Cloudlet Networks Performance Analysis and Improvement

Jameela Abdulla Hassan
Computer Sciences and Engineering
Umm AL-Qura University
Makkah, Saudia Arabia
Jameelaash@hotmail.com

Fahad Al-Dosari
The Dean, Faculty of Computer and Information Systems
Umm AL-Qura University
Makkah, Saudia Arabia
fmdosari@uqu.edu.sa

***Abstract*— Cloud computing is a Participation in the process and storage operations across distant servers that are shared by many organizations and users and thus be transferred from an application to a service. The organization can share data over the Internet and user can pay only for the resources that will be used only. While cloud computing has disadvantages, there are some advantages for cloudlets have over cloud computing which include: lower network latency and users having full ownership of the data shared. When the need of data to be stored in the servers grows quickly, the workload in every resource will grow too. So, we need a load balancing algorithm and the load balancing is important issue in the cloud environment. Load balancing defined as a technique that divides the extra load equally across all the resources to ensure that no one resource overloaded. . So the performance of the cloud can be improved by having an excellent load balancing strategy. For that we will discuss the existing load balancing algorithms in cloud computing and propose algorithm to improve round robin algorithm by CloudAnalyst simulator based on a factor of response time and processing time and the proposed algorithm was found to be best in response time and processing time when we compare it with round robin algorithms.**

***Index Terms*— Cloud Computing, CloudAnalyst, Load Balance, Mobile Cloud Computing, Cloudlet Networks.**

I. INTRODUCTION

Cloud computing is a participation -based service where you can use storage space and computer resources. One example of cloud computing is Drop box. It is a Web application service operates in a manner of cloud computing to store files on the user, and can use the service to share files between more than one user on the Internet and synchronize files between more than one computer or mobile phone. A Drop box program on the computer which appears in the form of a folder can be placed on the desktop, and treats like any other folder . In fact, it is in the server of Drop box but all we have is its image.

Another example is email (e.g. gmail , hotmail ,yahoo,etc.). When you want to use your email, go to the email service provider using your web browser, and log in. The fact of the matter is that your email is not in your computer; you access it over the Internet connection from anywhere. In other words, your email is not installed on your computer, but the mail service provider provides email servers to send, receive, accept, and store email for other organizations and/or end users [1].

II. DEFINITION OF CLOUD COMPUTING

The NIST defines cloud computing as "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".

A computing cloud is "a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing infrastructures on demand, which could be accessed in a simple and pervasive way " [4].

A cloud is a collection of IT resources, including hardware and software resources that a user accesses over a network. A cloud infrastructure is built, operated, and managed by a cloud service provider. Cloud computing is "a model that enables consumers to conveniently hire IT assets as a service from a providers cloud infrastructure". The cloud model is similar to a utility service such as electricity, wherein a consumer simply plugs in an electrical appliance to a socket and turns it on. The consumer is typically unaware of how the electricity is generated or distributed and only pays for the amount of electricity used. Similarly, to the cloud consumers, the cloud is an abstraction of IT infrastructure from which they hire IT resources as services without the risks and costs associated with owning the resources. Consumers pay only for the services that they use, either based on a subscription or based on resource consumption.

III. PROBLEM STATEMENT

The main purpose of the cloud computing is that its client can utilize the resources to have economic benefits. A resource allocation management process is required to avoid over utilization of the resources which may affect the performance of the cloud.

Round Robin algorithm has disadvantages:

1. At any point of time some nodes may be heavily loaded and others remain idle.
2. Longer average waiting time [2] [3].

IV. GOALS

The main goals of this study are:

- Propose a hybrid algorithm focused on the advantage of ESCE algorithm (less response time) to cover the

IRJECE

disadvantage of the Round Robin algorithm(long response time).

- Design it as at the start it works as RR after that when it has long response time it enter to concept of ESCE to reduce response time.
- Simulate it using CloudAnalyst Simulator.
- The results of proposed algorithm based on response/processing time will compare with RR .

V. MOBILE CLOUD COMPUTING

The mobile cloud computing is a development of mobile computing, and an extension to cloud computing . In mobile cloud computing, the previous mobile device based intensive computing, data storage and mass information processing have been transferred to cloud and thus the requirements of mobile devices in computing capability and resources have been reduced. Therefore, from both aspects of mobile computing and cloud computing, the mobile cloud computing is a combination of the two technologies. In mobile cloud computing networks resources are virtualized and assigned in a group of multiple distributed computers rather than in traditional local computers or servers [5].

As shown in Fig. 1, mobile cloud computing can be divided into cloud computing and mobile computing. Those mobile devices can be laptops, PDA, smart phones, and soon. Which connect with a hotspot or base station by 3G, WIFI, or GPRS? Mobile users send service requests to the cloud through a web browser or desktop application. Then the management component of cloud allocates resources to the request to establish connection, while the monitoring and calculating functions of mobile cloud computing will be implemented to ensure the QoS until the connection is completed [5]. The main aim of mobile cloud computing is to provide a suitable and fast technique for users to access and receive data from the cloud.

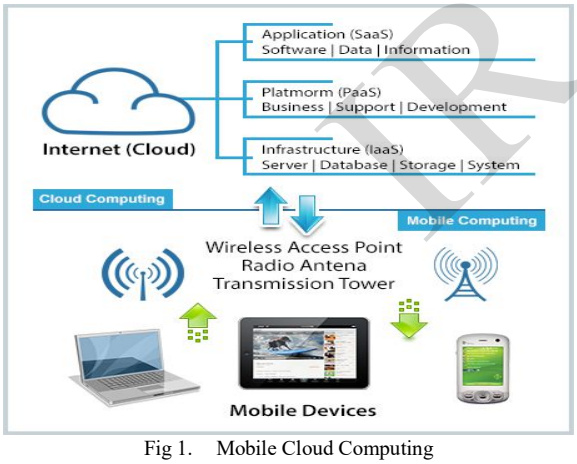


Fig 1. Mobile Cloud Computing

VI. CLOUDLET NETWORKS

When mobile cloud computing brings new types of services mobile users to take full advantages of cloud computing , there are also some disadvantages of mobile cloud computing. To overcome disadvantage cloudlet comes in the existence.

Cloudlet is a new architectural element that arises from the convergence of mobile computing and cloud computing. It represents the middle tier of a 3-tier hierarchy:

i. Mobile device
ii. Cloudlet
iii. Cloud.

Cloudlet can be viewed as a "data center in a box" whose goal is to "bring the cloud closer".

There are a handful of advantages cloudlets have over cloud Technology which include: lower network latency and users having full ownership of the data shared. Figure 2 shows the Cloudlet Architecture [6].

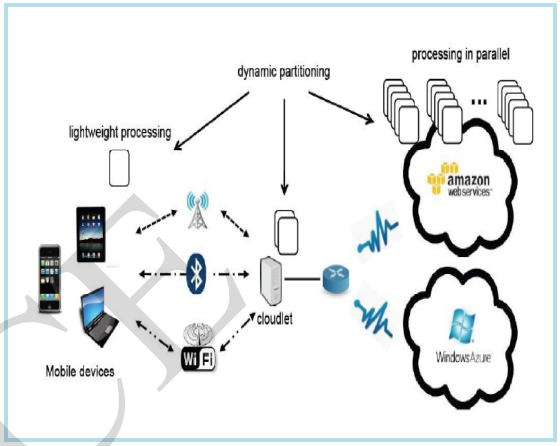


Fig 2. Cloudlet Architecture

There are two types of communication possibilities in cloudlet, the cloudlet-WiFi-based communication and the cloud-based communication. The Transmission Control Protocol (TCP) is used to transfer the data between the mobile device and the cloudlet server.

As shown in figure 1, cloudlet architecture used the both wired and wireless transmission methods to transmit the data from the user to cloud server and vice versa. For data transfer between mobile devices and cloudlet, the architecture uses the wireless network methods. These methods can be Wi-Fi network, Bluetooth or wireless sensor network. Different networks are used in cloud computing to make communication between the cloudlet and the cloud server. Different kinds of network protocols are also used by the cloudlet for the network communication.

Cloudlet follows Centralized Routing and Signaling Approach network. Each cloudlet sends its ID and the ID of its reachable (neighboring) cloudlets to the central server. The central server periodically computes the routing table for each cloudlet and installs the forwarding tables into the cloudlets. Once the routing tables are computed by the server, consequent routing table computation can be triggered by new changes in the cloudlet network. After mobile users register to a cloudlet, the cloudlet periodically sends the IDs (names) of its mobile users along with its ID to the centralized server.

IRJECE

This is because some nodes may move out of the coverage of the cloudlet and others may join in. The central server then keeps a big table of mobile node names (IDs) along with the ID of their respective cloudlets. If the table size grows big, a hierarchical approach where some servers are responsible for some cloudlets can also be used. When a node wants to communicate with another node, it sends the name of the node it wants to communicate with its cloudlet. Its cloudlet asks the central server (network of servers) to look up the name of the requested node. The central server responds with the cloudlet of the requested node. In this case the central server serves as a proxy server used in Session Initiation Protocol (SIP). Each cloudlet can also cache the list of nodes its users want to communicate along with the cloudlet ID [6].

VII. LOAD BALANCING ALGORITHMS

Load Balancing in clouds is a mechanism that distributes the excess dynamic local workload evenly across all the nodes to make sure that no single node is overwhelmed, hence improving the overall performance of the system [11]. Load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc [7].

The three existing algorithm to distribute the workload across multiple nodes over the network link to achieve optimal resource utilization, minimum data processing time, minimum average response time, and to avoid overload are:

A. Round Robin Algorithm

It is the simplest algorithm that uses the concept of time quantum or slices. Here, time is divided into multiple slices and each node is given a particular time quantum and within this time quantum the node will perform its operations.

Though the algorithm is very simple , there is an additional load on the scheduler to decide the size of quantum and it has longer average waiting time, higher context switches, higher turnaround time and low throughput[8][12].

1. Round Robin VM load Balancer maintains an index of VMs and state of the VMs (busy/available). At start all VM's have zero allocation.
2. a. The data center controller receives the user requests/cloudlets.
   b. It stores the arrival time & burst time of the user requests.
   c. The requests are allocated to VMs on the basis of their states known from the VM queue.
   d. The round robin VM load balancer will allocate the time quantum for user request execution.
3. a. The round robin VM load balancer will calculate
   the turn- around time of each process.
   b. It also calculates the response time and average waiting time of user requests.
   c. It decides the scheduling order.
4. After the execution of cloudlets, the VMs are de-allocated by the RoundRobinVmLoadBalancer.
5. The datacentercontroller checks for new /pending/waiting requests in queue.
6. Continue from step-2.

B. *Equally Spread Current Execution Algorithm*

In this technique, load balancer makes effort to preserve equal load to all the virtual machines connected with the data centre. Load balancer maintains an index table of Virtual machines as well as number of requests currently assigned to the Virtual Machine (VM). If the request comes from the data centre to allocate the new VM, it scans the index table for least loaded VM. In case there are more than one VM is found than first identified VM is selected for handling the request of the client/node, the load balancer also returns the VM id to the data centre controller. The data centre communicates the request to the VM identified by that id. The data centre revises the index table by increasing the allocation count of identified VM.

When VM completes the assigned task, a request is communicated to data centre which is further notified by the load balancer. The load balancer again revises the index table by decreasing the allocation count for identified VM by one but there is an additional computation overhead to scan the queue again and again[8]. But it is not fault tolerant and has the problem of single point of failure [9].

The balancer tries to improve the response time and processing time of a job by selecting it whenever there is a match. But it is not fault tolerant and has the problem of single point of failure [9].

C. Throttled Load Balancing Algorithm

In this algorithm the load balancer maintains an index table of virtual machines as well as their states (Available or usy) [11]. The client/server first makes a request to data centre to find a suitable virtual machine (VM) to perform the recommended job. The data centre queries the load balancer for allocation of the VM. The load balancer scans the index table from top until the first available VM is found or the index

IRJECE

table is scanned fully. If the VM is found, the data centre communicates the request to the VM identified by the id.

Further, the data centre acknowledges the load balancer of the new allocation and the data centre revises the index table accordingly. While processing the request of the client, if appropriate VM is not found, the load balancer returns -1 to the data centre. The data centre queues the request with it. When the VM completes the allocated task, a request is acknowledged to data centre, which is further apprised to load balancer to de-allocate the same VM whose id is already communicated [8].

The purpose of algorithm is to find the expected Response Time of each Virtual Machine because Virtual Machines are of heterogeneous capacity with regard to its processing performance, the expected response time can be found with the help of the following formulas [7]:

$$\text{Response Time} = \text{Fint} - \text{Arrt} + \text{TDelay} \quad (1)$$

Where, Arrt is the arrival time of user request and Fint is the finish time of user request and the transmission delay can be determined by using the following formulas:

$$\text{TDelay} = \text{Tlatency} + \text{Ttransfer} \quad (2)$$

Where, TDelay is the transmission delay, Tlatency is the Network latency and Ttransfer is the time taken to transfer the size of data of a single request (D) from source location to destination location.

$$\text{Ttranfer} = \text{D/Bwperuser} \quad (3)$$
$$\text{Bwperuser} = \text{Bwtotal/Nr} \quad (4)$$

Table 1 shows comparison of these Load Balancing Algorithms [10].

Table I: Comparison of Load Balancing Algorithms

Algorithm	Description	Pros	Cons
Round Robin	First request is allocated to a randomly picked VM. Subsequent requests are assigned in circular order.	Equal distribution of work load.	Job processing time is not Considered.
Equal Spread Current Execution	Request is assigned to any available VM that can handle it. If there is an overloaded VM then the balancer distributes some of the tasks to some idle VM to balance the load.	Response time and Processing time of a job is improved.	Not fault tolerant because of single point of failure.
Throttled Load Balancing Algorithm	Record of the state of each VM (busy/idle) is maintained. Request is accepted if a match is found in the table otherwise -1 is returned and the request is queued.	TLB tries to distribute the load evenly among the VMs.	Does not consider the current load on VM.

VIII. EXPERIMENTATIONS AND RESULTS

In this study, a proposed algorithm having concepts from Round Robin algorithm and ESCE algorithm has been proposed for improving long response time in Round Robin algorithm.

Round robin has longer average waiting time, so the response time becomes longer. But in ESCE algorithm the balancer tries to improve the response time and processing time of a job by selecting it whenever there is a match. The proposed algorithm was efficient in case of same data size per request and different requests per user per hour. In this proposed hybrid algorithm, the concept of circular way to allocate VMs to cloudlets has been taken from Round Robin algorithm and allocated the VM that have least loaded has been taken from ESCE algorithm.

At beginning, it works as round robin algorithm and if some nodes were heavily loaded and others remain idle , count loop increases too much and stop and enter a second loop that works as ESCE algorithm to again Allocated to the VM that have least loaded , which will improve the node that idle in round robin algorithm and waiting time. The proposed algorithm is implemented for an IaaS framework in simulated cloud computing environment and all the results are analyzed. All this work is done using a Cloud Analyst tool. This tool is completely based on Java.

Following versions of tools and software are used during work:

a) NetBeans: It is a Software Development platform written in Java.
b) Cloud Analyst (A tool based upon cloudsim)

The algorithm gave better results in terms of Response time, when compared with the results of Round robin algorithm and equally spread current execution algorithm. In order to evaluate the proposed algorithm. We run the experiments in Cloud Analyst simulator.

In the experiments we set the number of users. Each user base has different requests per user and we set the number of virtual machines in the data center to be 5 VMs. Simulated hosts is x86 architecture, virtual machine monitor Xen and Linux operating system. The Users are grouped by a factor of 10, and requests are grouped by a factor of 10. Each user request requires 100 instructions to be executed. The configurations file as in figure 3,4 and 5.

IRJECE

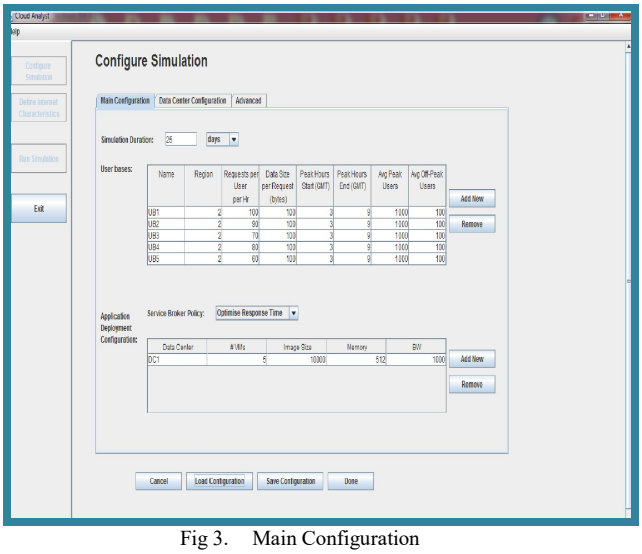


Fig 3. Main Configuration

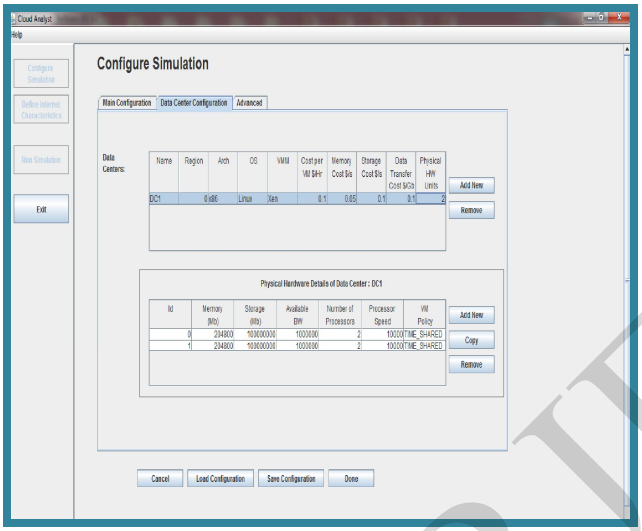


Fig 4. Data Center Configuration

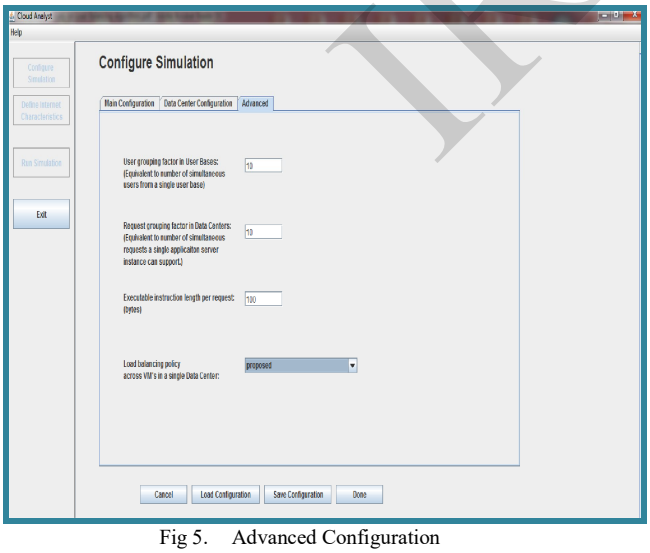


Fig 5. Advanced Configuration

The results and Comparison among these algorithms is given below in Table2.

Table II: The Result

	Round Robin	**Proposed**
Response Time(avg)	300.16	300.19
Response Time(max)	15311.51	**417.21**
Processing Time(avg)	0.39	0.42
Processing Time(max)	15032.50	**1.34**
DC Request Servicing Times(avg)	0.39	0.42
DC Request Servicing Times(max)	15032.50	**1.34**

VIII. CONCLUSION

In this study, a new algorithm is proposed and then implemented in a cloud computing environment using CloudAnalyst simulator in Java language.

As show in the table that the overall response time and data centre processing time is improved. We s noticed that was reducing the maximum of response time and processing time when each user base has different requests in this way the algorithm was efficient.

ACKNOWLEDGMENT

I cannot express enough thanks to my mother-may Allah have mercy on her- for her continued support and encouragement.

I offer my sincere appreciation for the learning opportunities provided by my supervisor.

My completion of this paper could not have been accomplished without the support of my supervisor, Dr. Fahad AL-Dosari – Thank You for your efforts in spite of you are busy.

Finally, to my parents and supervisor: my deepest gratitude for you for your encouragement. My heartfelt thanks.

REFERENCES

[1] https://en.wikipedia.org/wiki/Mailbox_provider

[2] S.K. Bagwaiya, V.; Raghuwanshi. "Hybrid approach using throttled and esce load balancing algorithms in cloud computing". in Green Computing Communication and Electrical Engineering (ICGCCEE),), 2014 International Conference, pages 1-6, 6-8 March 2014.K. Elissa, “Title of paper if known,” unpublished.

[3] Al Nuaimi, Klaithem, et al. "A survey of load balancing in cloud computing: Challenges and algorithms." Network Cloud Computing and Applications (NCCA), 2012 Second Symposium on. IEEE, 2012.first word capitalized,” J. Name Stand. Abbrev., in press.

IRJECE

[4] Wang, Lizhe, et al. "Cloud computing: a perspective study." New Generation Computing 28.2 (2010): 137-146.
[5] Yashpalsinh Jadeja and Kirit Modi. "Cloud computing-concepts, architecture and challenges.". Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on. IEEE, 2012.
[6] Jaiswal, A. S., V. M. Thakare, and S. S. Sherekar. "Study and Analysis of Architecture Components of Cloudlets in MCC." International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE) (2015): 376.
[7] Lamba, Sonia, and Dharmendra Kumar. "A Comparative Study on Load Balancing Algorithms with Different Service Broker Policies in Cloud Computing." International Journal of Computer Science and Information Technologies,(IJCSIT) Vol 5.4 (2014).
[8] Sharma, Tejinder, and Vijay Kumar Banga. "Efficient and Enhanced Algorithm in Cloud Computing." International Journal of Soft Computing and Engineering (IJSCE) ISSN (2013): 2231-2307.
[9] Shaw, Subhadra Bose, and A. K. Singh. "A survey on scheduling and load balancing techniques in cloud computing environment." Computer and Communication Technology (ICCCT), 2014 International Conference on. IEEE, 2014. [10].
[10] Shaw, Subhadra Bose, and A. K. Singh. "A survey on scheduling and load balancing techniques in cloud computing environment." Computer and Communication Technology (ICCCT), 2014 International Conference on. IEEE, 2014.
[11] Zaouch, Amal, and Faouzia Benabbou. "Load Balancing for Improved Quality of Service in the Cloud." International Journal Of Advanced Computer Science And Applications, Vol6, No7 (2015).
[12] Kundu, Anindita. International Journal of Engineering Science & Research Tecnology “An Efficient Fuzzy Load Balancing Algorithm for Public Clouds."

IRJECE