Performance Evaluation of Hybrid Round Robin Algorithm and Modified Round Robin Algorithm in Cloud Computing

Ilya Musa Adamu, A.Y Gital, S. Boukari, Ismail Zahraddeen Yakubu

Abstract: The scheduling Round Robin (RR) is an impartial algorithm that schedules cloud resources by giving static time quantum to all processes. Time quantum selection is very crucial as it determines performance of algorithms. This research paper suggests an approach to improve RR scheduling algorithm in cloud computing by considering the quantum to be equal to burst time of start request, which dynamically vary after each execution of a request. And also, if the remaining burst time of CPU for currently executing process is lesser than time quantum, then the CPU will be allocated again to the executing process for rest of CPU burst time. MatLab was used to implement the planned algorithm and benchmarked against MRRA available in literature. In comparison with the planned algorithm, Average Turnaround Time (ATAT) and minimal Average Waiting Time (AWT) was recorded. Based on the obtained simulated outcome, the planned algorithm should be preferred over modified round robin algorithm as it significantly improves the system efficiency.

Keywords: Cloud Computing, throughput, Cloud Services, Response Time, Turnaround Time.

I. INTRODUCTION

The popularity, technologies in storage, fast growth in processing and success of the Internet in recent years has contributed to computing resources to become cheaper, powerful and globally accessible [1,2]. The advanced trend is termed as cloud computing and has metamorphosed to offer solution to future and current Information Communication Technology (ICT) requirements [3, 4]. It is the latest paradigm for computing where data services and applications are provided over the Internet [5]. Currently, most of the educational institutions and business organizations use cloud environment for data storage and other computational task such as CPU and memory usage [6]. It is also service focused that offers low-cost and high quality information services based on pay-as-use model in which Cloud Service Providers (CSP) does offer guarantees [7]. Cloud computing is emerging as a commercial reality and recently been a booming area in the information technology domain. It is the use of computing resources such as software’s and hardware’s which are offered to customers as a service over the internet in various locations in globe [8]. Customers access the cloud services through internet by using Mobile devices, PDAs and PC, and Service providers provide the service to its customer [9]. These cloud services include Infrastructure as a service (IaaS), which typically involves virtualization technology to share hardware resources for executing cloud requests [10].

II. CLOUD RESOURCE PROVISIONING REQUIREMENTS

Scalability, quality of service, timeliness and flexibility are the basic requirements presenting the effectiveness of resource allocation techniques. Identifying the exact characteristics of a resource allocation technique is essential to establish cloud resource allocation requirements and develop a system in line with the nature of cloud computing [11]. Provisioning of equivalent or nearly-equivalent access to each cloud clients to its services is the major goal of cloud computing. In this context, the challenges confronting cloud computing is the ability of several services; like web objects or cloud applications, to efficiently manage the physical resource they’re running on. Application provisioning, processes and so on to an increasing numbers of clients is viewed as scalability [12]. Scalability in Cloud shall be considered as the performance report card. Quite a number of cloud users believed that all issues related to scalability are automatically addressed by the cloud itself, as a result of resources virtualization.
Certainly, this is at all not the case. It might be agreed because various cloud networks are established to grant certain clients access to vast computing resources than other clients of cloud. 

| Retrieval Number: A9139058119/19©BEIESP |
| DOI: 10.35940/ije.9139.078219 |

clarification to the question of cloud environment scalability, where services demand varies frequently, automated scheduling to cope up with this demand is becoming an issue in the industry, academic and research world. Proper resource management allows an optimal usage of computing resources, thus of the entire infrastructures that constitute the cloud, because mapping of resources to workloads is more efficient. Consequently, it assists in the achievement of better level of Quality of Service needed by clients. Providing more computational resources is easily achieved and feasible in cloud, as virtualized computing resources are considered limitless by clients. Furthermore, more computational resources provisioning to meet the client request steps up the total cost clients pay and demands the CSP employ efficient mechanisms. Quality of service performance, composed of individual features performance for the physical resources that are needed to satisfy the client’s requirement which determines the degree of client satisfaction. Ensuring quality of service in cloud however, is nontrivial task, as there are several clients having different service demands operating on internet based platforms. Moreover, several service providers can provide similar service by deploying different technologies. 

Timeliness refers to the expected time to access data. It can be measured as the time between when data is readily available for use since from when it is expected. The ability for the resource allocation solution to adapt to possible or future changes in its requirements can be described as flexibility in cloud computing. When you build or design a solution in the future, you should try to cater for these changes which arise inevitably. It should be catered for system design as a whole but there is also no reason not to also incorporate it with the smaller system aspects. The loose coupling of components is the key to building highly flexible systems. For example, there is no justification designing a solution that is firmly integrated with active directory if this part could change down the line. It is not a good idea even for dynamic strategies that do change in future. These requirements are measured using some metrics such as amount of storage provisioned, number of I/O request, and amount of CPU cycles and so on. To compare the performance of resource allocation techniques, some possible performance metrics are considered based on the requirements, these includes response time, turnaround time, wait time, context switch, throughput etc. These cloud resource allocation requirements metrics are described as below:

- **Throughput** is the most evident challenge with resource allocation in cloud. The number of completed processes within a time unit is referred to as throughput.
- **Turnaround Time (TAT):** As a challenge is total periods spent by an executing program waiting in ready list to get into memory, running on CPU and performing I/O. Minimization is necessary.
- **Waiting Time (WT):** Waiting time as a requirement is the amount of time a process waits in ready list.
- **Response Time:** Large response time is major setbacks in much resource allocation techniques as it reduces performance of the system. Response time as a characteristic is time taken to begin responding, not output response time. 

**Context Switch** is computation intensive that leads to time wastage, storage, scheduler overhead and many more resource. The aim of Majority of operating system design is optimization these features. Context switch is a feature that deals with storing and restoring CPU state to maintain state of execution at later time.

### III. RELATED WORKS

“Abdulrahim et al.” proposed to improved the work of (Manish & AbdulKadir, 2012). The planned algorithm was benchmarked against five other algorithms. From their experiment, the planned algorithm produces minimal TAT, average WT and number of context switches more than the other algorithms in their literature, but has a major drawback on response time. “Pandaba et al.” proposed an algorithm by modifying the round robin resource allocation strategy. Time equals to the time of first request was the first thing the algorithm starts with, that alternates after the end of first request. The algorithm calculates the average of sum of the times of requests found in the ready list when a new request is arrives into the ready list in order to be granted, including the new arrival request. Two registers are read; (i) AR: To store average of the burst time by dividing the value found in the SR by the count of requests found in the ready list (ii) SR: To store the sum of the remaining burst time in the ready queue. The modified algorithm performs reasonably regarding the wait time but fail in the response and turnaround time. “Dash et al.” proposes an algorithm that uses dynamic time quantum. The processes are arranged in an ascending order in the ready list base on their burst time. The average active processes in the ready list are calculated and set as the time quantum after each iteration. Their results improved performance of the average waiting time and context switching, but have a major drawback on response time and turnaround time. “Kathuria et al.” proposes an algorithm that modifies the RR algorithm with a dynamic time quantum. The algorithm calculates time quantum by maintain two different queues namely PRQ and RQ. The time quantum is obtained by taking the mean of the burst time of processes in the RQ. When it allocate CPU to first process present in RQ if the time quantum is greater than the remaining CPU burst time of the process that is currently executing then the CPU will still allocate the currently executing process for remaining CPU burst time and when the process complete its execution, it will be removed from the ready list and allocate CPU to next process. This mechanism helps in reduction of context switching and WT, but has a major drawback on response time and turnaround time. “Khatri, 2001” proposes an improve Dynamic Round Robin model in which the median of the set of processes in the ready list is considered as the optimal time quantum and if the median is less than 25 then its value must be considered as 25 to avoid the overhead of context switch. The leading process in the ready list is allocated to the CPU for a time interval of less than or equal to 1 time quantum.
If the remaining burst time of the currently executing process is within the time quantum interval, the processor again allocated to the same process. Else the process is preempted and placed at the rear of the ready list otherwise the process will be halted. Experimental result shows that the proposed algorithm IDRR, when compared with various variants of Round Robin algorithm it produces minimal

IV. METHODOLOGY

The proposed research will focus on improving the work of “Pandaba et al.\(^{[27]}\) by considering the quantum to be equal to burst time of start request, which dynamically changes after each request execution. And also, If the time quantum is greater than the remaining CPU burst time of the process that is currently executing, then the currently executing process of CPU will be allocated to the remaining CPU burst time.

A. Hybrid Round Robin Algorithm

a. Physical Model

This algorithm will consider the quantum to be equal to burst time of the request that comes first, which changes dynamically after each request execution. In order to be granted if a new request is arrives; the HRR algorithm computes the sum average of the times for all the requests in the ready list in addition to the newly arrived request(s), i.e. Quantum= total CPU Time/ no of Jobs. The total process Time refers to sum of remaining process Time at the end of each iteration, hence, no of Jobs represents the number of running jobs. Two registers will be needed: (i) ABR: To record the average burst times which is calculated by dividing BR value with total number of requests in ready list (ii) BR: To record the remaining burst time in ready list. When a request finishes its burst time after execution, the request is eliminated from the ready list BR is then recalculated by subtracting the amount of time request consumed at execution. Also, if the time quantum is more than the remaining CPU burst time of the running process, then CPU will be allocated again to the process that is currently executing for remaining CPU burst time. The process is eliminated from the ready list after completion of execution and the Processor is allocated to the subsequent process in the ready list. ABR is recalculated with respect to the burst time of the new request. While ready list is not empty, TAT and average WT time will be calculated.

b. Conceptual Model

The proposed algorithm is defined below:

Begin
Input: \(A_{\text{Req}}, P_{\text{new}}, \ ST_{\text{Rem}}, TQ, B_{\text{Time}}(P)\) Ready List
\(P \leftarrow P_{\text{New}}\) (New Request)
Ready List \(\leftarrow P(P\text{ submitted to ready list})\)
Calculate New ABR
Calculate New BR
CPU queue \(\leftarrow\) Ready List(P)
While (Ready List \(\neq \emptyset\)) do
Ready List \(\leftarrow P\)
Calculate New ABR
Calculate New BR
CPU queue \(\leftarrow\) Ready List (P)
end while
If (Ready List \(\neq \emptyset\)) then
\(B_{\text{Time}}(P) = TQ\)
Calculate New ABR
Calculate New BR
Else
\(\text{Avg} (B_{\text{Time}} \text{of Ready List Request}) = TQ\)
Calculate New ABR
Calculate New BR
// By TQ Time, CPU executes P
If (P terminated and not completed) and
BT (P) < TQ
Reload (P) // for execution
then
Calculate New ABR
Calculate New BR
Else
Return (P) To the Ready List with its updated Burst Time (BT)
Calculate New ABR
Calculate New BR
End if

V. SIMULATED OUTCOME OF THE PLANNED HRRA

MATLAB was used to implement the proposed algorithm. There is significant reduction in WT and TAT of processes in the planned algorithm when compared with MRRA. The quantum was considered to be dynamic in our proposed algorithm. Also, if the time quantum is more than the remaining CPU burst time of the process currently executed by the CPU, then the CPU will be assigned again to the currently executing process for remaining CPU burst time. Fig.3.3 below shows graphical the comparison by our proposed algorithm between AWT and TAT of each job. Also table 3.1 below shows the tabular comparison of the processes.
VI. CONCLUSION

In this paper we proposed new algorithm (HRRA) for scheduling in cloud computing. This algorithm is based on improvement and revised on the MRRA for scheduling resources in cloud. Time Quantum is always an important factor for the RRA. For the effectiveness and better performance of the algorithm, result of this work uses dynamic time quantum with re-allocation policy. MATLAB was used to implement the proposed algorithm (HRRA) together with the MRRA and result was compared based on the AWT, as well as the ATAT.

According to the result of this research work, the proposed algorithm (HRRA) is preferred as the system produces minimal Average Wait Time and Average Turnaround Time compared to MRRA.

REFERENCES


**AUTHORS PROFILE**

Dr. Souley Boukari, is a Professor of Computer Science from Abubakar Tafawa Balewa University, ATBU, Bauchi Nigeria where He is currently involved in teaching, research and supervision of students both undergraduates and Postgraduates in Computer Science. He has a B-Tech, M.Sc. and PhD in Computer Science from ATBU. His current research areas are Data Mining, Machine Learning, Artificial Intelligence, Software Engineering, cyber security and Digital Forensic.

Iliya Musa Adamu, is a PhD student of Computer Science from Abubakar Tafawa Balewa University, ATBU, Bauchi Nigeria and a Semester abroad Student at SRM Institute of Science and Technology, Tamil Nadu, India. He is currently involved in teaching, research and supervision of students both undergraduates and Postgraduates in Computer Science at the Federal Polytechnic Bauchi. He has a B-Tech in Computer, M.Sc. in Computer Science. His current research areas are cloud computing and Data communication and Network.

Dr. Abdulsalam Y Gital, is a Doctor of Computer Science from Abubakar Tafawa Balewa University, Bauchi Nigeria where He is currently involved in teaching, research and supervision of students both undergraduates and Postgraduates in Computer Science. He has a B-Tech, M.Sc. and PhD in Computer Science. His current research areas are Modeling and Simulation, Cloud Computing, Collaborative Virtual Environment and Computer Communication and Network.

Ismail Zahraddeen Yakubu, is a PG student of Computer Science and Engineering in SRM Institute of Science and Technology, Tamil Nadu, India. He is a Staff in Department of Computer Science at Federal Polytechnic Bauchi. He has a B.Sc. in Computer Science. His current research areas are Cloud Computing and Data communication and Networks, Image Processing, Database, Data Mining, Software Engineering and Internet of Things.