formula:

$$\text{Average Response Time (ART)} = \frac{\sum_{j=1}^{n} RT}{n} \quad (1)$$

Where:

$$Response\ Time\ (RT) = Finish\ Time - Arival\ Time + TDelay \quad (2)$$

$$Transmission\ Delay\ (TDelay) = Total\ Latency + Total\ Transfer \quad (3)$$

2. Average processing time (t): - as the name implies, its use to calculate the overall total average time taking to process cloud service. It calculates the overall time needed/taken by the datacenter to process cloud computing service task/request[9]. APT is denoted using the following formula:

$$\text{Average Processing Time (APT)} = \frac{\sum_{j=1}^{n} VM\ Load}{\sum_{j=1}^{n} VM\ Capacity} \quad (4)$$

Where:

$$VM\ Load = The\ sumation\ of\ all\ datacentre\ virtual\ machine\ load$$

VM Capacity = The sumation of all datacentre virtual machine capacity

VIII. Simulation Setup

For the simulation set-up, social media such as WhatsApp, Facebook, Twitter and Instagram etc. users are considered to be the cloud computing service users nationwide. The users worldwide comprise of seven (6) continents namely; Africa, Antarctica, Asia, Australia/Asian, Europe, North America and South America. Six (6) datacenters (DC1 to 6) was created one (1) for each region (UB1 to 6) to attain to cloud services as shows in Figure 2. Figure 3 and 4 shows the simulation results obtain.Table 2 shows the user base characteristics for the simulation which assume a similar system in a normalized scale of (1/200)th. Six (6) UserBase (UB) or regime was created to represents users from the six geo-political regions worldwide. We assumed 10% of cloud computing users at the off-peak hours and the peak hours are assumed to be in the evening time after office work or school activities with different interval rate. Table 3 shows the simulation parameters configured in the simulation setup and configuration.

Table 2: User base Configuration

Region	Region Id	Users (in millions approx.)	Time Zone (GMT)
North America	UB – 0	90	GMT – 6
South America	UB – 1	90	GMT – 4
Europe	UB – 2	220	GMT + 00
Asian	UB – 3	240	GMT + 8
Africa	UB – 4	340	GMT + 1
Australia	UB – 5	20	GMT + 8

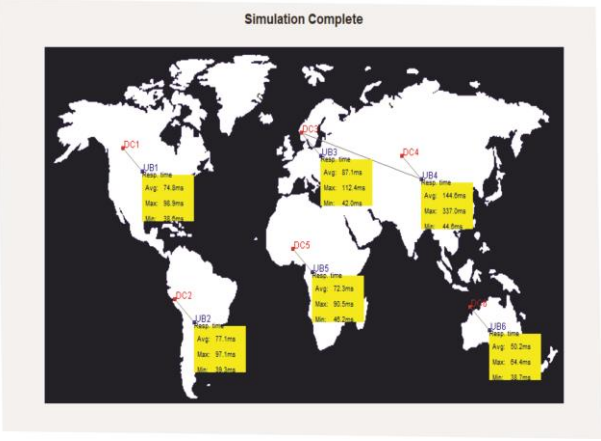


Figure 2: Data centers (DC) and region (UB) simulation result

IX. Performance evaluation, experimentationand Results

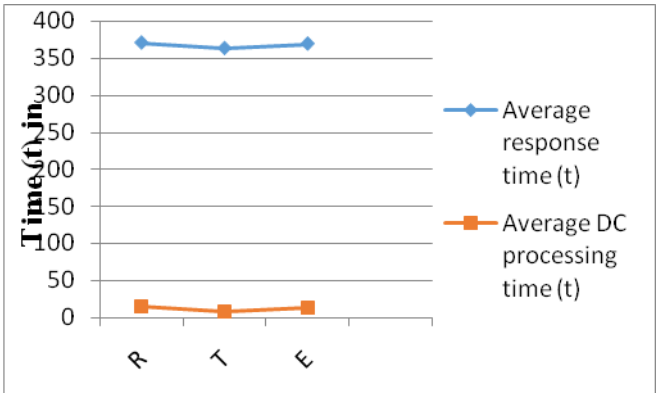
To test the performance of the three (3) selected load balancing algorithms for efficient virtual machine (VM) allocation. Three (3) scenarios were created under different parameters. Table 3 shows the performance parameters configured in CloudAnalysis application for the whole simulation process.

Table 3: Experimental parameters

Parameters	Values
Simulation Engine	CloudSim v3.0.1
Operating System	Linux
Front End Application	NetBeans IDE 8.2
Number of Data Centers	6
Number of customers (VM)	1000000
Types of work load	Random
Memory size	512mb
Bandwidth	1000bytes
VM scheduling policy	Time shared
Service broker policy	Optimize response (t)
Request per-user hour	60Minutes
Cloud service request size	100bytes
Cost per VM per hour	$0.1 dollar
Cost of memory	$0.05 dollar
Cost of data storage	$0.1 dollar
Cost of data transfer	$0.1 dollar/Gb

Scenario 1: Single Data center with 50 virtual machine (VM)

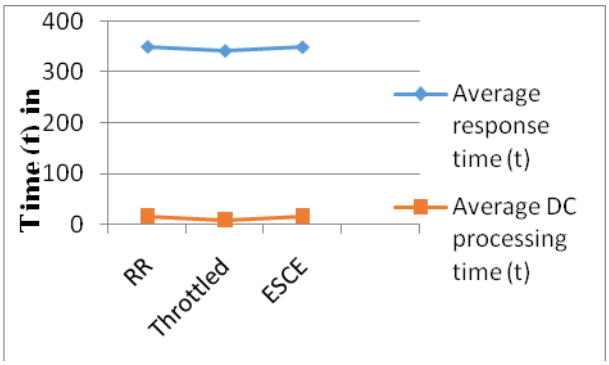
In this first scenario, a single datacenter is configured to accommodate fifty (50) different virtual machines at the same time. Figure 3 shows the simulation results, where Round robin (RR) and ESCE load balancing algorithms take longer time to process cloud service request and also take higher time to process cloud service when compared to throttled load balancing algorithms. RR load balancing algorithms take higher time to process and response to cloud service request due to it static nature of round robin load balancing techniques followed by some parameters such as RAM, Processor speed and network load.



Scenario 2: Two (2) Data centers with 50 users

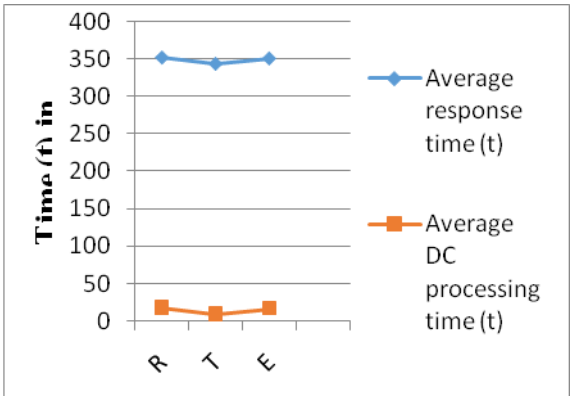
In this scenario, two data centers are configured to balance the work load between the datacenters. Fifty (50) different virtual machines are configured, twenty (25) on each datacenter to address cloud computing service request. Figure 4 shows the simulation results, where Round robin (RR) and ESCE load balancing algorithms take longer time to process cloud service request and also take higher time to process cloud service

when compared to throttled load balancing algorithms. But when compared to figure 3 with a single datacenter, both the RR and ESCE load balancing algorithms datacenter processing time and response time has reduce with the addition of a data center to balance the workload.



Scenario 3: Two (2) Data centers with 100 users

In this scenario, two data centers are configured to balance the work load between the datacenters. One hundred (100) different virtual machines are configured, fifty (50) on each datacenter to address cloud computing service request. Figure 4 shows the simulation results, where throttled load balancing algorithms still maintain almost the same response time with a little change in processing time after increasing virtual machine from 50 VM to 100 VM.



X. Discussion, Conclusion and future work

Load balancing is a vital issue in cloud computing that discuss the skilled and scalable distribution of network workload. The number of users requires and the need for service is increasing day by day. As such load balancing techniques is required to balance the network load on various computer system, servers and network. Load balancing hasprevented any single cloud computing server from breaking down due to high load or overloaded instructions/command. Different load balancing algorithm was introduce to optimized the response time and datacenter processing time among various clusters and avoiding one server overloading with a lot instruction while other servers are left idle.

In measuring the performance of load balancing algorithms for efficient VM allocation in cloud computing, different load balancing algorithm are compared in table 1 under different parameters. the three most popular

load balancing algorithm are simulated and compared to check the average response time and datacenter processing time under different workload. The analysis of the result shows that throttled load balancing algorithms performance is relative good under different workload as compared to round robin and ESCE. In the future, is to assess the suitability of each algorithm in relation to cloud computing quality of service and furthermore new algorithms will be proposed to compare with the traditional load balancing algorithms under different parameters and workloads.

REFERENCES

[1] Bhushan, Bharat and Mohit Kumar. "A methodological comparisons of the most efficient load balancing algorithms in Cloud Computing." International Conference on Innovative Computing and Communication (ICICC 2020) (2020): 1-6.

[2] Buyya, R, J Broberg and A Goscinski. Cloud Computing: Principles and paradigms. New Jersey: John Wiley & Sons Inc, 2011.

[3] Dhari, A and K I Arif. "An efficient load balancing scheme for cloud computing." Indian Journal of Science and Technology 10.11 (2017): 1-8.

[4] Ebadifard, F, S M Babamir and S Barani. "A dynamic task scheduling algorithm improved by load balancing in cloud computing." 6th International Conference on Web Research (JCWR 2020) (2020).

[5] Jena, U K, P K Das and M R Kabat. "Hybridization of meta-heuristic algorithm for load balancing in computing environment." Journal of King Saud University - Computer and Information Sciences (2020): 1-11.

[6] Kaur, Mandeep and Rajni Aron. "A systematic study of load balancing approaches in the fog computing environment." The Journal of Super Computing (2021).

[7] Kavis, M J. Architecting the cloud design decision for cloud computing service model: SaaS, PaaS and IaaS. New Jersey: John Willey & Sons Inc, 2014.

[8] Kumar, M and B Bhushan. "A methodological comparison of the most efficient load ballancing algorithm in cloud computing." International Conference on Innovative Computing and Communication (ICICC 2020) (2020): 1-6.

[9] Mulla, Badshaha P, C R Krishna and R K Tikoo. "Load ballancing algorithm for efficient VM allocation in heterogeneous cloud." International Journal of Computer Networks & Communications (IJCNC) 12.1 (2020): 83-96.

[10] Narayanan, S. "A study of load balancing algorithms in Software defined Network." International Journal of Trend in Research and Development 7.2 (2020): 8-9.

[11] Roy, Saumendu, et al. "Measuring the performance on Load ballancing algorithms." Global Journal of Computer Science and technology: B 19.1 (2019): 40-49.

[12] Saumendu, Roy, et al. "Measuring the performance on load balancing algorithms." Global Journal of Computer Science and Technology: B Cloud and Distributed 19.1 (2019).

[13] Siddiqui, S, M Darbari and D Yagyasen. "Evaluating load balancing algorithms for performance optimization in cloud computing." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8.12 (2019): 2217-2221.

[14] Tadapaneni, Narendra Rao. "A survey of various load balancing algorithms in cloud computing." (2020). SSRN Electronic Journal: Available at https://ssrn.com/abstract=3563342

[16] Tariq, Maria and Khadija Javed. "Comparison of various load balancing techniques in a centralized and distributed cloud computing environment." LGU Research Journal for Computer Sciences & Information Technology 4.1 (2020): 37-41.

[17] Verma, V K, et al. "Hybrid approach for optimized resource allocation for load balancing in cloud infrastructure." International Journal of Advanced Science and Technology 28.12 (2019): 224-235.

[18] Williams, M. A quick start guide to cloud computing: Moving your business into the cloud. New Delhi: London Philadelphia, 2010.