Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 85 (2016) 878 – 890

Procedia Computer Science

International Conference on Computational Modeling and Security (CMS 2016)

Modified Round Robin Algorithm for Resource Allocation in Cloud Computing

Pandaba Pradhan[1]
Department of CSE, ITER
SOA University, Odisha, India
ppradhan11@gmail.com

Prafulla Ku. Behera
Reader, Utkal University
BBSR, Odisha, India
p_behera@hotmail.com

B N B Ray
Reader, Utkal University
BBSR, Odisha, India
bnbray@yahoo.com

Abstract

Cloud computing is an attractive computing model since it allows for the provision of resources on-demand. Cloud computing has emerged as a new technology that has got huge potentials in enterprises and markets. Clouds can make it possible to access applications and associated data from anywhere. Companies are able to rent resources from cloud for storage and other computational purposes so that their infrastructure cost can be reduced significantly. Hence there is no need for getting licenses for individual products. Cloud Computing offers an interesting solution for software development and access of content with transparency of the underlying infrastructure locality. The Cloud infrastructure is usually composed of several data centers and consumers have access to only a slice of the computational power over a scalable network. The provision of these computational resources is controlled by a provider, and resources are allocated in an elastic way, according to consumers' needs. However one of the major pitfalls in cloud computing is related to optimizing the resources being allocated. The other challenges of resource allocation are meeting customer demands and application requirements. In this paper, modified round robin resource allocation algorithm is proposed to satisfy customer demands by

[1]* Corresponding author. Tel.: +91-9437628327;
E-mail address: ppradhan11@gmail.com

1877-0509 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the Organizing Committee of CMS 2016
doi:10.1016/j.procs.2016.05.278

reducing the waiting time.

Keywords: *Cloud; Cloud Computing; Cloud Users; Cloud Services; Resource Allocation, Customer Demand; Infrastructure; Resource Allocation Algorithm.*

1. Introduction

Currently Cloud Computing[3,4] is an emerging computing technology which is the big step in development and deployment of an increasing number of distributed applications. Cloud Computing is defined as the computing model that operates based on Clouds. In turn, the Cloud is defined as a conceptual layer[11] that operates above an infrastructure to provide services in a timely manner. Cloud computing emerges as a new computing paradigm which aims to provide reliable, customized and QoS (Quality of Service) guaranteed computing dynamic environments for end-users[2].Distributed processing, parallel processing and grid computing together emerged as cloud computing. The basic principle of cloud computing is that user data is not stored locally but is stored in the data center of internet.

According to the NIST definition[15], Cloud computing is a model for enabling ubiquitous, 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.

Cloud computing nowadays becomes quite popular among community of cloud users by offering a variety of resources. Cloud computing platforms[12], such as those provided by Microsoft, Amazon, Google, IBM, and Hewlett-Packard, let developers deploy applications across computers hosted by a central organization. Developers obtain the advantages of a managed computing platform, without having to commit resources to design, build and maintain the network.

There are numerous advantages of cloud computing the most basic ones being lower costs, re-provisioning[16] of resources and remote accessibility. Cloud computing lowers cost by avoiding the capital expenditure by the company in renting the physical infrastructure from a third party provider. Due to the flexible nature of cloud computing, we can quickly access more resources from cloud providers when we need to expand our business. The remote accessibility enables us to access the cloud services from anywhere at any time. To gain the maximum degree of the above mentioned benefits, the services offered in terms of resources should be allocated optimally to the applications running in the cloud.

Cloud computing, at its simplest, is a collection of computing software and services available from a decentralized network[4] of servers. The term “cloud” has long been used as a metaphor for the Internet, and there are many popular services and Web sites which you may already be enjoying, without being aware that they are cloud-based. Social networking sites, Web-based email clients like Yahoo! and Gmail, Wikipedia and YouTube, and even peer-to-peer networks like Skype or Bit Torrent are all applications that run in the cloud.

2. Cloud Deployment Model

Cloud computing deployment models are of 4 basic types:

2.1 Public Cloud

Public clouds[4,10], as shown in fig 1.1, are owned and operated by companies that use them to offer rapid access to affordable computing resources to other organizations or individuals. With public cloud services, users don't need to purchase hardware, software or supporting infrastructure, which is owned and managed by providers. Public cloud functions on the prime principle of storage demand scalability.

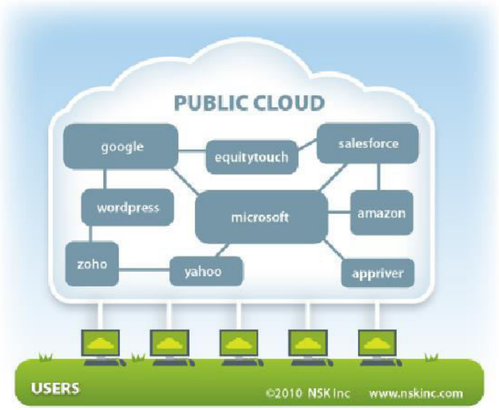


Figure 1.1: Public Cloud

Public clouds help in realization of characteristics like:

- Flexible and Elastic Environment
- Freedom of Self-Service
- Pay for what you use
- Availability and Reliability

2.2 Private Cloud

Private cloud[4,10] is a cloud infrastructure build exclusively for a single organization, deployed within certain boundaries like firewall settings whether managed internally or by a third-party and hosted internally or externally.

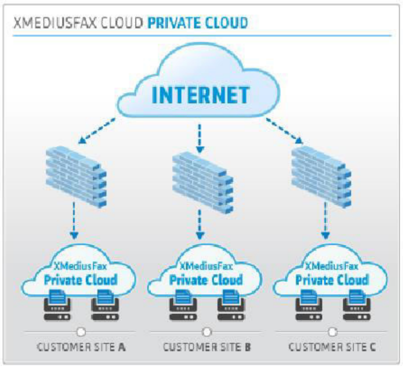


Figure 1.2: Private Cloud

Private clouds(as shown in fig 1.2) help in realization of characteristics like:

- Enhanced Security Measures
- Dedicated Resources
- Greater Customization

2.3 **Hybrid Cloud**

The cloud infrastructure consists of a number of clouds of any type,but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.
Hybrid Clouds[4,10](as shown in fig 1.3) realize the characteristics:

. Optimal utilization
. Data centre consolidation

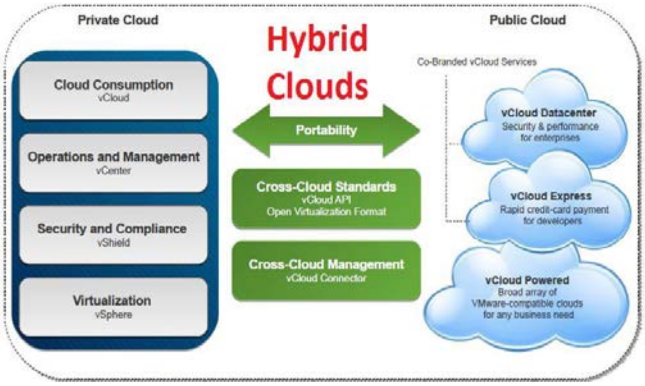


Figure 1.3: Hybrid Cloud

2.4 **Community Cloud**

The cloud infrastructure is shared between the organizations with similar interests and requirements whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so only some of the cost savings potential of cloud computing are realized and is shown in fig 1.4.

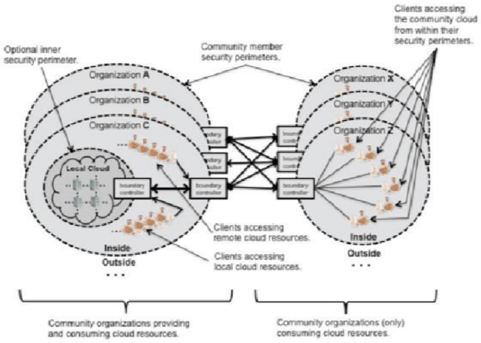

Clients accessing the community cloud from within their security perimeters.
Optional inner security perimeter.
Community member security perimeters.
Organization A
Organization B
Organization C
Organization X
Organization Y
Organization Z
Local Cloud
boundary controller
boundary controller
boundary controller
Clients accessing remote cloud resources.
Clients accessing local cloud resources.
Inside
Outside
Inside
Outside
Community organizations providing and consuming cloud resources.
Community organizations (only) consuming cloud resources.

Figure 1.4: Community Cloud

3. **Cloud Service Models**

Cloud computing providers offer their services according to several fundamental models[10,15](as shown in fig 2) : Infrastructure as a service (IaaS), Platform as a service (PaaS), and Software as a service (SaaS) where IaaS is the most basic model. Each higher model abstracts from the details of the lower models.

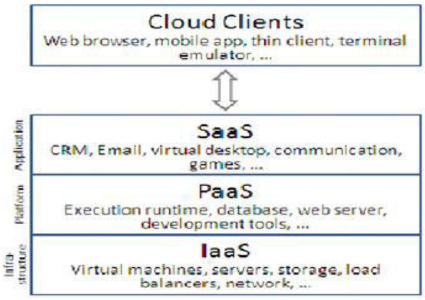

Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
Application
SaaS
CRM, Email, virtual desktop, communication, games, ...
Platform
PaaS
Execution runtime, database, web server, development tools, ...
Infra-structure
IaaS
Virtual machines, servers, storage, load balancers, network, ...

Figure 2: Cloud Service Models

3.1 **Infrastructure as a Service (IaaS)**

In the most basic cloud-service model, providers of IaaS offer computers - physical or (more often) virtual machines and other resources. IaaS clouds often offer additional resources such as a virtual-machine disk image library, raw (block) and file-based storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and

software bundles. IaaS-cloud providers supply these resources on-demand from their large pools installed in data centers. For wide-area connectivity, customers can use either the Internet or carrier clouds (dedicated virtual private networks). To deploy their applications, cloud users install operating-system images and their application software on the cloud infrastructure. In this model, the cloud user patches and maintains the operating systems and the application software. Leading vendors that provide Infrastructure as a service are Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.

3.2 **Platform as a Service (Paas)**

In the PaaS models, cloud providers deliver a computing platform,typically including operating system, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers. With some PaaS offers like Windows Azure, the underlying computer and storage resources scale automatically to match application demand so that the cloud user does not have to allocate resources manually. Typical players in PaaS are Google’s Application Engine, and Salesforce.com’s force.com .

3.3 **Software as a Service (Saas)**

In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as ”on-demand software” and is usually priced on a pay-per-use basis. SaaS providers generally price applications using a subscription fee. In this model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. Cloud users do not manage the cloud infrastructure and platform where the application runs.This eliminates the need to install and run the application on the cloud user’s own computers, which simplifies maintenance and support.The pioneer in this field has been Salesforce.coms offering in the online Customer Relationship Management (CRM) space.

4. **Resource Allocation in Cloud Computing**

In cloud computing,Resource Allocation[2,5,13] (RA) is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation starves services if the allocation is not managed precisely. Resource provisioning solves that problem by allowing the service providers to manage the resources for each individual module.

Resource Allocation Strategy (RAS)[2,6,14] is all about integrating cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment so as to meet the needs of the cloud application. It requires the type and amount of resources needed by each application in order to complete a user job. The order and time of allocation of resources are also an input for an optimal RAS. An optimal RAS should avoid the following criteria as follows:

i) ***Resource contention*** *situation arises when two applications try to access the same resource at the same time.*

*ii) S**carcity of resources** arises when there are limited resources.*

iii) ***Resource fragmentation*** *situation arises when the resources are isolated. [There will be enough resources but not able to allocate to the needed application.]*

iv) ***Over-provisioning*** *of resources arises when the application gets surplus resources than the demanded one.*

v) ***Under-provisioning*** *of resources occurs when the application is assigned with fewer numbers of resources than the demand.*

From the perspective of a cloud provider, predicting the dynamic nature of users[3,7] ,user demands, and application demands are impractical. For the cloud users, the job should be completed on time with minimal cost. Hence due to limited resources, resource heterogeneity, locality restrictions, environmental necessities and dynamic nature of resource demand, we need an efficient resource allocation system that suits cloud environments.

5. Different Resource Allocation Policies

Several resource allocation schemes[8,9,16] have come up in the literature of cloud computing. Researchers around the world have proposed and / or implemented several types of resource allocation policies. Few of the algorithms for resource allocation in cloud computing are covered here briefly.

5.1 Round Robin Algorithm

It is one of the oldest, simplest, fairest and most widely used scheduling algorithms, designed especially for time-sharing systems. A small unit of time, called time slice or quantum[11] is defined. All runable processes are kept in a circular queue. The CPU scheduler goes around this queue, allocating the CPU to each process for a time interval of one quantum. New processes are added to the tail of the queue.

The CPU scheduler picks the first process from the queue, sets a timer to interrupt after one quantum, and dispatches the process.

If the process is still running at the end of the quantum, the CPU is pre-empted and the process is added to the tail of the queue. If the process finishes before the end of the quantum, the process itself releases the CPU voluntarily. In either case, the CPU scheduler assigns the CPU to the next process in the ready queue. Every time a process is granted the CPU, a context switch occurs, which adds overhead to the process execution time.

An example of round robin algorithm is given below:

- CPU job burst times & order in queue
  - P1: 20
  - P2: 12
  - P3: 8
  - P4: 16
  - P5: 4
- The Gantt chart shown as in fig 3, and the average wait time are given as below:

(Let time quantum of 4)

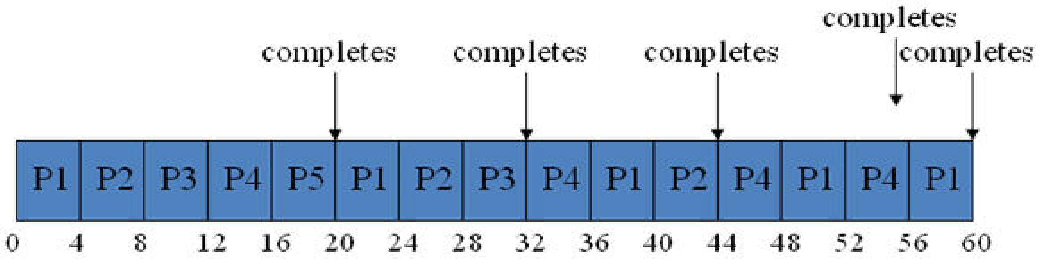


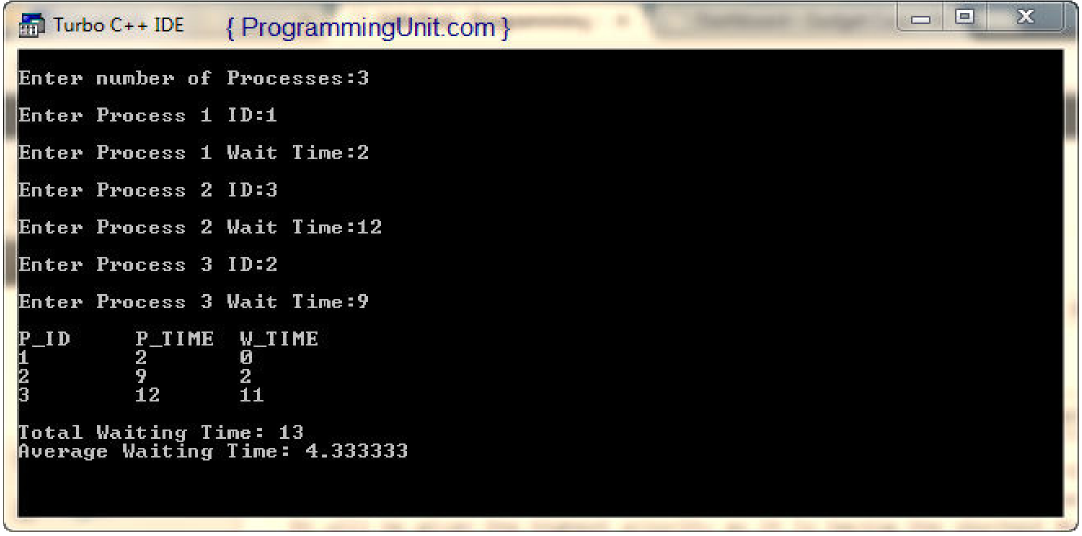
Figure 3: Gantt chart of RR Algorithm

Then the Waiting times are:

P1: 60 - 20 = 40
P2: 44 - 12 = 32
P3: 32 - 8 = 24
P4: 56 - 16 = 40
P5: 20 - 4 = 16

- Average wait time: 30.4

The execution of round robin algorithm snapshot is given as below:



```
Enter number of Processes:3

Enter Process 1 ID:1

Enter Process 1 Wait Time:2

Enter Process 2 ID:3

Enter Process 2 Wait Time:12

Enter Process 3 ID:2

Enter Process 3 Wait Time:9

P_ID      P_TIME  W_TIME
1         2       0
2         9       2
3         12      11

Total Waiting Time: 13
Average Waiting Time: 4.333333
```

5.2 SJF Algorithm

The SJF[17] scheduler is exactly like FCFS except that instead of choosing the job at the front of the queue, it will always choose the shortest job (i.e. the job that takes the least time) available. We will use a sorted list to order the processes from longest to shortest. When adding a new process/task, we need to figure out the where in the list to insert it.

An example of round robin algorithm is given below:

- CPU job burst times & order in queue
  - P1: 20
  - P2: 12
  - P3: 8
  - P4: 16
  - P5: 4
- The Gantt chart shown as in fig 4, and the average wait time are given as below:

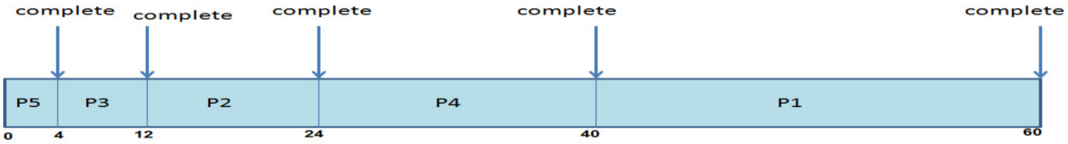


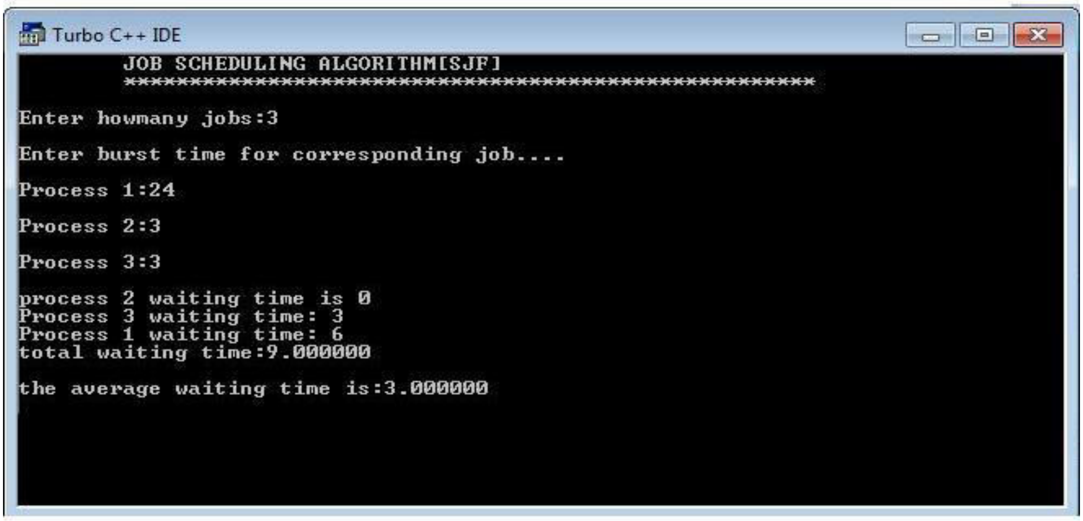
Figure 4: Gantt chart of SJF Algorithm

Waiting times:

P1: 60 - 20 = 40 ; P2: 24 - 12 = 12; P3: 12 - 8 = 4; P4: 56 - 16 = 40 ; P5: 4 - 4 = 0

Average wait time: 16.8

The execution of SJF algorithm snapshot is given as below:



```
        JOB SCHEDULING ALGORITHM[SJF]
        *****************************************************

Enter howmany jobs:3

Enter burst time for corresponding job....

Process 1:24

Process 2:3

Process 3:3

process 2 waiting time is 0
Process 3 waiting time: 3
Process 1 waiting time: 6
total waiting time:9.000000

the average waiting time is:3.000000
```

5.3 Comparision study of Round Robin Algorithm and SJF

SJF needs to know how long a process is going to run (i.e. it needs to predict the future). This run time estimation feature may be hard to implement, and thus SJF is not a widely used scheduling scheme.RR is a preemptive scheduler, which is designed especially for time-sharing systems. In other words, it does not wait for a process to finish or give up control. In RR, each process is given a time slot to run. If the process does not finish, it will "get back in line" and receive another time slot until it has completed.

The performance of RR[9,16] depends on the size of the time quantum, and if the time quantum is large, RR will behave jus like the FCFS policy. In general, we want the time quantum to be large with respect to the context-switch time (cs should be around 10% of time quantum).

6. Modified Round Robin Algorithm

Here a modified round robin scheduling algorithm is proposed for resource allocation in cloud computing.

METHODOLOGY:

This algorithm begins with the time equals to the time of first request, which changes after the end of first request. When a new request is added into the ready queue in order to be granted, the algorithm calculates the average of sum of the times of requests found in the ready queue including the new arrival request.This needs two registers: (i)SR: To store the sum of the remaining burst time in the ready queue (ii)AR: To store avg. of the burst times by dividing the value found in the SR by the count of requests found in the ready queue. After execution, if request finishes its burst time, then it will be removed from ready queue or else it will move to the end of the ready queue. SR will be updated by subtracting the time consumed by this request. AR will be updated according to the new data.

- The algorithm is as follows:

Algorithm 1: Modified Round Robin Algorithm

```
Begin
I/P : SR , AR , $P_n$ , BT(P) , TQ , Ready Queue
New request P arrives
```

```
        P Enters ready queue
        Update SR and AR
        Request P is loaded from ready queue into CPU queue to be executed
        While (Ready Queue! = NULL) do
          Ready Queue     P
          Update SR& AR
          Load P     // For Execution
        end while
    If (Ready Queue = NULL) then
        TQ   =BT (P)
        Update SR & AR
      else
        TQ = AVG (BT of all request in Ready Queue)
        Update SR & AR
 // CPU executes P by TQ Time
If (P terminated) then
        Update SR & AR
      else
        Return P    // To the Ready Queue with its updated Burst Time (BT)
        Update SR & AR
end if
---------------------------------------------------------------------------------------------------
```

7. Simulation and Result:

The proposed algorithm is implemented in MATLAB. The turnaround and waiting time of different processes in our algorithm are significantly reduced.In our proposed algorithm, we have consider the quantum to be dynamic i.e. quantum =totalCPUTime/noOfJobs.Where totalCPUTime is the sum of remaining CPUTime after each iteration, and noOfJobs rep-resents the total no. Of active jobs (i.e. jobs having CPUTime greater than 0). Fig.5 shows the avg. waiting time and turn around time of each job by our proposed algorithm.

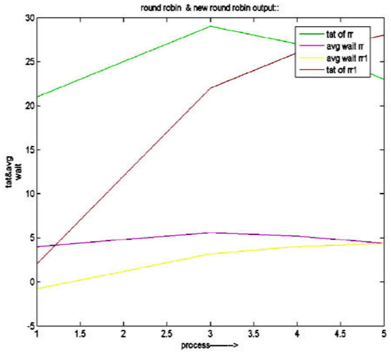


Fig5: Comparison Result

8. Conclusion and Future work

In this paper, we have discussed the Modified Round Robin algorithm that is a simple step for obtaining an optimal scheduling model. Scheduling algorithms should not affect the behaviour of the system(same results regardless of schedule) .However, the algorithm do impact the system's efficiency and response time .Time quantum is the bottleneck facing round robin algorithm and was more frequently asked question: What is the optimal time quantum to be used in round robin algorithm? In light of the effectiveness and the efficiency of the Round Robin algorithm, this paper provides an answer to this question by using dynamic time quantum instead of fixed time quantum. It will need much more efforts and researches to score a goal.

References

1. V.Vinothina, Dr.R.Sridaran, Dr.PadmavathiGanapathi," A Survey on Resource Allocation Strategies in Cloud Computing", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012.

2.Ya gız Onat Yazır, Chris Matthews, Roozbeh Farahbod Stephen Neville, Adel Guitouni, Sudhakar Ganti and Yvonne Coady," Dynamic Resource Allocation in Computing Clouds using Distributed Multiple Criteria Decision Analysis",2010 IEEE DOI 10.1109/CLOUD.2010.66 91, 2010 IEEE 3rd International Conference on Cloud Computing.

3. Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, Zhong Ming.(2011), "Adaptive Resource Allocation for Pre-empt able Jobs in Cloud Systems",in 10th International Conference on Intelligent System Design and Application, Jan. , pp. 31-36.

4. Mohiuddin Ahmed, Abu Sina Md. Raju Chowdhury, Mustaq Ahmed, Md. Mahmudul Hasan Rafee," An Advanced Survey on Cloud Computing and State-of-the-art Research Issues**",** IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012 ISSN .

5.Ronak Patel, Sanjay Patel," Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, February- 2013 ISSN.

6. Dorian Minarolli and Bernd Freisleben," Uitlity –based Resource Allocations for virtual machines in cloud computing"(IEEE,2011).

7.Xindong YOU, Xianghua XU, Jian Wan, Dongjin YU:RAS-M ,"Resource Allocation Strategy based on Market Mechanism in Cloud Computing"(IEEE,2009).

8.Zhen Kong et.al : Mechanism Design for Stochastic Virtual Resource Allocation in Non-Cooperative Cloud Systems: 2011 IEEE 4th International Conference on Cloud Computing :pp,614-621.

9.Goudarzi H., Pedram M.(2011),"Multi-dimensional SLA-based Resource Allocation for Multi-tier Cloud Computing Systems",in IEEE International Conference on Cloud Computing, Sep. ,pp. 324-331.

10.Qi Zhang,Lu Cheng, Raouf Boutaba."Cloud Computing: State-ofthe-art and research challenges". The Brazilian Computer Society 2010.

11. Atsuo Inomata, TaikiMorikawa, Minoru Ikebe, Sk.Md. MizanurRahman: Proposal and Evaluation of Dynamin Resource Allocation Method Based on the Load Of VMs on IaaS(IEEE,2010),978-1-4244-8704-2/11.

12. Hien et al. ,'Automatic virtual resource management for service hosting platforms, cloud'09,pp 1-8.

13. AndrzejKochut et al. : Desktop Workload Study with Implications for Desktop Cloud Resource Optimization,978-1-4244-6534-7/10 2010 IEEE.

14.Tram Truong Huu& John Montagnat: Virtual Resource Allocations distribution on a cloud infrastructure (IEEE, 2010), pp.612-617.

15.P. Mell, T. Grance.(2011), "The NIST Definition of Cloud Computing", NIST Special Publication 800-145, Department of Commerce, USA, pages: 2-3 .

16.M. A. Vouk.(2008), "Cloud Computing - Issues, Research and Implementations", Journal of Computing and Information Technology - CIT 16(4),)pages: 235-236 .

17.Tram Truong Huu & John Montagnat.(2010), "Virtual Resource Allocations distribution on a cloud infrastructure" ,IEEE,pp.612-617.