Dynamic Job Scheduling Algorithms Based on Round Robin for Cloud Environment

Bossy Mohamed, Noha E. AL-Attar, Wael Awad and Fatma A. Omara
Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt
Faculty of Engineering, Delta University for Science and Technology, Dakahlia, Egypt
Department of Computer Science, Faculty of Computers and Information, Cairo University, Cairo, Egypt

Abstract: This study attempts to solve the problem of the static scheduling algorithms by developing a dynamic version of Round Robin scheduling algorithm; Dynamic Priority Round Robin and Enhanced Dynamic Priority Round Robin. The proposed algorithms have been developed based on a dynamic manner of choosing the quantum time according to the current status of the requested jobs in attempting to fulfill the user's requirements and improve the overall system performance and resource utilization. The implementation of the developed algorithms is done by the Cloudsim simulator. The results record that the two versions of dynamic scheduling algorithms achieve high performance and resource utilization for the Cloud system comparing with the static scheduling algorithms like Round Robin and others. Accordingly, they decrease the idle waiting, computational and turnaround time of the requested jobs. By comparing the proposed algorithms with their corresponding static Round Robin versions, it is found that; Dynamic Priority Round Robin (DPRR) algorithm has enhanced the saving in idle waiting time, the response time and turnaround time are by 25, 51 and 32%, respectively. Similarly, the idle waiting time, response time and turnaround time are decreased in the proposed Enhance Dynamic Priority Round Robin (EDPRR) algorithm by 51, 44 and 30%, respectively. Furthermore, the resource utilization has also improved by 18% and 5% for the both of developed algorithms (DPRR and EDPRR) respectively.

Keywords: Cloud computing, dynamic scheduling, round robin, static scheduling

INTRODUCTION

Over the last few years, there is a persistent increase in the size and specifications of data used by different users. This usage increasing led to the appearance of what is called big data (Hashem et al., 2015). Usually, the big data is characterized by five main features:

- Volume; which represents the size of the data that need to be processed
- Velocity; which refers to the rate of the data growth and usage
- Variety; which means that there are different types and formats of the data used in processing,
- Veracity; which means that both data results and analysis have to be accurate
- Value; which is the added value and contribution offered by data processing (Chen et al., 2012).

Due to the spread use of the big data in many fields (e.g., healthcare, science, engineering, finance, business and different society problems), it is a big challenge for data users and providers to handle this growth rate by the traditional computing infrastructure. Thus, Cloud Computing may be one of the best solutions to face this challenge (Hashem et al., 2015; Janakiraman, 2016).

Now a day, Cloud Computing gains popularity in processing, managing and analyzing the big data. In other words, it is considered as a way to get enough storage and computing power to handle big data without purchasing infrastructure assets (Chang and Wills, 2016). Due to the pay-as-you-go fashion in Cloud Computing, the user can ask for the resources only as he needs and the cloud providers will provision the required resources dynamically and process large data sets in a parallel way (Katyal and Mishra, 2013).

Generally, when a user requests a service from the Cloud environment, the cloud provider treats this service as a job to be executed. In order to execute the required job, many processes need to be carried out in order to deliver the requested service as; job scheduling, resources provisioning and resource allocation. Simply, the master job controller in the
Cloud manages the order of executing the existed jobs according to their priority, then the provisioning and allocation processes begin to allocate the suitable resources to the selected jobs (Vernier, 2016).

The job scheduling is a decision making process which manages the order of executing the user's jobs using a set of strategies. The main goal of the job scheduling process is to increase the resource utilization and minimize processing time of the jobs, which in turn increases the system throughput (Salot, 2013; Singh et al., 2014).

Generally, job scheduling algorithms can be divided into two main types; Preemptive and Non-Preemptive algorithms. In Preemptive algorithms; the current running jobs are temporarily suspended according to a specific condition. These jobs wait to be re-executed after a specific predefined time called "time quantum". This type of algorithms focuses on decreasing the response time for the required jobs. Round Robin (RR) is one of the significant examples of the preemptive algorithms (Singh et al., 2014).

In Non-Preemptive algorithms; once a job is selected to start its execution, it continues in running until it finishes completely. Non-Preemptive algorithms try to decrease the idle waiting time for the user's jobs. First-Come-First-Served (FCFS) and Shortest Job First (SJF) are examples of the non-preempt algorithms (Katyal and Mishra, 2013). First-Come-First-Served is considered the simplest and the default scheduling algorithm. The idea of FCFS is that it does not determine any priorities or conditions to execute the jobs; just the first arrived job will be the first finished job without any interruption. This means that the short jobs may wait for a long time before starting its execution (Chen et al., 2013) Shortest Job First (SJF) algorithm tries to solve the problem of executing the short jobs. In SJF algorithm, the jobs with small processing time will have the highest priority to be executed, so they do not have to wait the early arrived large jobs to finish their execution (Siahaan, 2016).

In Cloud Computing, the Master Virtual Machine (MVM) controller is responsible for scheduling the incoming jobs and sending them to Slaves Virtual Machines (SVMs) depending on the chosen scheduling algorithm (Chen et al., 2013). Due to the existence of multiple independent users who usually are more than the Cloud available resources, the job scheduling process becomes one of the main challenges that face the Cloud Computing community (Ernemann et al., 2002) Choosing the suitable scheduling algorithm depends on six main parameters: Resource Utilization; the ratio between the total busy time of the virtual machine and the total execution time of the parallel application, Throughput; the number of completed jobs within a certain period of time, System Efficiency; the busy time of each virtual machine, Response time; the time taken to start responding to the job request), Idle Waiting time; the waiting time taken by each job to complete its execution and Turnaround time; the total amount of time taken by the job (Samal and Mishra, 2013).

One of the most famous scheduling algorithms in Cloud Computing is Round Robin (RR) algorithm because it is suitable for job scheduling in time-sharing systems due to its simplicity, fairness and generality (Siahaan, 2016).

In Round Robin, jobs are queued with the same style in FCFS algorithm, but RR assigns a fixed equal quantity of execution time to each job which is known as Time Quantum (TQ). TQ is a fixed interval of time used to interrupt the long jobs to give a chance to other jobs to start their execution (Pinedo, 2012). After the end of TQ, the uncompleted jobs will be stopped and traced to the tail of the job queue (Salot, 2013).

RR algorithms are categorized into two types; Static RR and Dynamic RR (Aravind and Chelladurai, 2016). In Static RR, a fixed Time Quantum is set. It starts from the beginning of the jobs execution time till they are finished. Static RR ignores any information about deadline time, computation time, precedence constraints and any expected delay time (Stankovic et al., 1998). Thus, Static RR may suffer from many drawbacks due to the fixed choice of TQ as it may affect the performance of the system and resources' utilization (e.g., If TQ is large, small jobs may finish their execution after a small amount of time and unutilized resources cannot be used until the end of TQ. On the other hand, if TQ is small, the big jobs will be surely interrupted after the end of the TQ and trace to the tail of the job queue waiting for their new turn, this may decrease the system efficiency) (Samal and Mishra, 2013).

Dynamic RR is basically based on calculating TQ according to the real processing time of the current jobs, which means that the TQ will be dynamically changed every turn (Noon et al., 2011). Dynamic RR tries to enhance the static RR algorithm by increasing the performance and stability of the system and provides a self-adaptation system in which the system adapts itself according to the incoming jobs (Matarneh, 2009).

In this study, two enhanced algorithms for dynamic Round Robin algorithm have been proposed in order to overcome RR drawbacks in the Cloud Computing environment.

LITERATURE REVIEW

Many researches have been conducted in the field of Cloud Computing scheduling algorithms. Some of the researchers handled only the static scheduling algorithm as Aravind and Chelladurai (2016). They have focused on addressing scheduling fairness between CPU-bound jobs and I/O-bound jobs by developing a Fair-Share Round Robin Scheduling Algorithm (FSRR). FSRR algorithm is considered an
enhanced version of the classical round robin to decrease the unfairness problem in classical RR.

Other researchers tried to enhance the job scheduling specially in Cloud Computing as Li (2016) who has developed the Priority Based Weight Round Robin (PBWRR) algorithm. According to PBWRR algorithm, a weight for each job has been suggested according to its priority in the waiting queue in order to avoid long waiting time and minimize the response time.

Liu et al. (2013) has presented a Generalized Priority Algorithm for efficient execution of jobs. This algorithm categorized each node's computing capacity into two types; foreground virtual machines (e.g., high CPU priority) and background virtual machines (e.g., low CPU priority). This algorithm aims to increase resources utilization in the system.

Another enhancement for RR algorithm has been introduced by Mohamed et al. (2016) They have introduced an Enhanced Priority Round Robin (EPRR) which is based on listing the uncompleted jobs in a priority queue depending on the remaining execution time (i.e., the job with small remaining time will have high priority). The advantage of this algorithm is to decrease the average waiting and turnaround time of the job.

Although, using the static job scheduling algorithms is the most popular in cloud computing, it cannot handle the rapid increase in the complicated applications that are handled by the Cloud environment. Thus, there is a significant need to use efficient dynamic scheduling algorithms to manage and control the available resources, as well as, increase the system throughput.

Lee et al. (2011) have proposed a Dynamic Priority Scheduling Algorithm (DPISA) which can change the priorities of the available jobs during the scheduling time. According to DPISA algorithm, the scheduler sets a threshold value that limits the waiting time for jobs in the job queue, if the job exceeds this threshold then it will be sent to execute immediately.

Matarneh (2009) has presented a Self-Adjustment Time Quantum in Round Robin (SARR) algorithm. The SARR algorithm depends on Burst Time of the now running processes. The idea of this algorithm is to adjust the Time Quantum according to the burst time of the running processes. It tries to minimize idle waiting time, but it still needs better adjustment criteria for Time Quantum.

**MATERIALS AND METHODS**

The goal of this study is to enhance the Round Robin algorithm in such a way that it can deal with the dynamic nature of job scheduling. The original form of the RR algorithm is to use a static time quantum to execute the job within it using the preemption style.

Determining this fixed time quantum is considered the main drawback in the RR algorithm. So, the work in this study will try to handle this problem by developing two versions of the Round Robin algorithm using dynamic quantum time. The two algorithms which have been introduced are "Dynamic Priority Round Robin" and "Enhanced Dynamic Priority Round Robin".

**Dynamic Priority Round-Robin (DPRR) algorithm:**
Dynamic Priority Round Robin (DPRR) is based mainly on the dynamic change of the time quantum. At the beginning of the scheduling process, the time quantum is calculated according to the expected processing time of the requested jobs and then it is modified in every time slice depending on the processing time of the existed jobs. The Time Quantum (TQ) calculation is defined in Eq. (1):

\[
TQ = \frac{\text{AVG} (\sum PT(J))}{N}
\]

where,

- \(PT(J)\) = The processing time of every job in the system
- \(N\) = The number of requested jobs

DPRR uses two types of queues; Primary Job Queue (PJQ) which stores the incoming jobs which need to be processed in the Cloud system and Priority Queue (PQ) which is used to list the uncompleted jobs that will be executed depending on their priority which is determined according to the condition in Eq. (2):

\[
T_{\text{Remain}}(j) \leq TQ/2
\]

where, \(T_{\text{Remain}}(j)\) is the remaining time of the uncompleted job \((j)\).

Thus, DPRR algorithm is based on executing the uncompleted jobs firstly depending on their remaining time. According to Eq. (2), if the \(T_{\text{Remain}}\) of the uncompleted job is less than or equal half of the current Time Quantum, then these jobs will be listed in the PQ, otherwise, it will be queued in the tail of the PJQ.

Also, DPRR algorithm cares about achieving good resource utilization during the run time. It determines any virtual machines that become free before the end of the Time Quantum. If a free VM is existed, the scheduler will assign a new job from the head of PJQ to be executed on the free VM.

The pseudo code of DPRR algorithm is illustrated in Fig. 1. DPRR algorithm assumes that, \(m\) is the number of the available Virtual Machines (VMs) and \(N\) is the total number of the requested jobs over the Cloud system. While \(TQ\) represents the calculated Time Quantum and \(PT\) is the expected Processing Time of the requested jobs.

Unfortunately, DPRR does not concern any criteria for selecting the jobs during the runtime to be run. The
Fig. 1: Pseudo code of DPRR algorithm

```
1. Begin
2. Arrived Jobs store in PJQ*N Store as FCFS*
3. Initialize PQ by NULL* PQ initializer *
4. S.R = 0 * R is No. of waiting jobs in PQ, S is No. of processed jobs on the available VM*
5. While (PQ not equal NULL* PQ has jobs *)
6. Do
7. TQ=AVGrze PT/Job/N* Calculating Time Quantum */
8. For each job J to VM
9. Issue_Job onto VM to execute* Execute jobs *
10. S = S+1 * Count No. of executing jobs (S) *
11. End For
12. For each processed job i to VMs* locate the jobs on the free VM*/
13. If Tremain (i) is equal 0* If a job ends before the end of TQ */
14. Set VM to null* free up the VM that has finished job*
15. Issue_Job from PJQ onto VM*Put new jobs from PJQ on the free VM*
16. S = S-1
17. End If
18. End For
19. Until(TQ_equal 0)* Until the Time Quantum finishes */
20. For each processed job i to S* Selected jobs *
21. If Tremain (i) not equal 0 and less than or equal TQ/2 */ if the processed Job doesn't finish and its remaining time is less than TQ average */
22. Add i to PQ* Add job i to Priority queue */
23. Else
24. Add i to PJQ* Add job i to the tail of Primary job queue */
25. End If
26. End For
27. While (PQ is not NULL)
28. For each Job i in PQ
29. Sort jobs in PQ in ascending order according to T (i)
30. R = Count (i) * Count uncompleted processed jobs in PQ*/
31. While (R is greater than VMs)
32. TQfin= AVG(Σ T(1)/R) * Calculate Time Quantum for jobs in PQ)
33. Issue_Job from PQ onto VM to execute
34. If Tremain (i) is equal 0* If a job ends before the end of TQ */
35. Set VM to null* free up the VM that has finished job*
36. Go to 33*Put new jobs from PQ on the free VM*/
37. End If
38. S = S-1
39. End While
40. For each job i in PQ to VMs
41. Issue_Job from PQ onto VM to execute
42. End for
43. VMfree = VMs-VMs* calculate No. of free VMs */
44. TQfin= AVG(Σ T(i)/R)+ Σ T(i)/N)
45. Goto 8
46. End For
47. End While
48. End While
49. End
```

scheduler in DPRR only checks for the free VMs and then selects the job from the head of PJQ without taking into consideration if the selected job has a proper processing time to be executed on this resource or not, which may lead to decrease the performance of the system.

**Enhanced Dynamic Priority Round-Robin (EDPRR) Algorithm:** Enhanced Dynamic Priority Round Robin (EDPRR) algorithm is considered an enhancement of DPRR algorithm. It is developed to overcome the main drawback (i.e., the scheduler selects the new job to assign in to free VM from the head of the PJQ without considering any criteria for choosing the suitable job). In EDPRR algorithm, after checking if there is an existed free virtual machine during the runtime and before the end of TQ, the scheduler selects a new job from PJQ to be executed but under a condition. This
condition concerns with selecting the suitable job that can be executed on the free resource (i.e., VM) during the remaining time of the \( TQ \). Calculating the time of the free virtual machine (VM) is defined in Eq. (3):

\[
T_{\text{free}}(VM) = TQ - PT_{\text{finished_job}}
\]  

(3)

The pseudo code of the enhanced part in the proposed EDPRR algorithm is described in Fig. 2.

![Fig. 2: Pseudo code of the enhanced part in the proposed EDPRR algorithm](image)

### RESULTS AND DISCUSSION

In order to implement and evaluate the performance of the proposed scheduling algorithms (i.e., DPRR and EDPRR), CloudSim simulator is used (Buyya et al., 2009). It is a simple toolkit that is used to simulate the cloud environment. It gives the ability of a quick and easy change of the needed parameters to be assumed in the simulation that provides a high degree of configurability and flexibility (El-Attar et al., 2014).

The experimental study considers some assumptions as follows:

- It is operated on a single Data Center with 5, 8 and 10 homogenous VMs. Each VM has the same type of operating system and the same amount of processing and capacity (e.g., ‘1GHz’ processor and ‘2GB’ of RAM).
- The experiment is conducted with a varying number of jobs ‘100’, ‘150’ and ‘200’.

Each job has 2 parameters; Name and Required Capacity (Table 1). Job’s computation time will be calculated using job’s required processing capacity which varies from one job to another.

According to the results of implementing the proposed algorithms (DPRR and EDPRR), the performance can be evaluated through some of the essential parameters that measure the efficiency of the scheduling process using the proposed algorithms. The evaluation parameters are; the response time, idle waiting time, turnaround time and utilization performance. In this evaluation, the comparison has been done among the two proposed algorithms (DPRR and EDPRR) as dynamic scheduling algorithms from a side and a static scheduling algorithm EPRR (i.e., it depends only on listing the uncompleted jobs in a priority queue with a static amount of time quantum \( TQ \)) which have been presented before in Matarneh (2009) on the other side. The comparative results of the three algorithms based on the four evaluation parameters (i.e., Idle Waiting time, Response Time, Turnaround Time and Resource Utilization) are presented in Table 2 to 5 respectively. Also, the enhancement percentages of the proposed algorithms are calculated according to Eq. (4):

\[
\text{Improving percentage} = \frac{P1J - P2J}{P2J} \times 100 \% 
\]  

(4)

where, \( P1 \) is the default scheduling algorithm and \( P2 \) represents the enhanced algorithm to be compared. While \( i, j \) are the jobs and the VMs respectively.

Resource utilization has been evaluated using Eq. (5) which defines the utilization function as the ratio between the total busy time of the virtual machines and the total finish execution time of the parallel application (Abdelkader and Omara, 2012):

\[
\text{Utilization} = \frac{\text{VM Busy Time}}{\text{Make Span}} \times 100
\]  

(5)

As shown in Table 2 to 5, the performance evaluation has been done over two phases. First, the static scheduling (EPRR) algorithm is compared with the first proposed dynamic (DPRR) algorithm. According to this phase, it is found that the performance of the proposed DPRR outperforms the static EPRR algorithm by 79% and 65% for both response and turnaround time respectively. Furthermore, the proposed DPRR algorithm utilizes the resources that are freed up during the run time by selecting a new job from the PJQ. So, the idle waiting time is decreased in the proposed DPRR algorithm by 75% with respect to the static EPRR algorithm. Also, the resource utilization is enhanced by 17%.

On the other hand, for the dynamic algorithms, the second proposed EDPRR algorithm has achieved a
better performance than that the first proposed (DPRR) algorithm. The overall performance of the system after using EDPRR algorithm is improved by decreasing the idle waiting time, the response time and the turnaround time of jobs with 5, 13 and 19%, respectively with respect to the proposed DPRR algorithm. Also, EDPRR algorithm improves the resources utilization by 8% with respect to DPRR algorithm.

The performance evaluation also handles the effect of changing the number of the available virtual machines. The proposed DPRR and EDPRR have been implemented using 5, 8 and 10 VMs, with the same amount of requested jobs (i.e., 100, 150 and 200). According to the implementation results, it is found that the increasing of the number of virtual machine enhances the scheduling process, where by increasing the number of virtual machine to 8 instead of 5, the idle waiting time, the response time and turnaround time are decreased in the proposed DPRR algorithm by 25, 51 and 32%, respectively. Similarly, the response time and turnaround time are decreased in the proposed EDPRR algorithm by 51, 44 and 30%, respectively. Furthermore, the resource utilization has also improved by 18% and 5% for DPRR and EDPRR algorithms respectively (Fig. 3 to 6).

Finally, as it is shown in the above comparison, the dynamic scheduling algorithms have enhanced the performance of the system better than the static ones.
The performance comparison between the static EPRR, DPRR and EDPRR algorithms is plotted in Fig. 7.
REFERENCES


