Job Scheduling Algorithm For Computational Grid in Grid Computing Environment

A Dissertation Proposal submitted by

Seema Kumari

to

Lovely School of Computer Science Engineering

In partial fulfilment of the Requirement for the

Award of the Degree

Master of Technology in CSE

Under the guidance of

Mr. Gulshan Kumar (Assistant Professor)

UID: 16865

(May 2015)

LOVELY PROFESSIONAL UNIVERSITY

School of: CSE

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the Student: Sonu Kumari

Registration No: 11210463

Batch: 2012

Roll No: B16

Session: 5th

Parent Section: K2213

Details of Supervisor:

Name: Gulshan Kumar

Designation: Asst. Professor

UID: 18265

Qualification: M. Tech

Research Experience: 3 years

SPECIALIZATION AREA: Grid Computing (pick from list of provided specialization areas by DAA)

PROPOSED TOPICS

1. Job Scheduling using hybrid algorithm in Grid Computing

2. Design and Analysis of Job Scheduling algorithm in Grid Computing to enhance the Scheduling process

3. Developing of Improved fault tolerance algorithm in order to meet QoS in Grid Computing

Signature of Supervisor

PAC Remarks:

Topic 2 is approved

APPROVAL OF PAC CHAIRPERSON: Signature: Date:

Supervisor should finally encircle one topic out of three proposed topics and put up for approval before Project Approval

CERTIFICATE

This is to certify that Seema Kumari has completed M.Tech Dissertation Titled **Job Scheduling Algorithm for Computational Grid in Grid Computing Environment** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the Dissertation proposal has ever been submitted for any other degree or diploma.

The Dissertation proposal is fit for the submission and the partial fulfillment of the conditions for the award of M.Tech Computer Science & Engg.

Date: 8 June 2015

Signature of Advisor

Name: Gulshan Kumar

UID: 16865

ABSTRACT

On job scheduling problem in grid computing there are a lot of research has been done but still further analysis and research required to improve the performance of job scheduling algorithm in computational grid. In this research we will deal with algorithm to schedule the number of jobs and to check the implementation by running number of jobs on various nodes. Then we will compare the proposed work with previous one in terms of such performance parameters like average turnaround time, waiting time, resource utilization ,cost factor, performance and complexity and number of executed jobs so that the resources will be allocated effectively and efficiently. It is research that the resource allocation model can proficiently and efficiently allocate workloads to proper resources. In this thesis we have mainly concentrate on the improving computational performance in terms of different parameters like as average waiting time and total computation time. A Cloudsim toolkit is used. for scheduling in this research work. In the proposed work of job scheduling, we have developed a unique and novel algorithm making use of hybrid as well as parallel approach of handling and processing the jobs in sequence. As in the research problem formulation, the Grid is considered as a complete infrastructure for the sharing of resource. A lot of scheduling algorithms have been developed and evaluated for Grid scheduling. In the previous work mentioned in the base literature, the analysis is done on the already developed algorithms. We have researched and implemented the new scheduling algorithms that named as IHSAG for the utilization of all the processors in a Grid computing infrastructure.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my mentor Mr. Gulshan Kumar for this excellent guidance, caring and providing me with an excellent atmosphere for completing my dissertation work. I am thankful for his aspiring guidance and friendly advice the dissertation work. I am sincerely grateful to him for sharing his truthful and brighter views on a number of issues related to the research**.**

Thank You,

Seema Kumari

DECLARATION

I hereby declare that the dissertation proposal by Seema Kumari entitled, **Job Scheduling Algorithm for Computational Grid in Grid Computing Environment** process is submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date: 8 June 2015 Investigator

Regn.No. 41200443

Contents

LIST OF FIGURES

Chapter 1

INTRODUCTION

Grid computing is a type of distributed system that is used multiple resources that are located at different locations for sharing information, storage data. Grid computing is a group of methods and techniques that are used on the multiple servers for execute the task easily and quickly.The main purpose of Grid computing is to distribute the computational power, storage data and network resource to solve the particular problem across dynamic and geographically organizations. Grid is a dispersed system that enables the distribution, selection, and geologically dispersed resources enthusiastically at runtime i.e depending on their usability, ability, execution and quality of service according to the user's specification. The capability of current computer and processor networks has amplified exponentially as compared with customary mainframes. This enhancement in their performance, in general terms wastage of computational grid resources because mostly the central processing unit sits free. Grid utilizes this free central processing unit cycles to processing the computation when it is requested by the grid applicant. Grid computing makes an optimized utilization of the resources the CPU cycles otherwise would have wasted. Grid computing is useful for reducing the computing cost, waiting time and utilization of the resources of the jobs that are processed in the job queue. The problem in the grid computing occurs such as security, complexity and resource scheduling. Nowadays Grid computing is a latest IT technologies that are used for lower cost enterprise information system. In grid infrastructure group of hardware and software components such as memory, processors, Virtual machine, networks, application server, and storage data management are used for need of the business. Energy conservation is also a major concern in grid computing systems because it can bring several important benefits such as reducing operating costs, increasing system reliability, and prompting environmental protection. According to this the users can get greater computation resource and can process their large-scale computational issues thus solving a complex problem to the computational level of a supercomputer. A scheduler plays a very important role in grid computing. Scheduler is helpful for managing the jobs and available resources in grid computing. The main motive of the scheduler should be maximize utilization of the resources and minimize the time of the job execution. The scheduler select suitable resources for job to complete its execution.

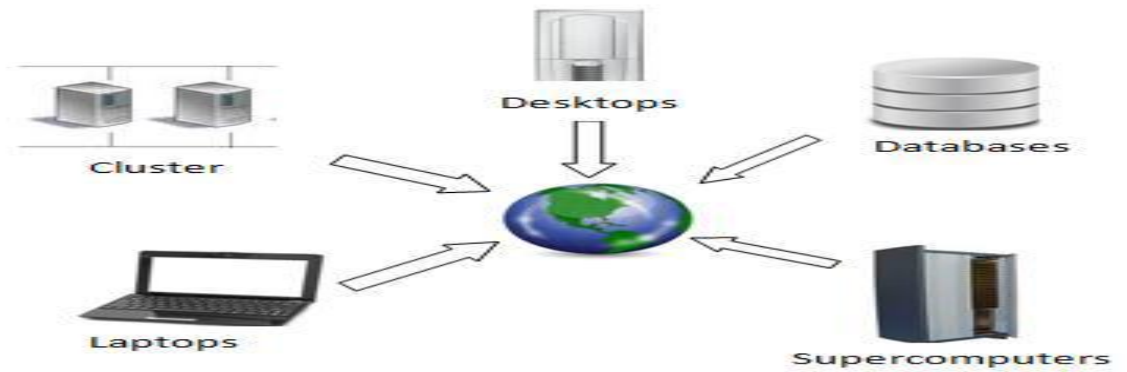

Figure 2: Grid Computing

How Grid Computing Works

A research scientist search that proteins records into the computer to analyze the information. Grid computing is a distributed computer network in which a different type of resources are used to share the information ,processing power, data storage with every resources in the grid system with the same network. Most of the Researchers, Application Programmers and Engineers are still working in the grid environments for implementing the standard protocol Enterprise scheduling, Workload management and optimized performance across all business processes. The grid computing theory is a special type of distributed computing heterogeneous computer with in homogeneous network share number of the resources. In this each resource is shared and it is used to turning a computer n/w into the supercomputer. For accessing the grid system to the authorized user, there is no difference than that of the local resource machine. For the each authorized user would have access to huge storage capacity and executing power.

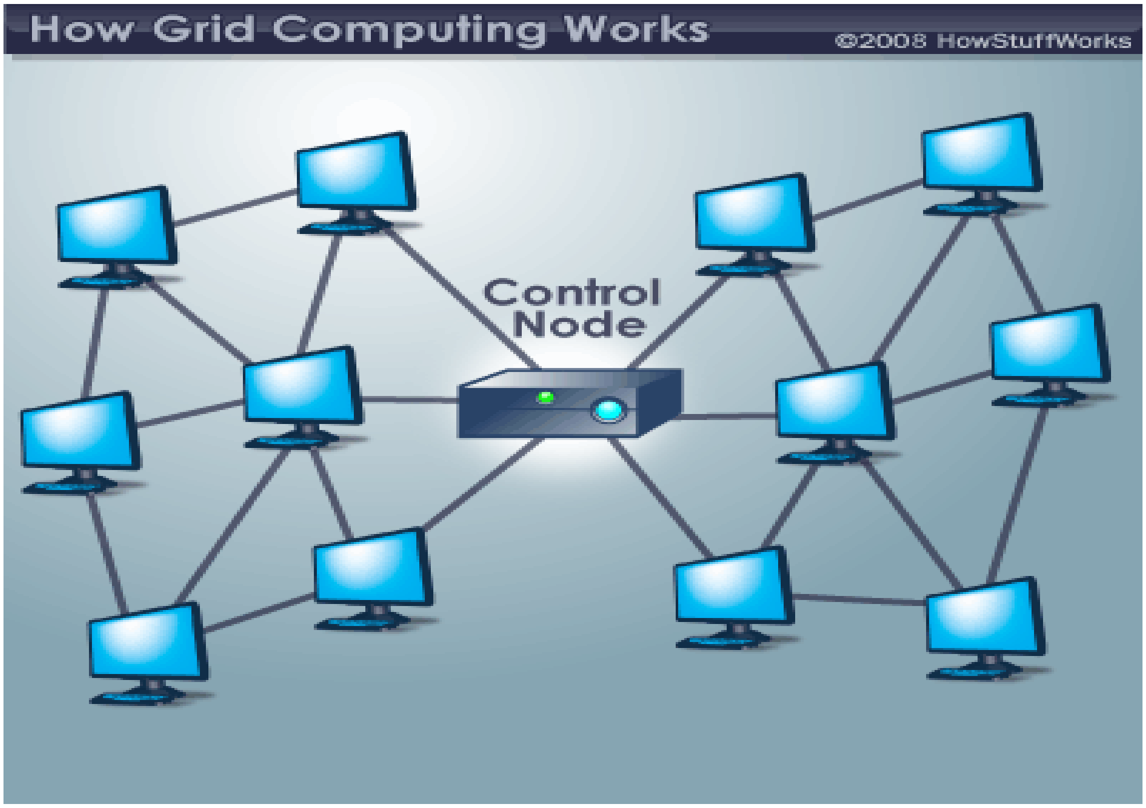

Figure 2: How Grid computer Works

1.2 Currently Distributed Environments

1.2.1 Cluster Computing: A cluster computer is a type of computing that is a group of tightly coupled computers works team-wise so that it can be viewed as a single computer.

1.2.2 Grid Computing: Grid computing is a pool of computational no. of resources, resources are different type of computers that is interconnected with a network to solve computational problem.

1.2.3 Cloud Computing: Cloud computing is a form of computing because of this storing and access your knowledge on the web instead of client's hard disk.

1.3 Grid Architecture

To understand the grid computing we have to understand the grid layers and architecture. In architecture user wants to do some kind of application for processing. The grid user send the application for processing to the grid but user can't interact with grid system. So this application is send to the grid broker and broker goes to GIS (Grid Information System) i.e a kind of directory information of all the resources .Here grid broker knows about the resources and select the suitable resources for the processing of application. Grid broker send all the data to the resources, resources process the data and send back result to the broker. The broker send back result to the grid user

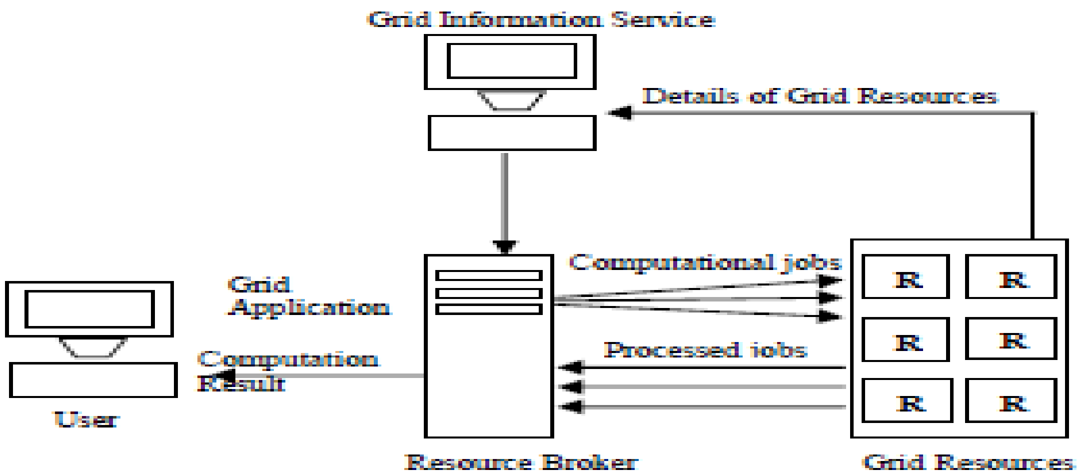

Figure 3: Grid Computing Architecture

1.3.1 Grid scheduling are mainly categorized on the basis of two types:

Static scheduling: In this scheduling process the jobs for execution are submitted to the different type of the available resources before the execution and continue the execution without interruption for these resources.

Dynamic Scheduling: In this scheduling rescheduling is allowed and executing no. of jobs can be transferred according to their dynamic status about the workload of the resources.

1.3.2 Scheduling Fundamentals:

The Scheduling fundamentals is a procedure that is used to manage the execution of the inter-dependent tasks on heterogeneous resources. It launch the assigning best resources to workflow the tasks in order that the execution may also be completed to fulfil target objective services that's pointed out through grid users. When an activity should start and end, depending on its (1) duration (2) Processing-time (3) Start-time (4) Finish-time (5) Expected execution time

1.4 JOB SCHEDULING

Scheduling is the criteria where some tasks are assigned for the processing to the resources. The assigned resources to the different type of the various tasks is known as the job scheduling. It maintains the priority queue, waiting time of the jobs for the CPU time and it is useful to decide the which job is select from the queue and how much time take to assigned to it ,then all the jobs are processed completed in a timely manner. The main target of the scheduling is that it achieves the maximum possible throughput and the application that is matched needed with the available resources in grid computing. The resource is also a digital computation factor that is matches to the threads in grid. A scheduler makes a possible multi task as a technique for the processing a various processing task at the same time on a single machine. The purpose of the scheduler is that to maximize the throughput, reduce the response time and workload of every approach. Scheduling is an approach where various tasks are allocated to the available resources by the resource broker. The resources is also have a digital computation factors that is similar to the produced or develop data flows which can be turnover scheduled on hardware like as processor. Nowadays the job scheduler mostly provide a GUI platform and a single point of control for the monitoring and defining of the background processing in a distributed network of different type of computers. Growing job schedulers are required to arrange coordination of the real system environments, business activities with conventional background for IT processing, across various type of machine platforms and real time business application environments. So the scheduler is very helpful to create a schedule for a work, if the lot of works are spread across the multiple operating system, it is also helpful to create a complex or compound job chains and job dependencies.The scheduler is an approach that selects the jobs to be submitted into the procedure to the run. The working technology could function up to three distinctive scheduler forms: a long-term scheduler (also known as an admission scheduler or high-degree scheduler), a mid-term or medium-term scheduler, and a brief-time period scheduler. The names endorse the relative frequency with which their services are carried out.

Scheduling deals with the problem of deciding which of the outstanding requests is to be allocated resources. There are many different scheduling algorithms.

1.5 GRID COMPUTING SCHEDULING PROBLEM

A computational grid is a network consisting of various computers and that are loosely coupled to each other to perform grid computing. Grid computing is define the mechanism in which same problem is solved at one time by giving resources to every computer in distributed way. In other words grid computing is like a shared environment. Requirement of resources to execute problem are extra memory, hardware, applications etc. Resources are owned by various organizations that paid some charges to use it and they have authorization that who will access it or who will not. Grid computing is truly resource sharing problem. In scheduling systems, it can be seen as problem in which some useless components can be ignored like authentication, access control etc. Basic terms used are defined in grid computing are as following:

A task/ job is an activity that has to be completed in specific time called deadline.

The properties of a task can be named as name of task, execution time, starting time, memory requirement etc.

A resource is define as something that is required to carry out an operation.

1.5.1 The Grid contains a three type of the layers:

i. Resource Layer
ii. Middleware Layer
iii. Application Layer

In grid environment the application layer is a top level of the layer that is used to the job execution on the resources level. The application layer is exist where user define the jobs for processing in the grid system. The second layer is a middleware layer that is used to link the application layer and resource layer. This layer consist the understanding required to search, allocate and monitor resources for the jobs. It sends the job for processing to the resource level and when it submits the jobs at resource level it get a message from the resource level that a job is submitted or finished, it updates the message to the application layer. If job is fail at resources level a then it defines a fault handling and the middleware layer again try to allocate it for the execution in resources layer. The third layer is a resource layer that is used to hosting a data and execute the jobs. The resource layer update to the middleware layer the acknowledgments, and the message on the current state of the resources, new data factor, and completed jobs

1.6 Benefits of Grid Computing

1.6.1 Improved User Productiveness — Organizations can dramatically improve the quality and speed of the products and services they deliver, while reducing IT costs by enabling transparent collaboration and resource sharing.

1.6.2 Scalability — Grids will grow seamlessly over time, allowing a number of hundreds of thousands of processors to be built-in into one cluster. Components can be updated severally and extra assets shall be price-introduced, reducing highest one-time fees.

1.6.3 Flexibility — Grid computing provides computing strength at any place it is required most, Grids will include heterogeneous determine nodes, allowing assets to be further and removed as desires dictate

1.6.4 Better Utilization Of Resources — The Grid computing system are used to dispensed resources additional expeditiously and gives you additional usable computing and storage power. This may increasingly shortage time-to-market, permit alternation, or modify extra checking out and simulation for improved product excellent. Through using the current type of the resources, grid computing is helpful for the shield IT investments, contain the costs of the resources whereas providing extra capability.

1.7 Various Scheduling Algorithms

1.7.1 FCFS

In FIFO, when the job scheduler wants to run the process, then it will first of all pick the process that is at the top of the queue. The scheduler is not having any kind of priority. The most common method of job scheduling is FIFO (First Come First Out) in running queue. In the job scheduling, number of processes is present in this queue. Then the scheduler picks that process, which is on the first position of the queue, then executes it then took another process from the queue. If the job scheduler is a preemptive then a process can be gone in waiting state for some time.

Process	Duration	Order
P1	10	1
P2	3	2
P3	4	3

Giant Chart

P1	P2	P3	
0	10	13	17

Calculate Waiting Time of P1=0
Calculate Waiting Time of P2=10
Calculate Waiting Time of P3=13

Average Waiting Time: (0+10+13) / 3= 7.6

Advantage: FIFO development is easy to implement.

Disadvantage: The greatest drawback of FIFO is that it is not preemptive. As a result of which it is not suitable for interactive jobs. Another drawback is that a process which is running with long time will delay all jobs behind it.

1.7.2 ROUND ROBIN

Round Robin is a pre-emptive version of the FIFO. They will execute in FIFO manner but for some specific time. This time interval is caused time slice. If process does not get completed in time slice then it has to wait for another time slice. With Round Robin scheduling, there is improvement in the system time that is used for context switching.

Advantage: It is easy to implement and each process gets fare time for processing.

Disadvantage: highly interactive processes will get scheduled no more frequently than CPU- processed.

1.7.3 SJF Scheduling (Shortest Job First)

In this scheduling the scheduler first indexing the processes with the least approximates executing time remaining to be next in the queue. This needs promoted information required time to complete a process. It is a best and suitable approach to decrease the waiting time.

This scheduling is very difficult to implement for process of the scheduling. Processor should know in advance how much time process will take.

Processors	Arrival Time	Burst Time
P1	0	4
P2	1	8
P3	2	5
P4	3	9
P5	4	3

Giant Chart

P5	P1	P3	P2	P4

0 3 7 12 20 29

Average Waiting Time: 0+3+7+12+20/5=8.4

1.7.4 Multi-Level Queue Scheduling

This scheduling is used for situations in which processes are easily divided into different groups. For example, a common division is made between foreground (interactive) processes and background (batch) processes. These two types of processes have different response-time requirements and so may have different scheduling needs. It is very useful for shared memory problems. In this Scheduling a multiple queues are used for the processors. Each and every queue can have its own scheduling algorithms**.** In this scheduling The Priorities are allocated to each and every queue.

Figure 4: Multi-level queue Scheduling

1.7.5 EDF (Earliest Deadline First) Scheduling: The earliest deadline first scheduling algorithm simply schedules all the jobs according to the earliest deadline first. In EDF Scheduling Algorithm the jobs which have assigned highest priority with minimum deadline. When one task is complete successfully according to its deadline then the new jobs is fetched in the ready queue which have minimum value of the deadline. But this scheduling algorithm is not suitable for that jobs which have same deadline then it selects the job randomly. It is a dynamic priority scheduling algorithm. In this algorithm that job is the first preference which has a highest priority with the earliest deadline. But in case of EDF to compute the response time is very difficult.

1.8 Difference: Grid Computing Vs. Cloud Computing

Cloud is a development to the object programming language and meaning of the cloud is Internet. For the client, it is a receiving the outputs for different type of the inputs. Cloud computing is based on the virtualized concept or virtual resources that is installed on the server of the clusters. In cloud computing it is also known as the SPI model. SPI model are the services that is available on the cloud. Cloud computing terminate the complexity and costs, configuring, managing the hardware and software that is useful for make a needed to build and establish the applications. This establish application are service over the web browser.

Grid computing is a type of distributed computing system where a multiple computer are interconnected with a network to solve computational problem very quickly and easily. In grid computing, user submits the job for execution with the help of resource broker. Grid is process the capability of different type of computing system for executing the jobs. Then the jobs are divided into different sub-jobs, each resource on a grid system is submit a job. When the all the sub-jobs are processed completely then it send the results to the broker that is used to manage the jobs for the resources. .

Figure 5: Cloud Vs Grid Computing

1.9 Simulator for Grid Computing

1.9.1 CloudSim Toolkit - Cloudsim is GUI based software that is developed in cloud laboratory at the department of the Computer Science and Software Engineering.in the Melbourne University.

1.9.2Architecture

The CloudSim is useful modeling and simulation of cloud environments together with committed management interfaces for memory, storage, bandwidth and VMs. It also provisions hosts to VMs, application execution management and dynamic system state monitoring. A cloud service provider can implement customized strategies at this layer to Cloudlet: This component is handled by the scheduling policy that is implement by datacenter broker component. In the cloudsim the cloudlet is used to set the user location that generate the application and datacenter location. In this configuration various parameter is used such as no. of request, no. of vms, no. of processors and various type of other parameters that is used for simulation result and that parameters shown in the graphical form. Using these parameters result is based on cost, efficiency, performance, awaiting time, turnaround time, response time, Process time etc.

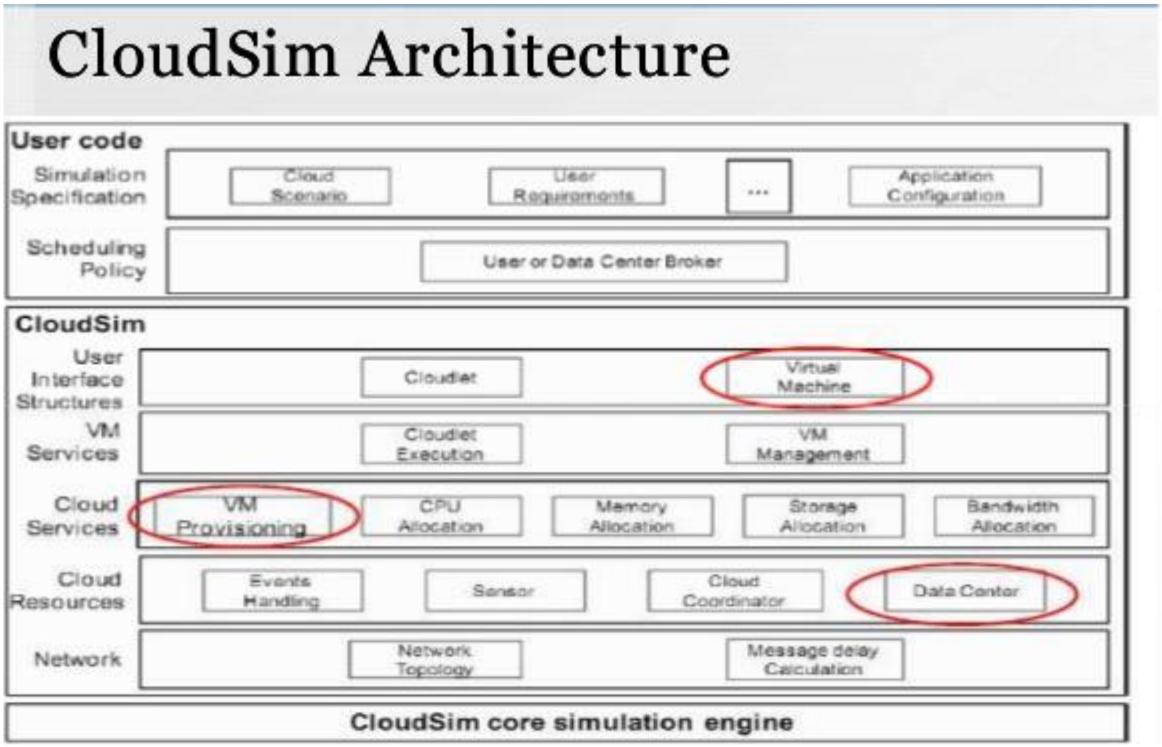

Figure 6: CloudSim Architecture

The user code layer exposes basic entities such as the number of machines, their specifications, etc, as well as applications, VMs, number of users, application types and scheduling.

1.9.3Components:

Regions: This component is a graphical regions in which resources are allocated to the their client.

Datacenter: It is component of the hosts and it is used to manage the virtual machine. It is useful for accept the request for virtual machine from the datacenter broker.

Datacenter Broker: This component is helpful modify the two type of mechanism i. e submit the virtual machine request to the datacenter and assigning the tasks to the virtual machine.

Host: This component executes the action of virtual machine creation and destruction and it is also used for processing the virtual machine.

VM: This component is used to execute the application on the machine i.e called as Vms which is lies on the physical machine.

VmLoadBalancer: This component is used to assign the load on the different type of data centers that is used in the cloud analyst. The load is assign according to the user request.

Internet Characteristics: This component includes various parameters such as amount of traffic, bandwidth and latency that is used between the regions in cloud analyst.

Cloudlet: This component is handled by the scheduling policy that is implement by datacenter broker component.

1.9.4 Cloudsim Feature

It is support and helpful for Modeling and simulating large scale data centers
It is used for Modeling and simulating virtualized server hosts
It is support for Modeling and simulating energy-aware computational resources
It I support for Modeling and simulating federated clouds
It is used for User-defined policies for allocation of hosts to virtual machines

Chapter 2

REVIEW OF LITERATURE

To propose and defend the research work, a number of research papers are analyzed. Following are the excerpts from the different research work performed by number of academicians and researchers.

Arora, S., Singh, S., Singh, S., & Sharma (2014) In this algorithm each firefly have same parameter, the value of the parameter is changes from iteration to iteration and in this algorithm there are few of the fireflies are associated with in the mutation, not all. This algorithm analyze the search space by adding attribute to less brighter firefly from more brighter firefly .This algorithm proposed the movement of every firefly using different type of the probability for each firefly implement mutation on particular firefly according to their mutation probability and after then perform the mutation on the every fireflies to better explore search space. This proposed mutated firefly algorithm performance is much more efficient than that of the firefly algorithm.[5]

Arsuaga-Rios & Vega-Rodriguez (2014) Job scheduling is critical problem and challenge task for execution time and cost. In this research we represent a cost optimization based on the multi-objective Brain Storm algorithm and this algorithm is based on a novel algorithm i.e called as a Brain Storm Optimization (BSO) which is inspired human behaviour. The proposed algorithm minimize the cost and processing time. MOBSA includes a multi type of properties to the BSO for the development of more than one objective. This is used for execution time and cost of the grid scheduling problem in grid computing.[6]

Dr. Sanjay. K. Dwivedi and Ritesh Gupta (2014) In MFQ algorithm provides the preemptive and non-preemptive algorithms can be used together in Multi-Level queue scheduling and it divides the ready queue into a number of queues with different scheduling policies. The proposed MFQ algorithm in multilevel scheduling ,with the help of simulator for various group of jobs and scheduling policies, In this algorithm preferred RR for the fair use of CPU in 1st queue and SJFRR in next queues (which we called as SJFRR) to reduce the average waiting time and turned around time.[10]

Nithya, G., and R. M. S. Engels (2014) Multi-agent broking approach proposed algorithm is based on firefly algorithm. The main work of the firefly algorithm is helpful to reduce the make span time. Now in this algorithm addition jump option was introduced to jump to new situation to arrive high probability i,e called as a jumper firefly and it is more efficient than that of other heuristic algorithm and firefly algorithm.[19]

Prajapati, Harshadkumar B., and Vipul (2014) The author uses multiple machine model to simulate the grid computing environment and clustering of machines have been done at University of Calgary. The experimental data has been collected for seven days and examined thoroughly the job scheduling in computing cluster. A single site centralized algorithm for scheduling has been simulated where first all the jobs enter in a single que and 13 then machines are assigned to them based on different criterions. It is concluded that the by thorough analysis with a job arrival rate of 3.125 jobs per half an hour with Poisson distribution a mean of 496.3 minutes for running time is obtained following a negative exponential distribution. A GPSS-H system is simulated in which most of the jobs requires less than four processors and a constrained of 4 processor per job is applied.[20]

Dipti Sharma Mr. Pradeep Mittal (2013) Grid computing is studied in this paper and it is proposed that the resources are shared in Grid computing from various locations which can be heterogeneous or homogeneous in nature. The variety of applications areas are covered in Grid computing ranging from science to engineering areas including medical field too. The major challenge is to schedule the job for various processes and an algorithm for the job scheduling is proposed which can allocate the resources in an efficient way for improving the performance of the Grid.[11]

Manimala, R.,and P.Suresh (2013) In this technique the scheduler gets a job which is allocated by a client, it will send a request to the Information Service to attain the compulsory information like the idleness of the Central processing unit percentage of each resource, clusters average load and system's average load. After the section of the cluster, it will evaluate the average load of the selected cluster with the average load of the organization. The load average of the cluster is defined by the average load of each resource in particular cluster. Weighted sum of the square method for the measure of the load is used.[16]

Abba, Haruna Ahmed (2012) In Prioritized Deadline based scheduling algorithm most of the researchers have not completely considered the deadline perspective for job processing But PDSA generate a new concept that meet the deadline as indicated by users and users are extremely to processed the no. of jobs in appropriate way. In this algorithm input is determined by its deadline, arrival time and burst time.. After then calculate the time delay in increasing order for per job and then select the job for the processing on the basis of least time delay. The performance of the PDSA as comparison to the previous algorithm i.e EDF scheduling algorithm and RR scheduling algorithm is more efficient. and greatest tardiness.[1]

Run-ze, Ming-shan , Xiang (2012) In this researcher proposed technique for task scheduling in grid computing, which is an enhanced task scheduling algorithm using triangle module operator to integrate Quality of Service of a user with system's based performance of the computations? The result are simulated on Grid Simulator, and the results powerfully show the proposed algorithm performs better, resource load balancing and Quality of Service effectiveness than minimum heuristic which does not regarded as good Quality of service.[23]

Wang, Jing, Gongqing, Zhang, and Xuegang (2012) This proposed research is a type of computational method that modifies a problem with the help of iterations trying to improve a performance of the individual. This heuristic formulates the fitness of each character and based on their fitness character are added in the list of optimization .The new population is examined for next hop or iteration of algorithm. When acceptable fitness level is reached then algorithm stops.[27]

Adil Yousif, Abdul Hanan Abdullah, Sulaiman" MohdNor and Adil Ali Abdelziz., (2011) The Firefly algorithm proposed a meta heuristic algorithm which is focused for find the global optimal solution to complete the allocated job within minimum make span and flow time. It propagates based on intensity level which have higher knowledge that more glow. All firefly to move towards each other on basis on intensity level and the amount of movement depends upon attractiveness parameter which is further depend on distance between them. It focus to the efficiently utilization of the processor in the grid. In firefly algorithm first of all the No. of available resources ,no. of jobs allocated and parameters are identified .after then it found the brightness of each firefly using fitness function then calculate the distance and light intensity is updated. Now the current global best is recognized and fireflies are ranked. [3]

Amudha, T., & Dhivyaprabha (2011) this proposed algorithm is based on the quality of service priority scheduling. In this algorithm priority value is taken as a parameter of the tasks. When jobs are assigned for the execution then the jobs are grouped into the sets and sets are divided into two different groups. In one group set the value of higher priority jobs and second group set the value of lower priority jobs. Now the jobs are divided into different categories by the Grid Scheduler. Now that job is assigned first for scheduling which have higher priority value. The Grid system are divided into two groups i.e cluster 1(high processing system) and cluster2 (high and low processing system). The scheduler assigned the jobs for processing to cluster1 which have high complexity and cluster 2 which have low complexity jobs. First of all high priority jobs are assigned for processing after then low priority jobs are assigned and this procedure will keep on until all the jobs are not executed.[4]

Gomathi, S., & Manimegalai (2011) Adaptive algorithm proves a genetic algorithm that is based on evolutional algorithm (G-EA).G-EA provides a algorithm for solving the scheduling problem for reducing the job execution time in grid computing-EA as compared with Round-Robin, Min-Min scheduling algorithms and perform better throughput. In adaptive scheduling the parameters that is used to make a decisions of the scheduling changed according to the last, resource status in grid computing. [12]

Xhafa, F., Kolodziej, J., Barolli, L., & Fundo (2011) In this research paper represented a hybrid algorithm i.e GA+TS which is useful for the batch scheduling in grid environment. In this research a GA is executed first which have high level hybrid approach to find a appropriate solution and flow of GA is executed first until the final result is not found and then TS will executed. Here we figure out that the hybrid algorithm using various grid scheme perform by a grid simulator. The computational final results present that the hybrid GA+TS algorithm are better than the both of GA and TS for few cases of instances of the scheduling problem in grid computing. [26]

(Sharma, 2010) In this paper the importance of Grid computing is shown and Grid computing is considered as a variant of distributed computing in which the computational power and data storage along with the other resources are shared for improved performance of the grid. Since scheduling is an NP-complete problem, an optimal solution of the scheduling is not possible to optimal solutions are obtained in terms of various objectives. The aim is to achieve the highest possible performance. An easy way of understanding the Grid computing is presented in this paper so as to facilitate new researchers with the basic concepts of Grid computing. [22]

Ku-Mahamud, K. R., & Nasir, H. J. A. (2010) The ant colony optimization algorithm is dependent on the behavior on ants moving in search of food. Each ants share their experience by releasing a chemical called as pheromone. This pheromone level evaporates moves with time and ants follow the patch in which there is maximum pheromone level. The pheromone is helpful to find the at least way and the value of pheromone of each resource is depends on their job processing time.. The higher processing time of the resource will receive a large no. of pheromone. When a resource is successfully done its job then its higher value of the pheromone will be added each time for processing of the next coming job .If the resource fails to complete its job execution then it will be penalty by adding lower pheromone value. The problem is that where the possibility of jobs are assigned to same resources having high pheromone value .The algorithm is follows as :-

i. The applicant will forward a job for execution
ii. The resource broker will accept the request sent by the applicant and find a appropriate
iii. Resource based on higher value in pheromone for the job.
iv. When job is assigned to the resource then the local pheromone update is done.

v. When a resource is successful to complete its job then the global pheromone is updated.

vi. After then the result will be sent to the client.[15]

Murugesan, G., & Chellappan (2010) It is shown that the major issues in job scheduling are the execution of workflow. To map and manage the execution of dependent jobs on resources job scheduling is applied. Suitable resources are allocated to workflow tasks in order to complete the execution under the given constraint and also achieving the objective function. It is very important to have proper scheduling strategy because it can have severe effect on the performance of the system. Various algorithms employed in literature are surveyed and explored in detail. [17]

G.Kannan and Dr.S. Thamarai Selvil (2010) In this proposed work this algorithm is combination of the two algorithm i.e Backfill algorithm and Earliest Deadline algorithm. Job Scheduling and Resource Management are the fundamental issues in Grid Computing. This algorithm is constituted with Care Resource Broker (CRB) that maintains virtualized. CRB helps to entire decisions to prioritizing the jobs and reducing the average waiting time, average turnaround time and response time of the jobs that are available in the job queue scheduling and the user's quality of service has enhance to make as large the virtualized grid environment.[13]

Raksha Sharma,Vishnu Kant Soni,Manoj Kumar Mishra (2010) This proposed work plays a very crucial role to increase the resource utilization and how much no. of the jobs is processed in grid system This concept is based on the hierarchical structure that is divides the level into three different categories i.e. named as user level, global level and cluster level. First of all the grid user allocate the jobs for processing at the first level i.e. user level, then scheduler gets the information about the resource in the grid from the GIS .When the application is ready to submit into the grid at the global level then it will search the clusters that are available in the system and then selects the cluster whose computational power is higher. If the computational power of the cluster is higher than that of enters application then the next application enter at same time .Now comparison of the total computational power of combination of first and second application with the selected cluster. It‘s value will be less then or can be equal to the computational power of selected cluster. The local scheduler works to maximize the resource utilization and reducing the processing time with in the cluster.[22]

Nithyapriya, D., & Krishnamoorthy (2010) Meta Scheduling plays a important role in the scheduling system in grid computing. In this scheduling, Meta Scheduling system is used to utilize the grid power of the jobs and resources and uses a maximum jobs with minimum available no. of resources. In this Grid system assign the jobs based on the priority of the user account and system account. After then which job has a common location add them and prioritized them based on the location. The high priority job can be allocated first for the computation with the available resources based on their computing power and it is useful to improve the performance. [18]

G. Murugesan, Dr.C.Chellappa (2010) The author addresses the problem of scheduling and focuses on resource utilization as it is the most important part which affects the grid computing performances. Since a variety of resources are available, the complication of the scheduling process increases which is the main challenge in grid system. An allocation model based on divisible load theory has been proposed in which various variety of workloads are considered with a unique originating processor. The resource allocation model which is proposed is quite novel and an economic model based on multiple load originating processors is formulated. The costs of grid users are minimized using LP techniques by solving the equations of fractional load allocation to nodes in the system. The model is shown to work quite effectively with high efficiency and provides better solution when cost and time are considered. [14]

Chang, R. S., Chang, J. S., & Lin, P. S. (2009) An ant algorithm is also known as the well-balanced job scheduling algorithm in grids. In the atmosphere, the ants have a fearsome ability to group up to find an optimum way to food resources. In this paper the ant algorithm emulate the behavior of ants. In this paper, we research a BACO techniques for job scheduling in the Grid Computing. The main role of our research are to balance the completely load system while it is trying to minimize the makespan of the jobs. Now it is comparison with the other implemented job scheduling algorithms, BACO can checkmate to the jobs according to the exploratory results. [8]

Dong, F., & Akl, S. G. (2006) The major challenges of Grid Scheduling are identified in this paper and the architecture of different components which are involved in the process is given to enable the readers with an approximate idea of scheduling process. Literature has been surveyed thoroughly and the various algorithms are compared on their merits and demerits in terms of objective functions and policies, Quality of services, strategies, limitations, performance etc. Some common issues are addressed and explored in detail.[9]

Srinivasan, S., Kettimuthu, R., Subramani, V., & Sadayappan, P. (2002) In this paper that backfilling offer valued profit in the scheduling of parallel type of the jobs, In this no concept is clear which backfilling concept is most suitable e.g. .there should be least backfilling be used or the more hostile EASY backfilling plan; there should be a first-come first-served (FCFS) priority be used. In this paper we studying a trace base simulation that is used to address these queries and extract new comprehension, the main features of backfilling scheme for job scheduling. In this paper we calculate the performance come to terms with slow and turnaround times of jobs within the various parameters that is depend on their width (processor request size), length (job duration) and precision of the user's evaluation at run time, some of the contingent trends may be discovered.[25]

Chapter 3

PRESENT WORK

3.1 Problem Formulation

There are a lot of scheduling algorithm already developed in grid computing environment. The scheduling problem becomes more challenging because of some unique characteristics belonging to Grid computing. There is need to analyze the performance of job scheduling algorithms on cloud as well as grid environment as priority, load balancing and related aspects can affect the performance of cloud and grid to a huge extent.

In this research proposal, we proposed a virtual machine provisioning to analyze reduce the consumption time, cost, performance as well as resources functionalities along with their parameters and process the more no. of jobs in comparison of the previous implemented algorithm. In previous my base paper conclude a new algorithm i.e. EDSRTF which is based on this diagram.

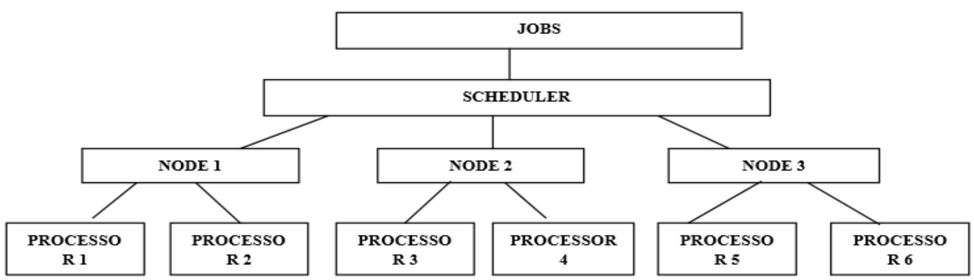

Figure 3.1 Diagram of EDSRTF

The basic solution obtained without swarm intelligence is not efficient in terms of the turnaround time and optimal results.In my research work, we have analyzed the performance of assorted algorithms based on new parameters and dimensions.

3.2 Objectives

The objectives are following:-.

i. To perform a detailed comparative study on different algorithms in grid as well as cloud infrastructure

ii. To propose the implementation of algorithms on new parameters so that the hidden aspects and attributes of the algorithms can be analyzed.
iii. To perform the implementation / simulation and analysis on algorithms on new parameters beyond the work in the base literature.
iv. To perform various scheduling algorithms in the field of grid computing.
v. To implement enhance the scheduling process in grid computing.
vi. To check implementation by running jobs on various nodes.
vii. To compare the results of proposed algorithm based on parameters i.e. execution time, resource utilization and throughput.

3.3 Research Methodology

Research is process or investigation of approaches, methods, techniques to produce new facts or improves the previous conclusions. The main focus of Research is to uncover the facts from previous searches. I proposed a new algorithm and compare the result It is found and concluded from the results that EDSRTF algorithm is having effective results in cumulative way if we consider all the parameters in investigation as a whole level.

In this research I have integrate CloudSim with Ecllipse and then start programming work. I use JAVA for deploying Scheduling algorithm and then later on I can simulate the environment using CloudSim. I executed some examples and now I have implement my own algorithm. This research work is focus on development as well as implementation of a job scheduling algorithm for effective scheduling and prioritization. By this way and the proposed technique, the effective results in terms of less cost, higher performance and less execution time is obtained. In this research work, we have analyzed the performance of assorted algorithms based on new parameters and dimensions. There is need to analyze the performance of job scheduling algorithms on cloud as well as grid environment as priority, load balancing and related aspects can affect the performance of cloud and grid to a huge extent. The work done in this manuscript includes to propose the implementation of algorithms on new parameters so that the hidden aspects and attributes of the algorithms can be analyzed. The work includes to perform the implementation / simulation and analysis on algorithms on new parameters beyond the work in the base literature.

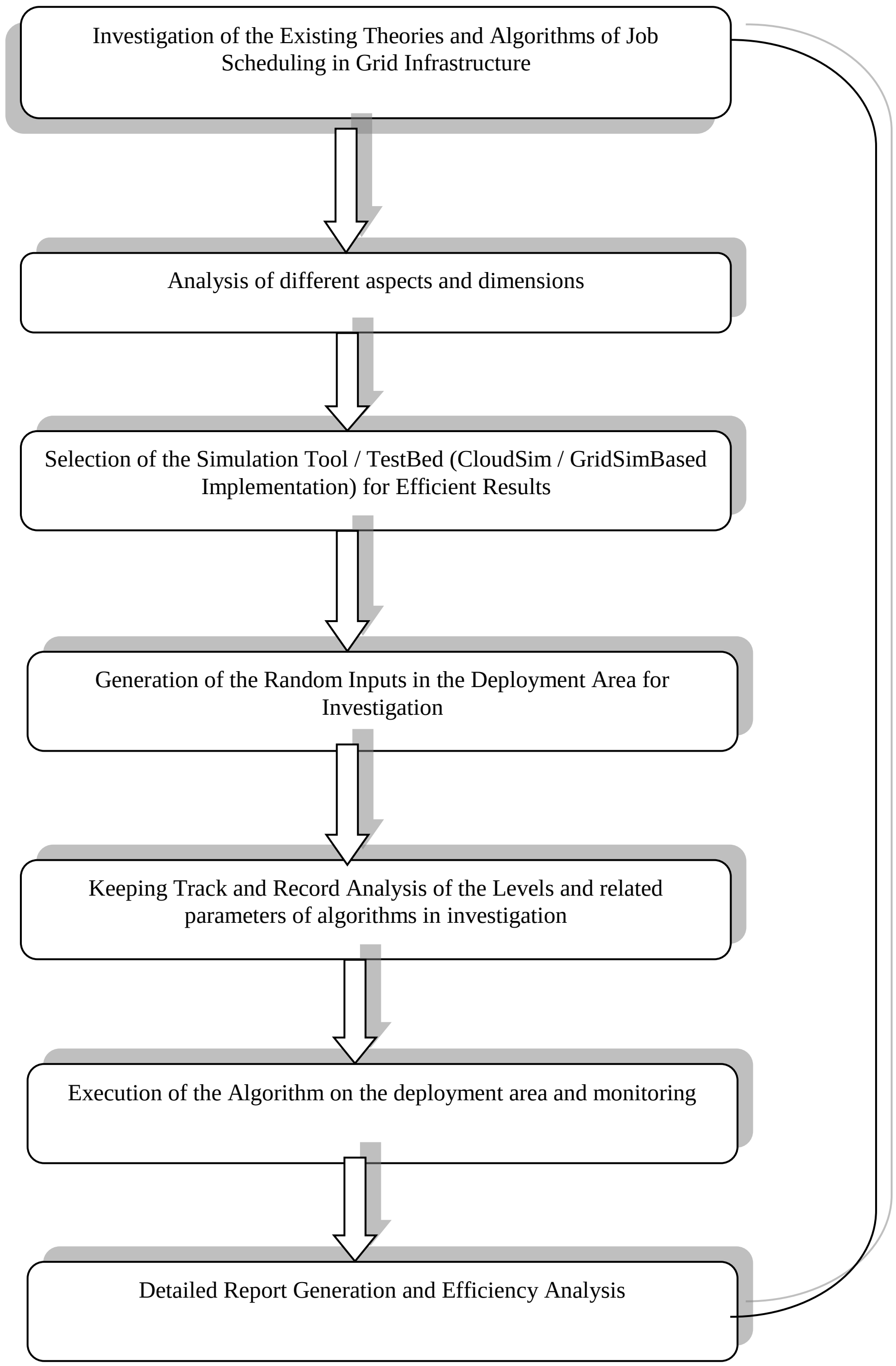

Figure 3.2: Proposed Flow of Work

3.3.1 Implementation Steps:

Step1:

Set the cloudsim configure like number of Vm's, host, san etc.

Step 2:

Each server six processors and maximum of 10 jobs can be entered at a time.

Step 3:

Considered three nodes Node 1, Node 2, Node3 and each node consists of 2 processors P1, P2, P3, P4, P5 and P6 respectively.

Step 4: implement algorithms are EDSRTF, LJF and FCFS

Step 5: Improve Hybrid Algorithm (IHSAG) and Merge Using Parallel Approach in Grid Infrastructure.

Step 6: run cloudsim

Step 7: Finally in result show the evaluation result and graph.

3.3.2 Technical implementation and Approach

To design the matrices and blocks of the computation cost as well as communication.

Assignment of priorities to tasks.

Calculate EST=0 & EFT=0 of task 1 on each processor, set process avail time for p1, p2, p3=0. Assign processor to task1 on which EFT is minimum

either p1=EFT (T1or task1) or p2=EFT (T1) or P3=EFT (T1)

Activation of the Genetic Algorithm and its aspects

For t2 to tn repeat above

Calculate if (parent of t1 executed on p1)

EST (tn) on p1=max (parent executed)

p2=max (parent completion + Communication cost), processor avail time)

EST (tn) on p3= same as above

EFT=EST + Computation cost

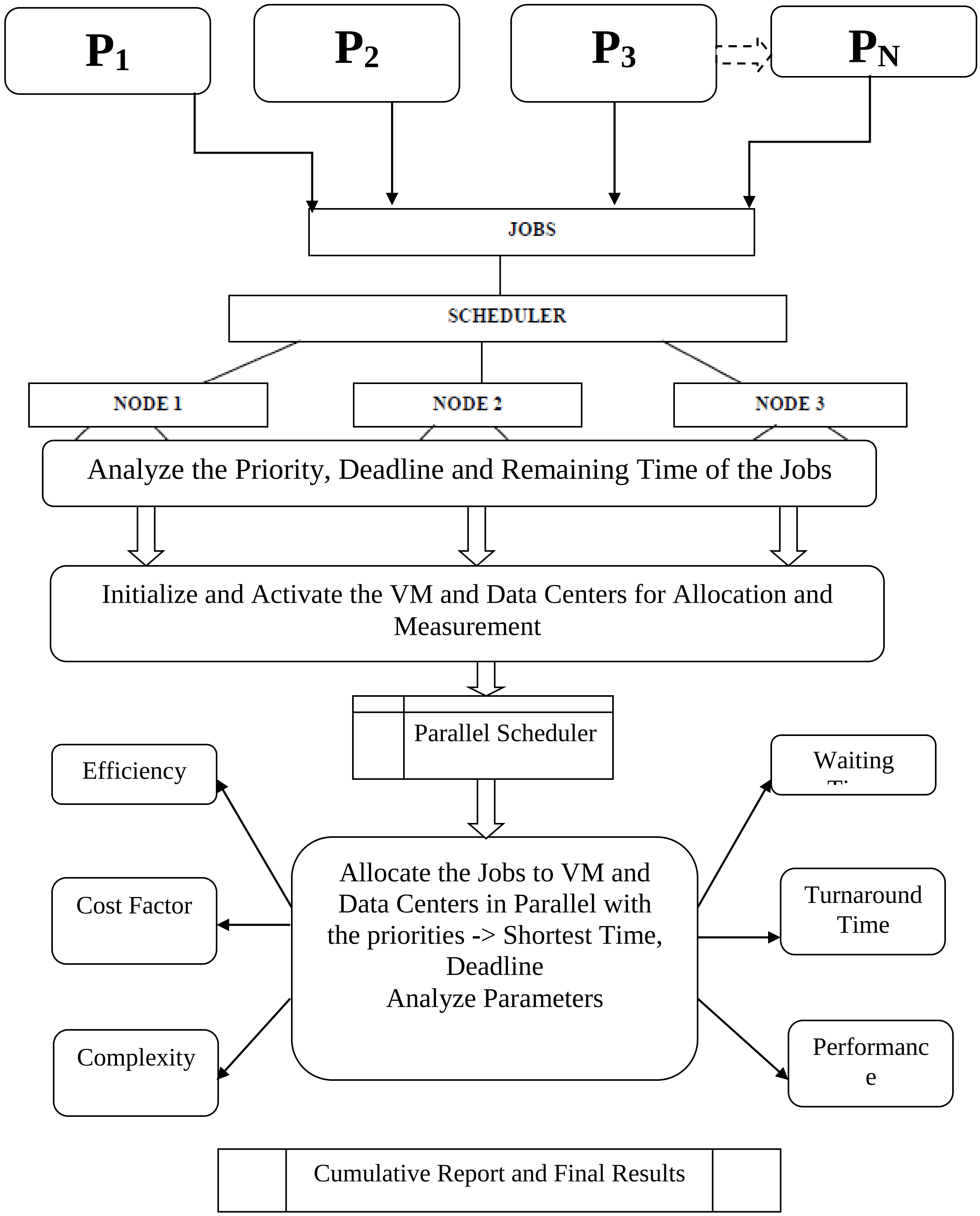

Figure 3.3 Flow Chart of Research Methodology

In the proposed work of job scheduling, we have developed a unique and novel algorithm making use of hybrid as well as parallel approach of handling and processing the jobs in sequence. As in the research problem formulation, the Grid is considered as a complete infrastructure for the sharing of resource .A lot of scheduling algorithms have been developed and evaluated for Grid Computing. In previous research work mentioned in the base literature, the analysis is done on the already developed algorithms. We have researched and implemented the scheduling new algorithms IHSAG for the utilization of all the resources in a Grid infrastructure or environment. In this research work, we propose and implement the Hybrid as well as parallel approaches of scheduling algorithm and evaluate the proposed work. The major key concept of the proposed algorithms in this paper is to execute the jobs optimally so that there is the best combination of average waiting, turnaround and efficiency and cost. The extensive and detailed performance comparison is done and presented using cloudsim and gridsim libraries to calculate the efficiency, performance, cost of scheduling algorithms. To facilitate and defend the research work, the software based simulation tool is deployed that produces a extensive and detailed simulation of the no. of various type of Grid scheduling algorithms. The output that is fetched is in the form of scheduling performance and giving effective and optimal results. In this paper we implement the different algorithm according to the base paper and improve the algorithm .In this paper we proposed a new algorithm based on different type of the parameters and find the output that is shown in graph and table form. In this research the implementation is done in Eclipse IDE with Cloud and Grid Simulators in the External JAR Libraries for compatibility with Cloud and Grid Infrastructure.

It is found and concluded from the results that EDSRTF algorithm is having effective results in cumulative way if we consider all the parameters in investigation as a whole level.

Hardware Requirements

Processor	Pentium & Above
Ram	256 Mb
Hard Disk	10 Gb

Software Requirements

Operating System	Windows 7 Or Window 8
Front End	Java

Tools Used

I. Eclipse IDE

II. CloudSim

Chapter 4

RESULT AND DISCUSSIONS

We simulated the proposed algorithm using CloudSim toolkit to verify the improvement of the proposed approach over other scheduling strategies. In this chapter, we have discussed implementation details of the simulation environment. The experimental results are analyzed based on performance criterions.

The implementation is done in Eclipse IDE with Cloud and Grid Simulators in the External JAR Libraries for compatibility with Cloud and Grid Infrastructure. It is found and concluded from the results that EDSRTF algorithm is having effective results in cumulative way if we consider all the parameters in investigation as a whole level.

Following are the screenshots and results obtained from the simulation run.

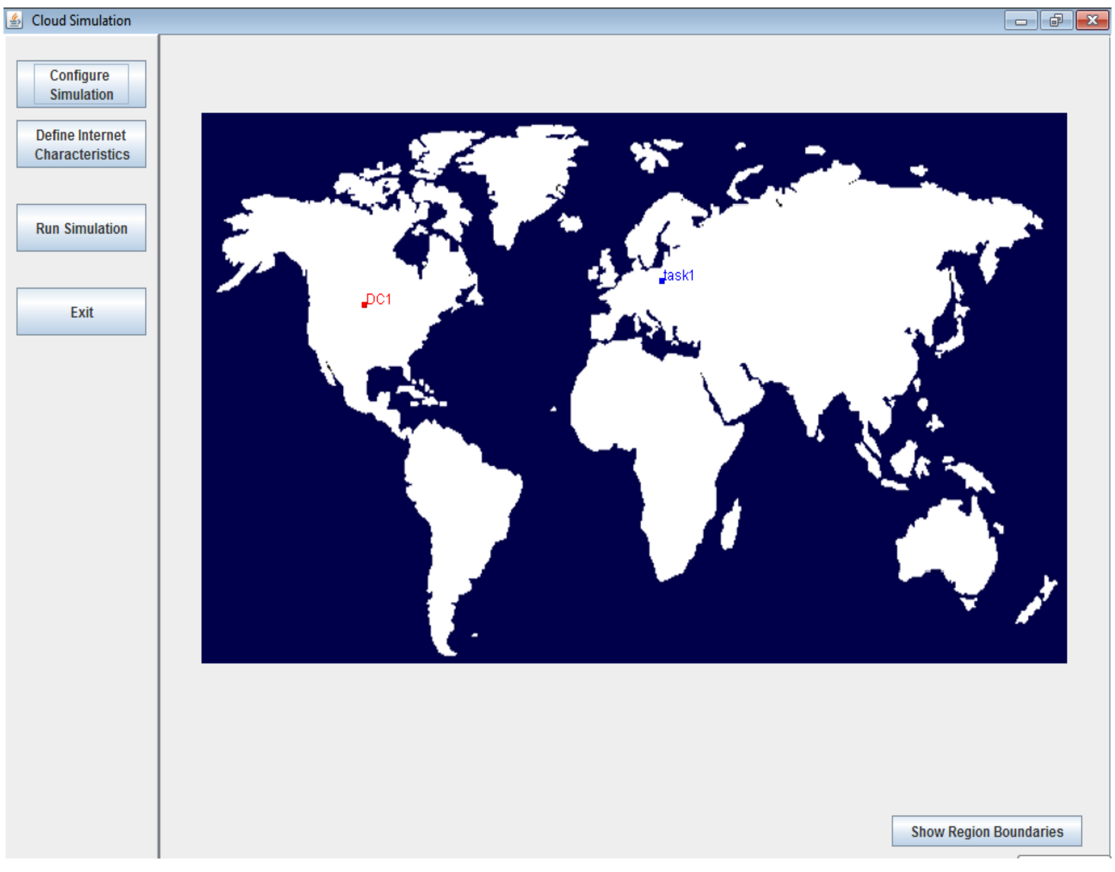

Figure 3.1: Initialize the Simulation Scenario

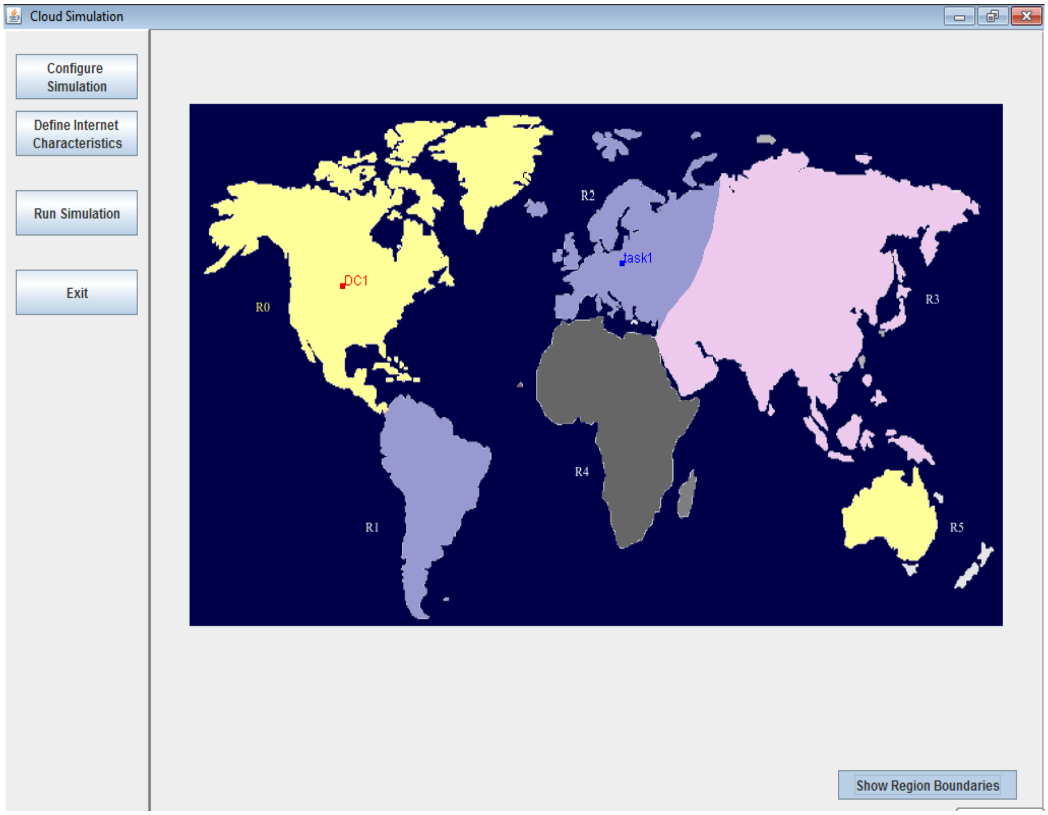

Figure 4.2: Display the Region Boundaries of the Processes in the Simulation Scenario

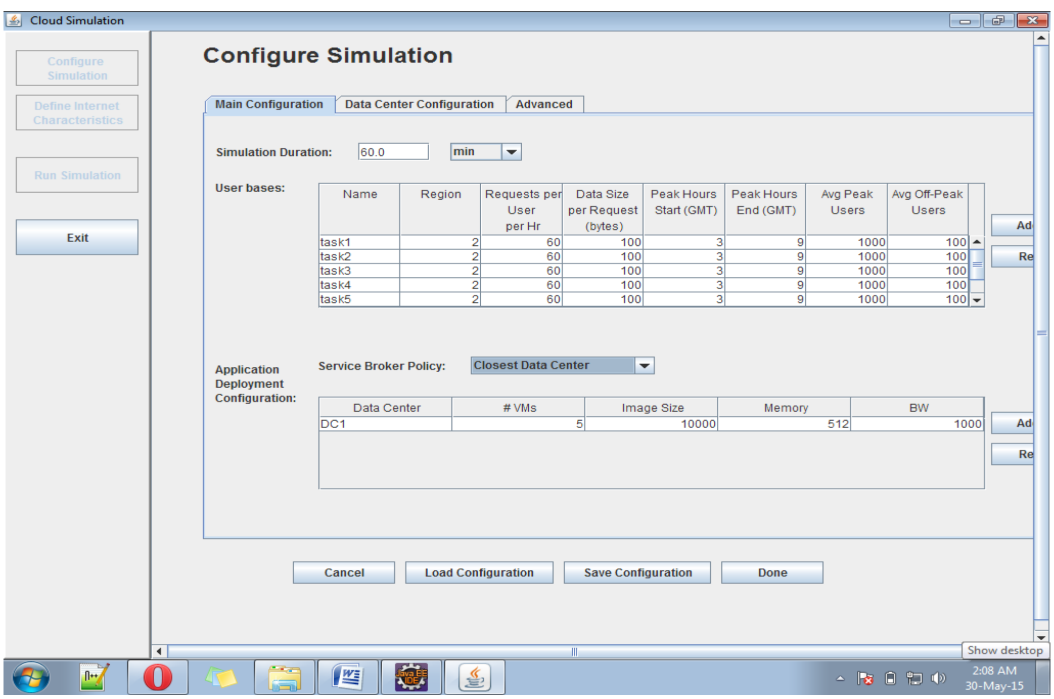

Figure 4.3: Configuration of the Simulation Attributes

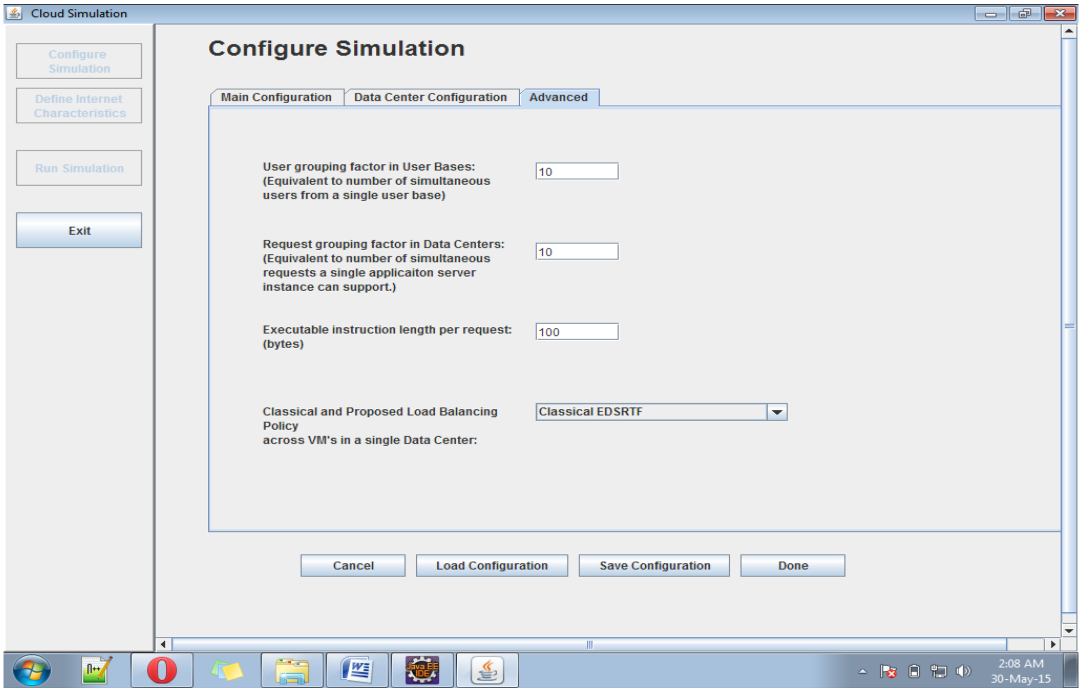

Figure 4.4: Configure Simulation with Algorithm 2

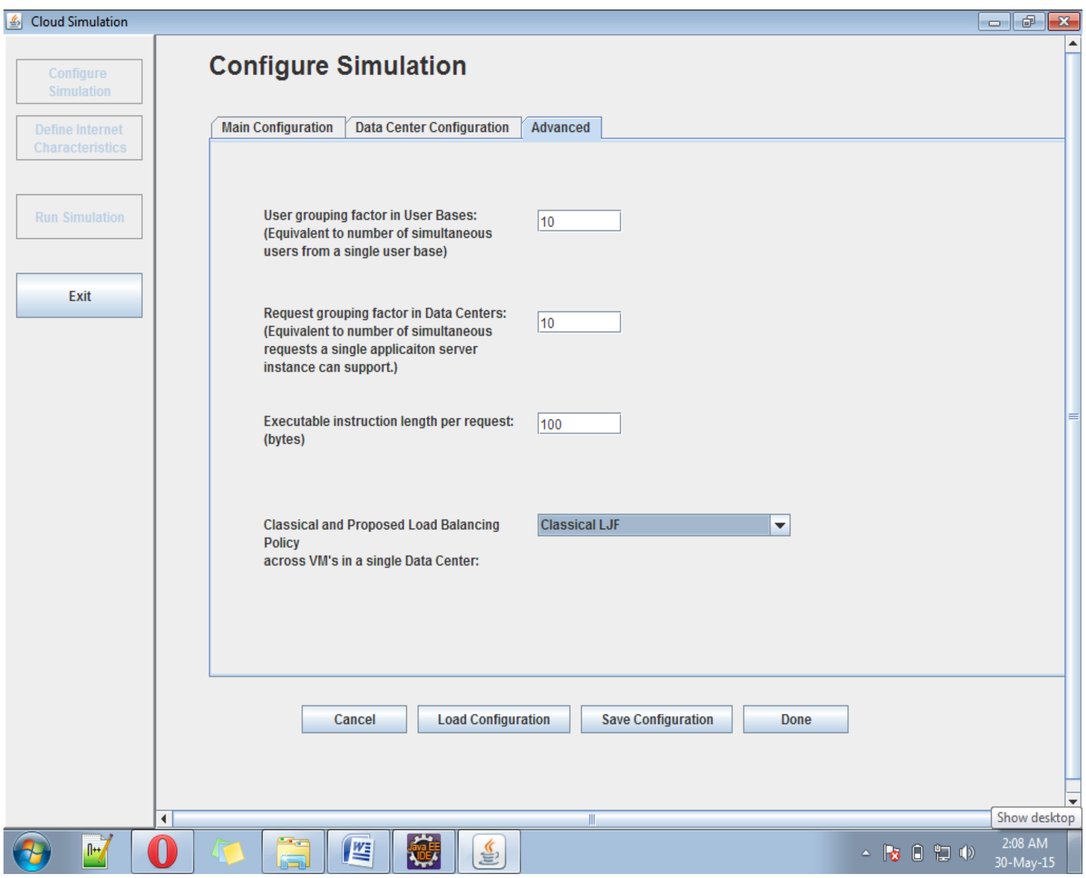

Figure 4.5: Configure Simulation with the Algorithm 2

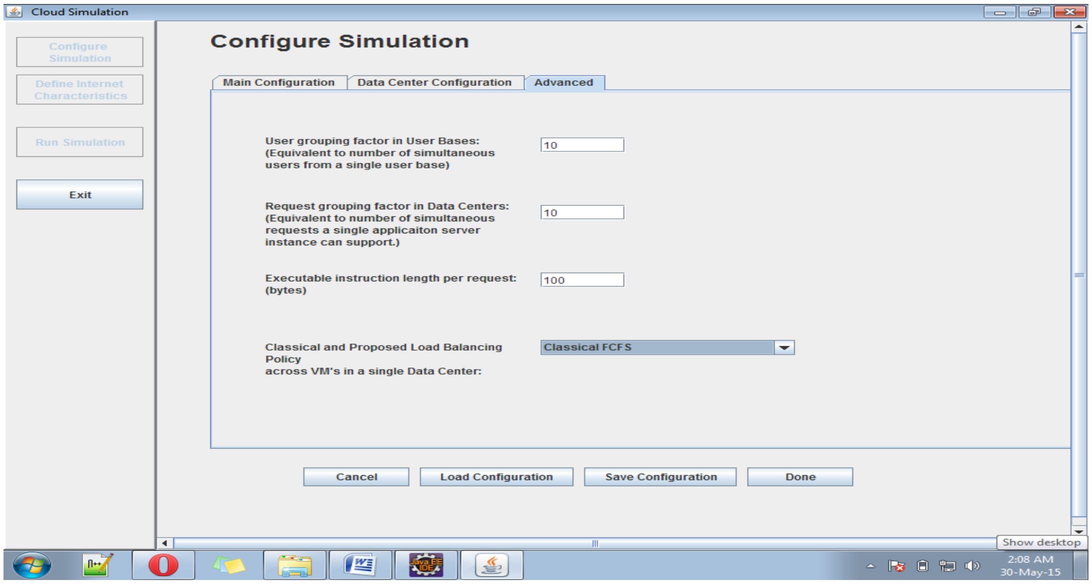

Figure 4.6 – Configure Simulation with the Algorithm 3

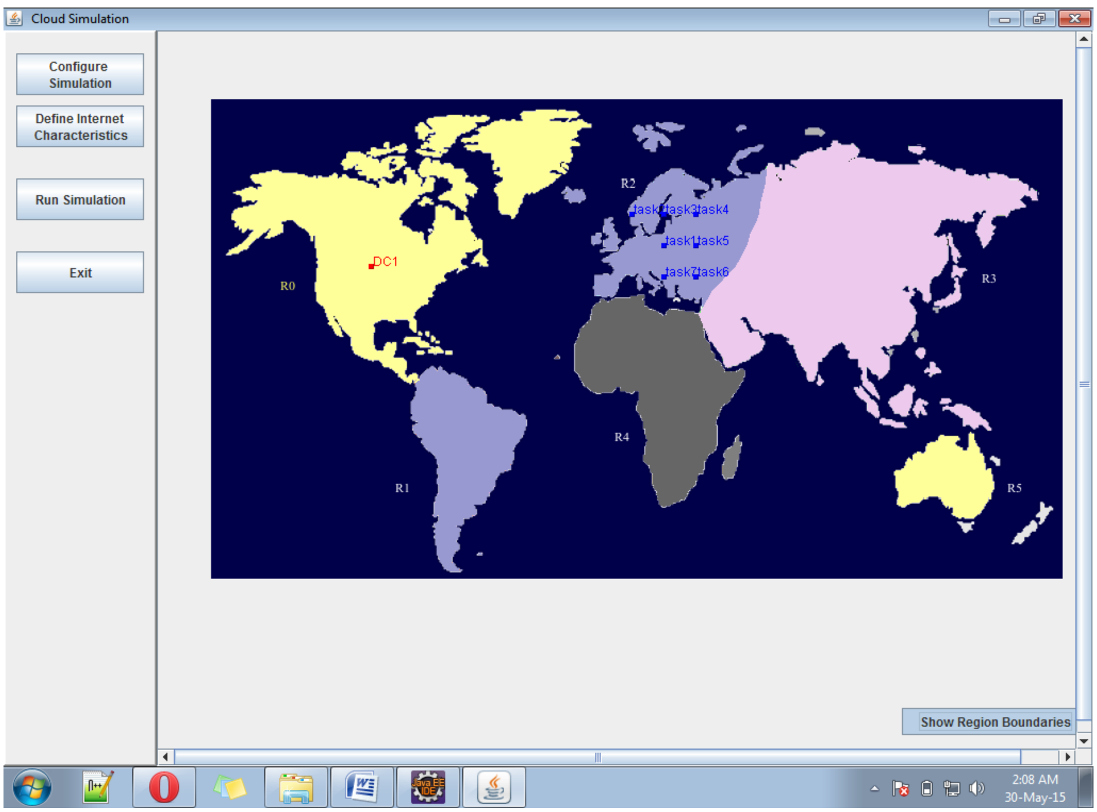

Figure 4.7 – Show Region Boundaries after setting up the Configure Simulation

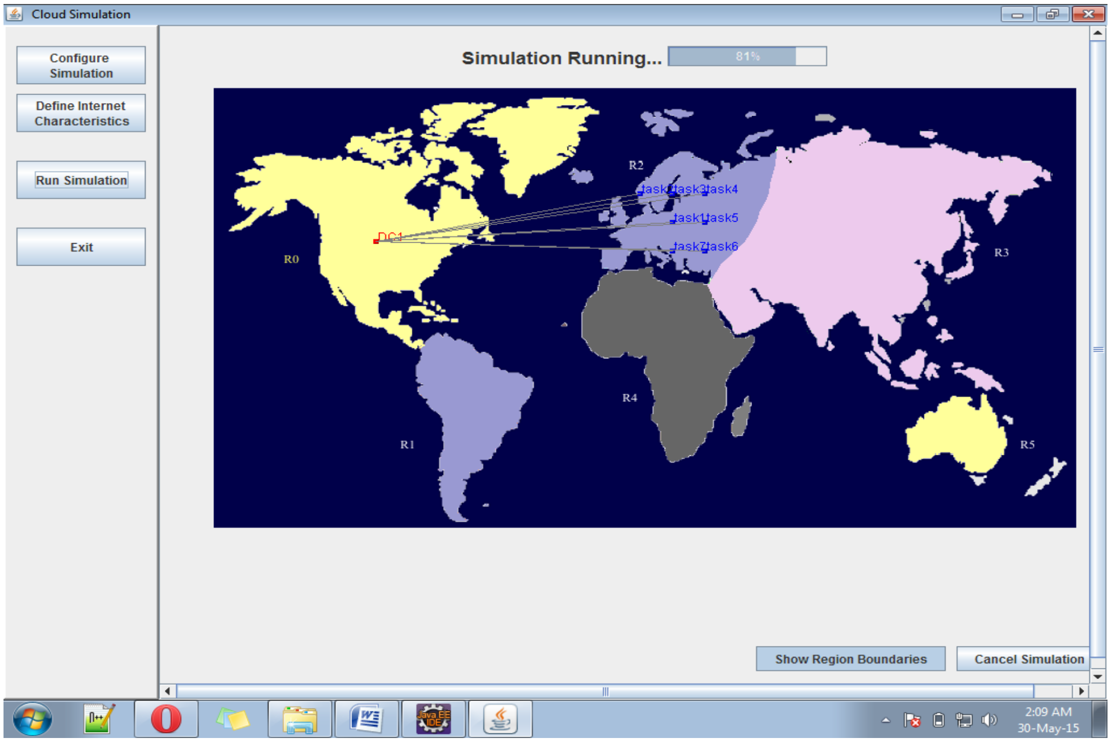

Figure 4.8 – Running the Simulation Scenario

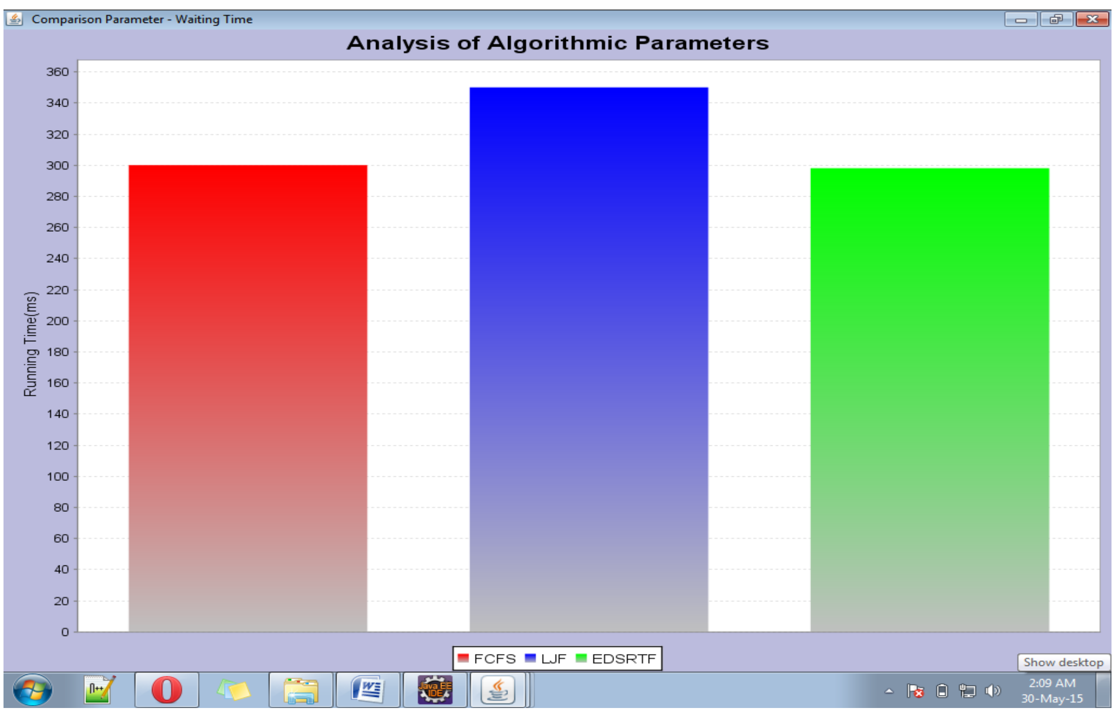

Figure 4.9 – Performance of Algorithm on the Parameter - WAITING TIME

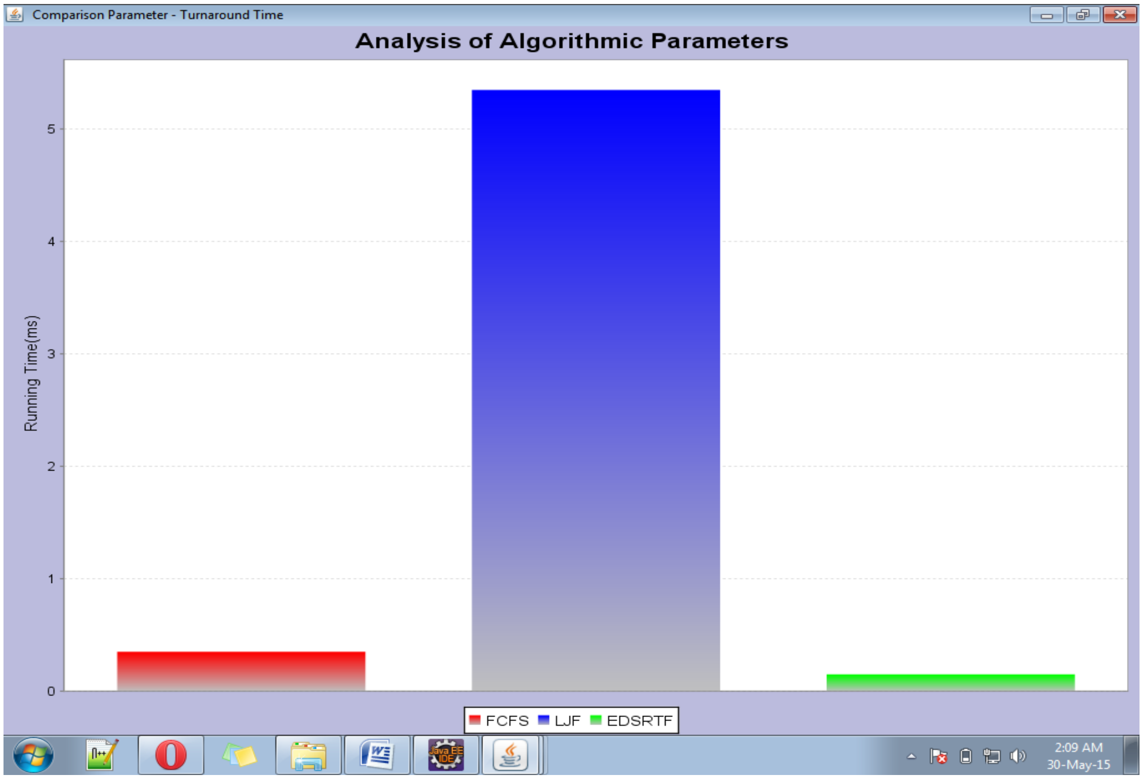

Figure 4.10 – Performance of Algorithm on the Parameter – TURNAROUND TIME

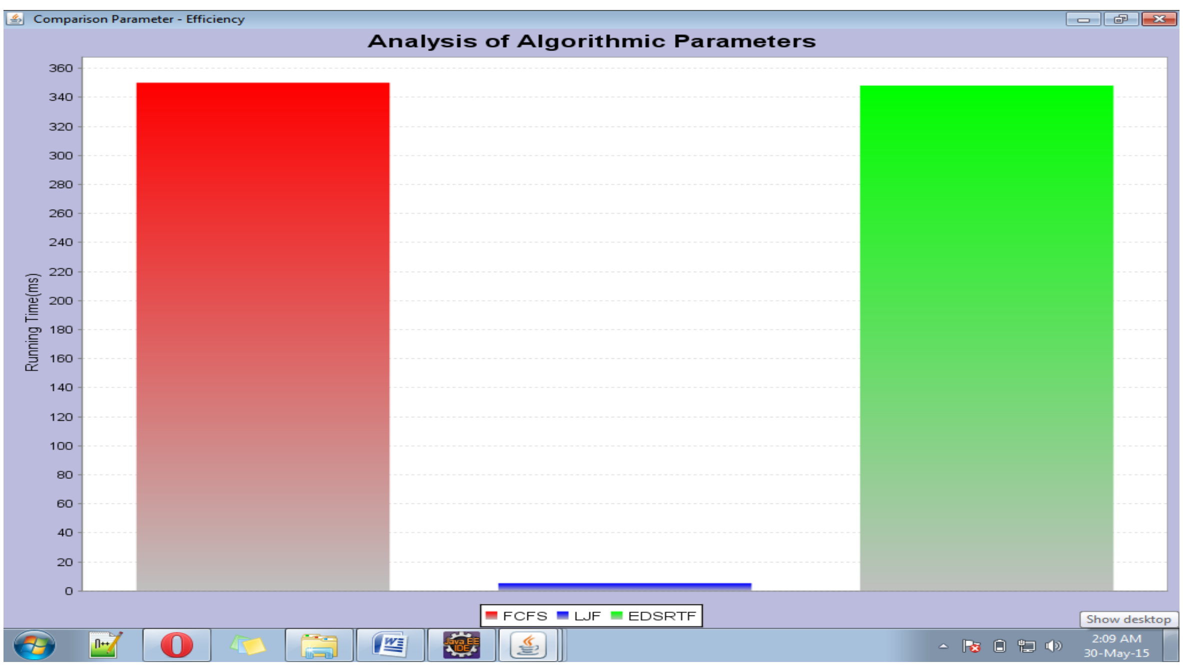

Figure 4.11 – Performance of Algorithm on the Parameter - EFFICIENCY

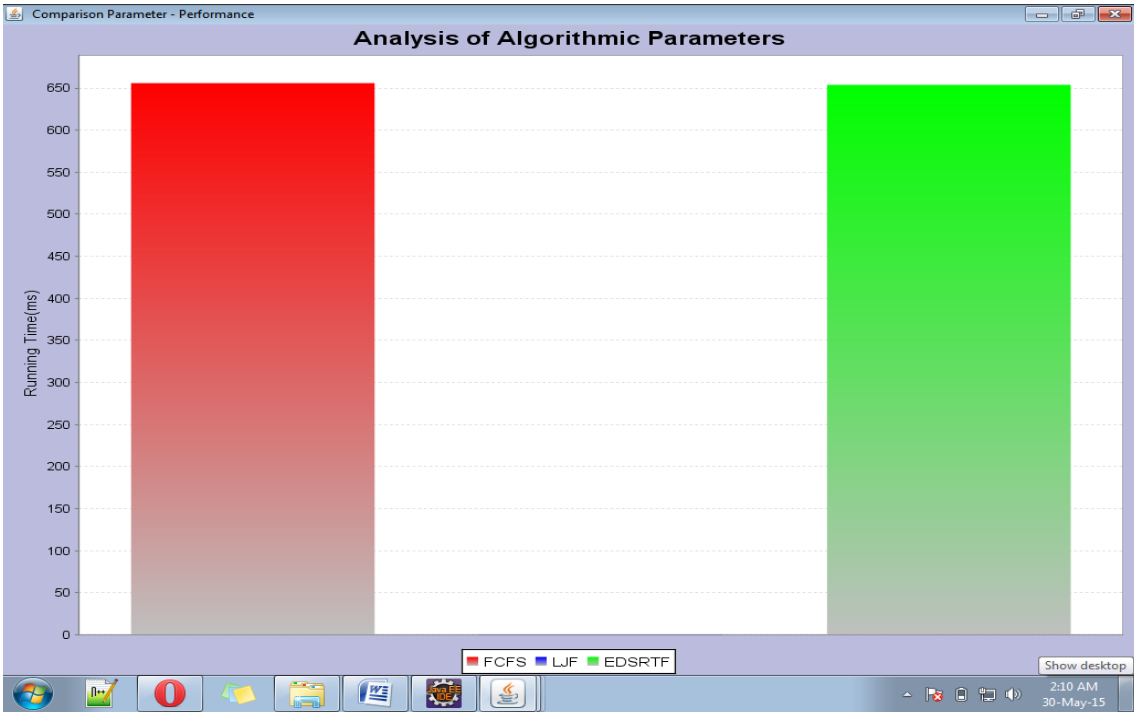

Figure 4.12 – Performance of Algorithm on the Parameter – OVERALL PERFORMANCE

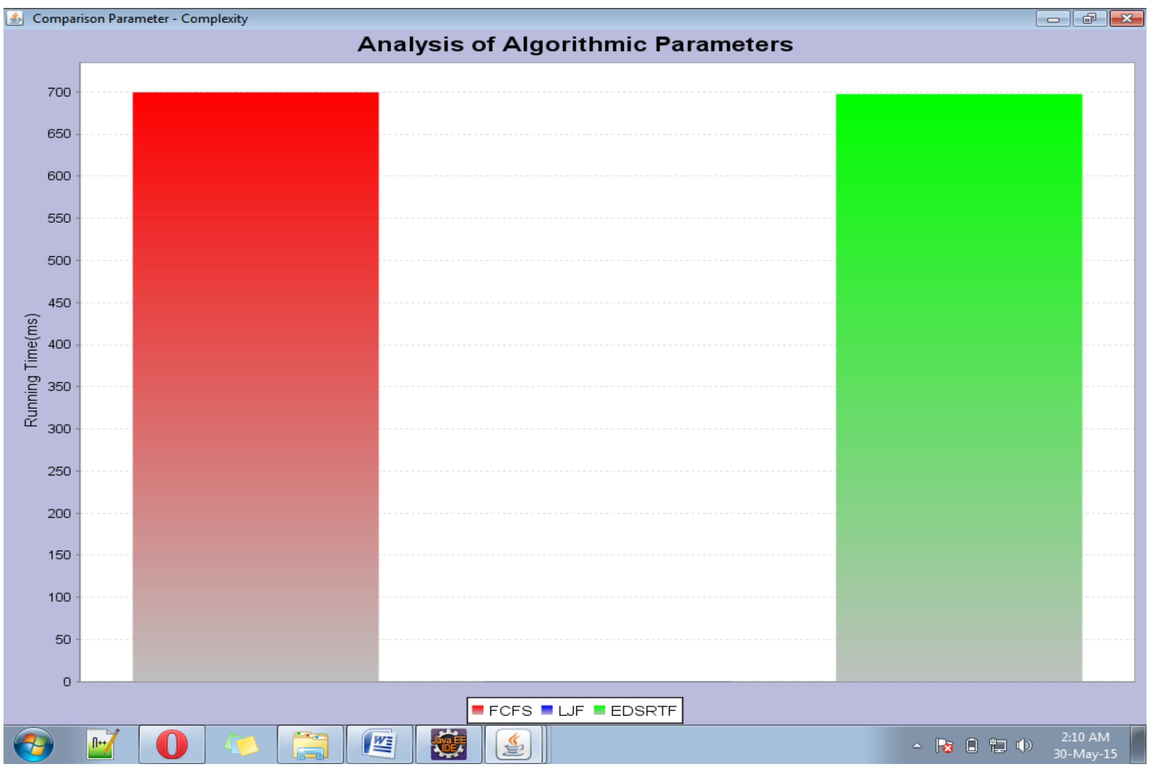

Figure 4.13 – Performance of Algorithm on the Parameter - ALGORITHMIC COMPLEXITY

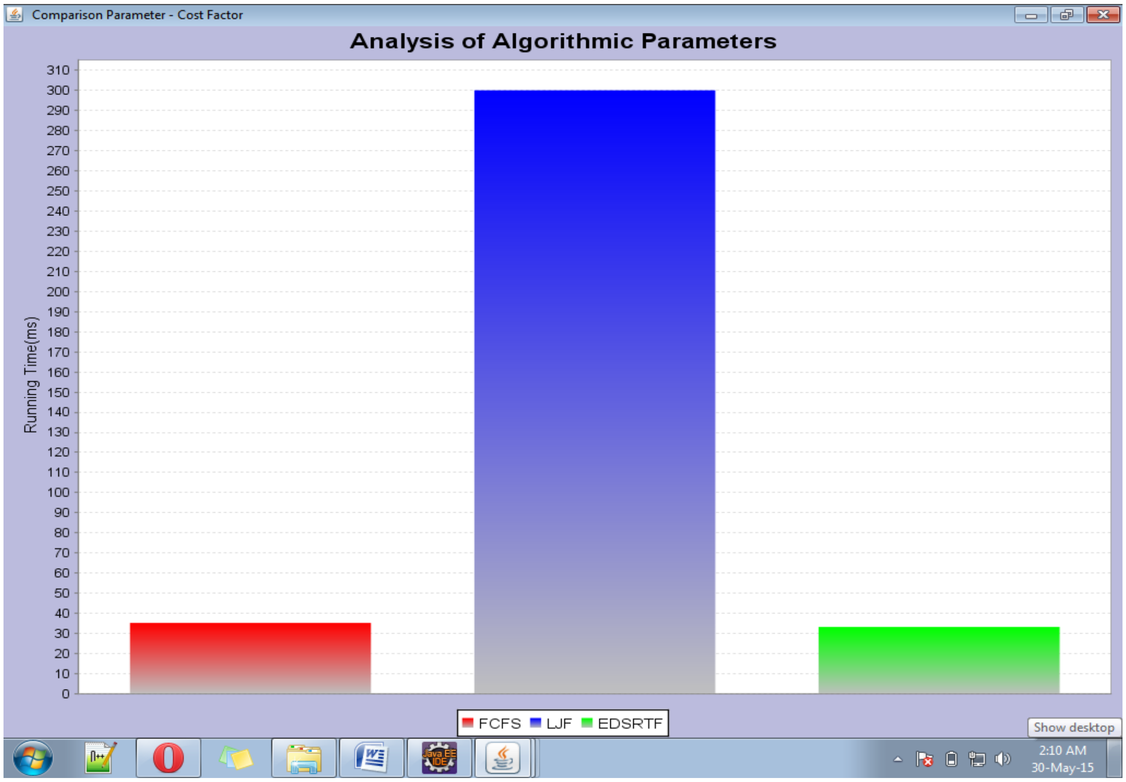

Figure 4.14 – Performance of Algorithm on the Parameter – COST FACTOR

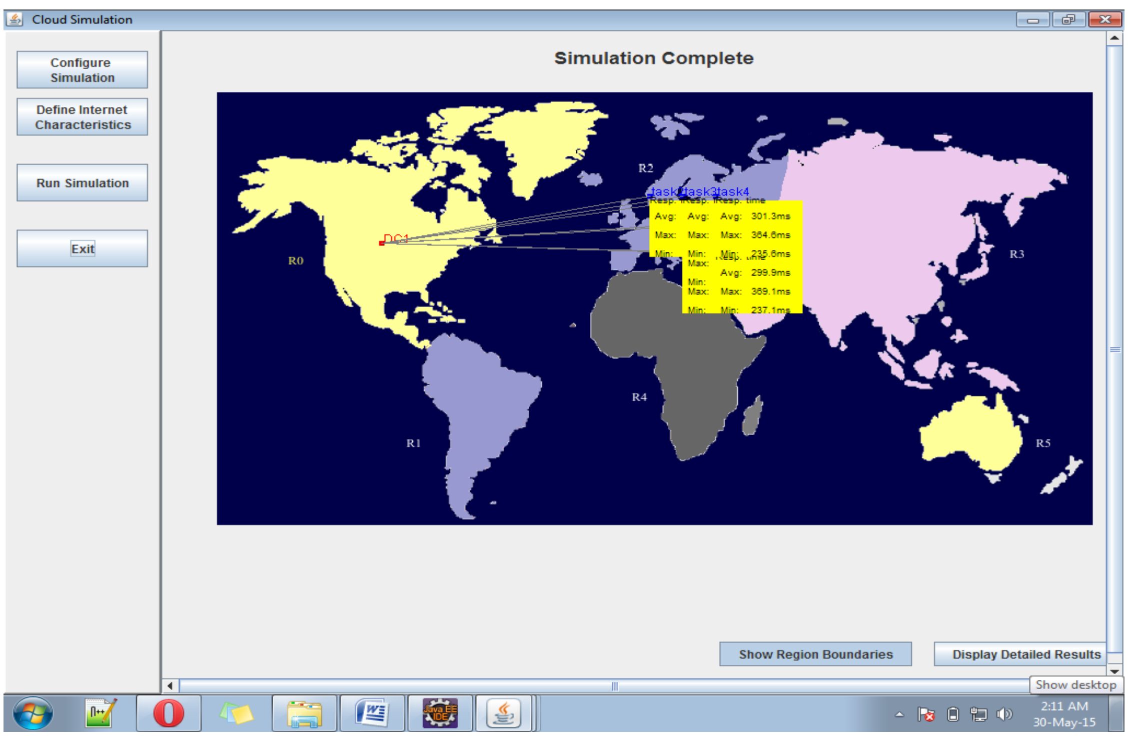

Figure 4.15 – Complete Simulation and Show Data Centers with Task Attributes with Time

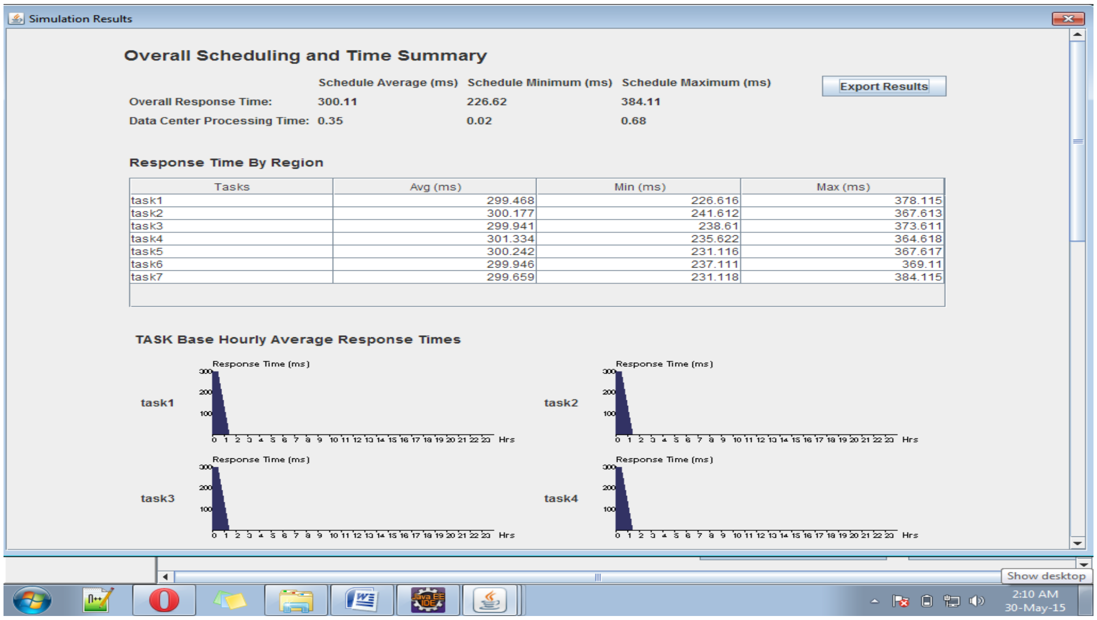

Figure 4.16 – Detailed Results of the Simulation - 1

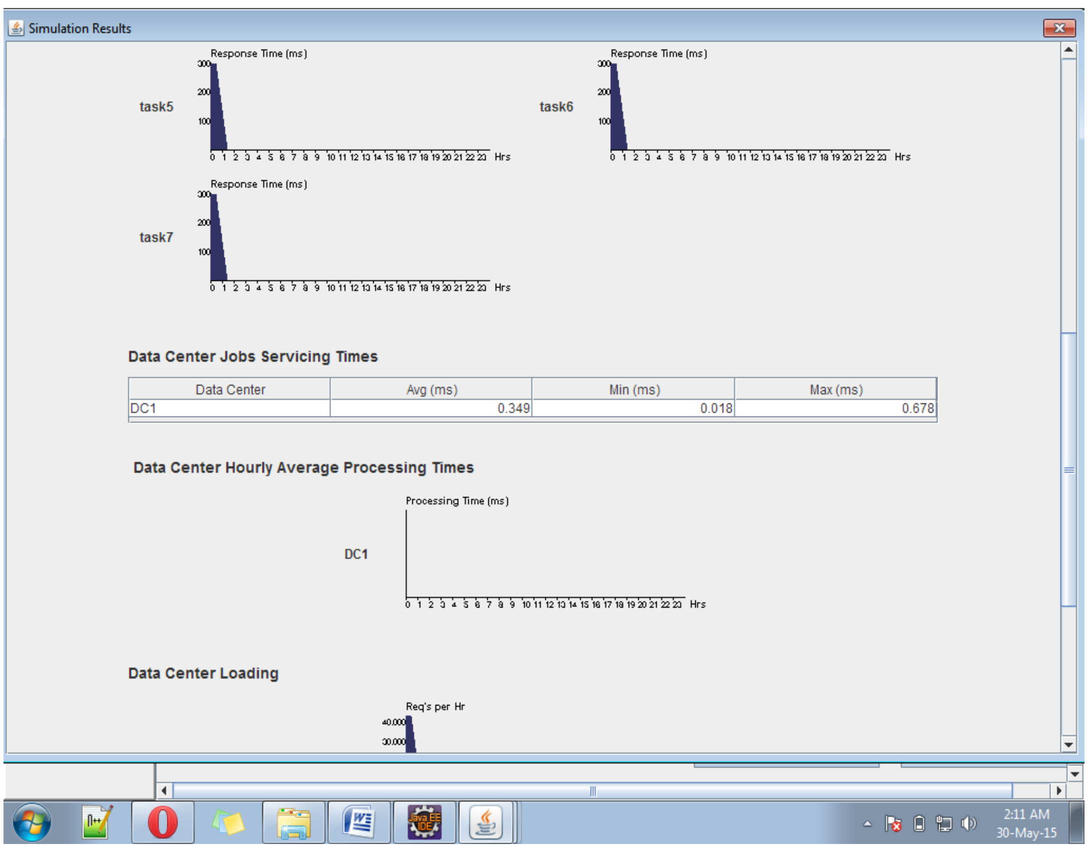

Figure 4.17 – Detailed Results of the Simulation - 2

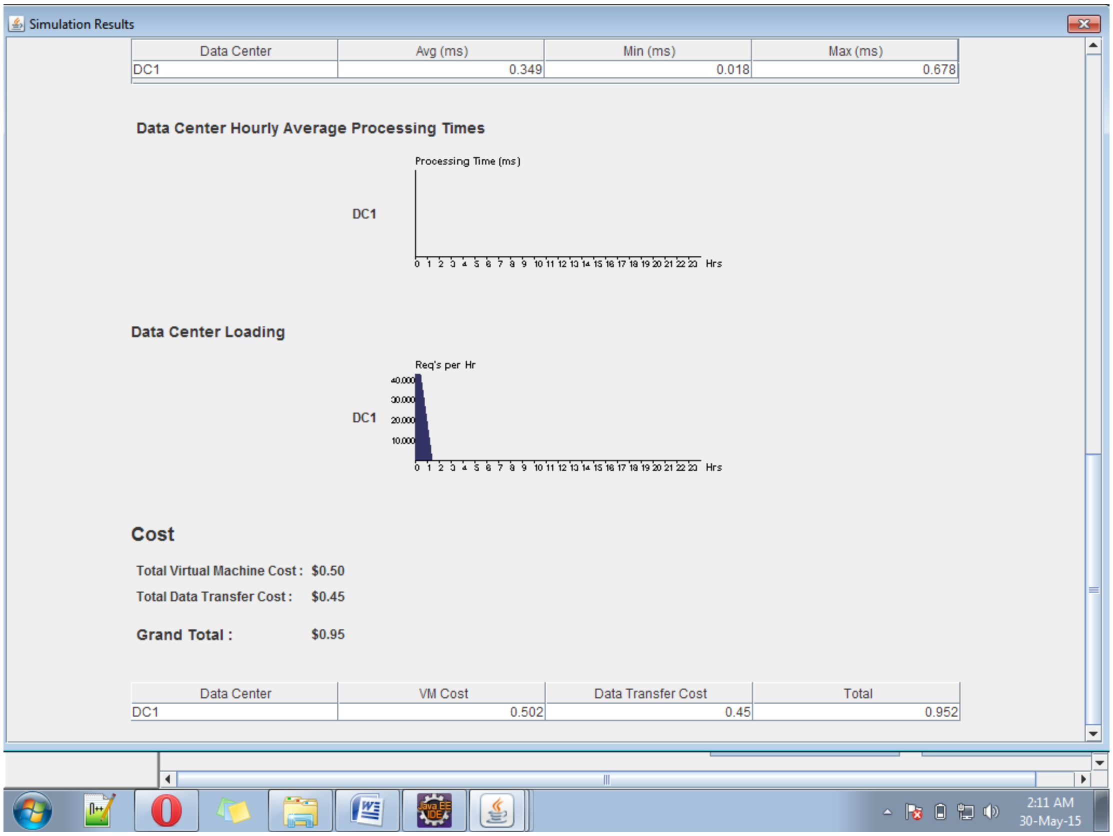

Figure 4.18 : Detailed Results of the Simulation – 3

SCENARIO -2

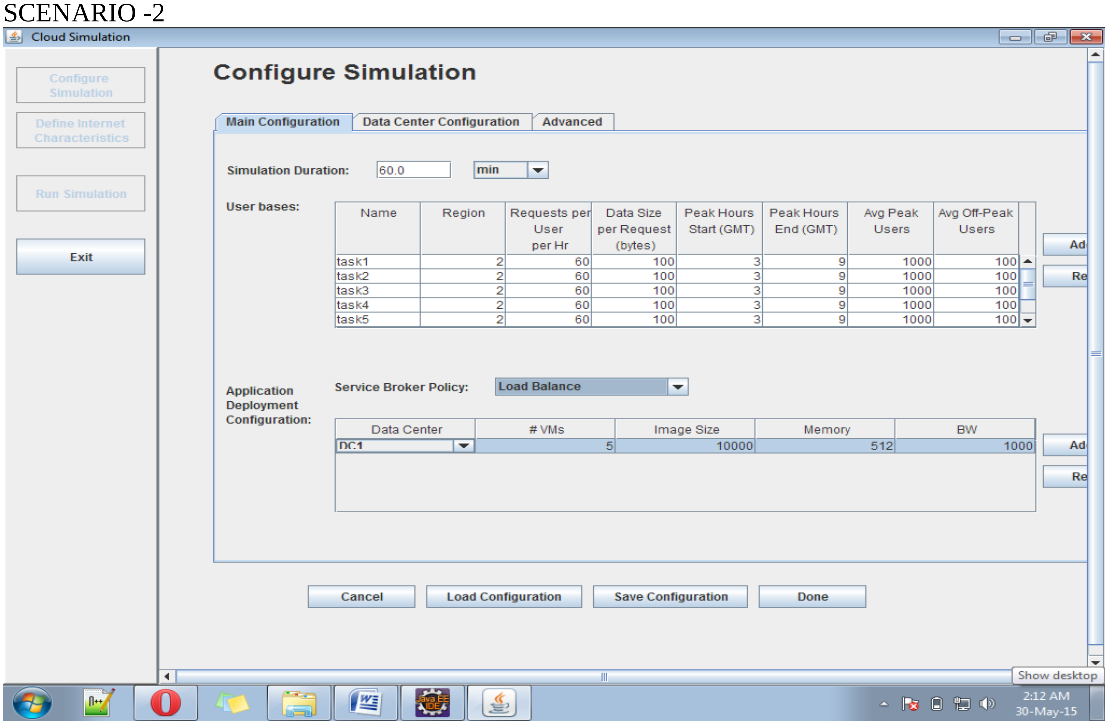

Figure 4.19: Configure Simulation 2

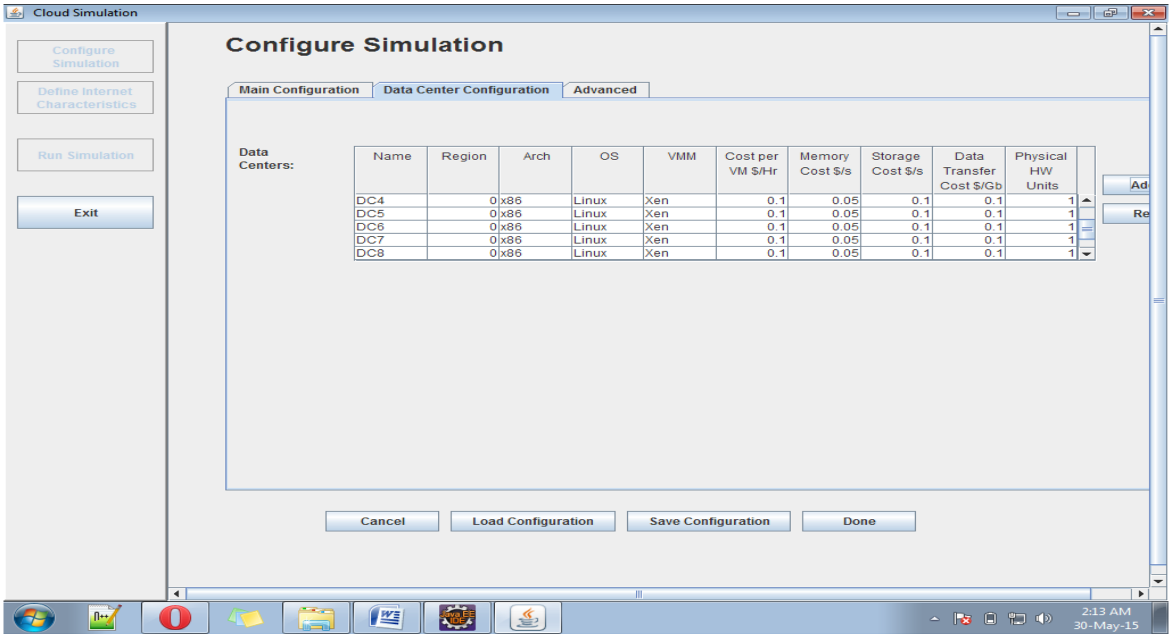

Figure 4.20 – Configure Simulation with more data centers

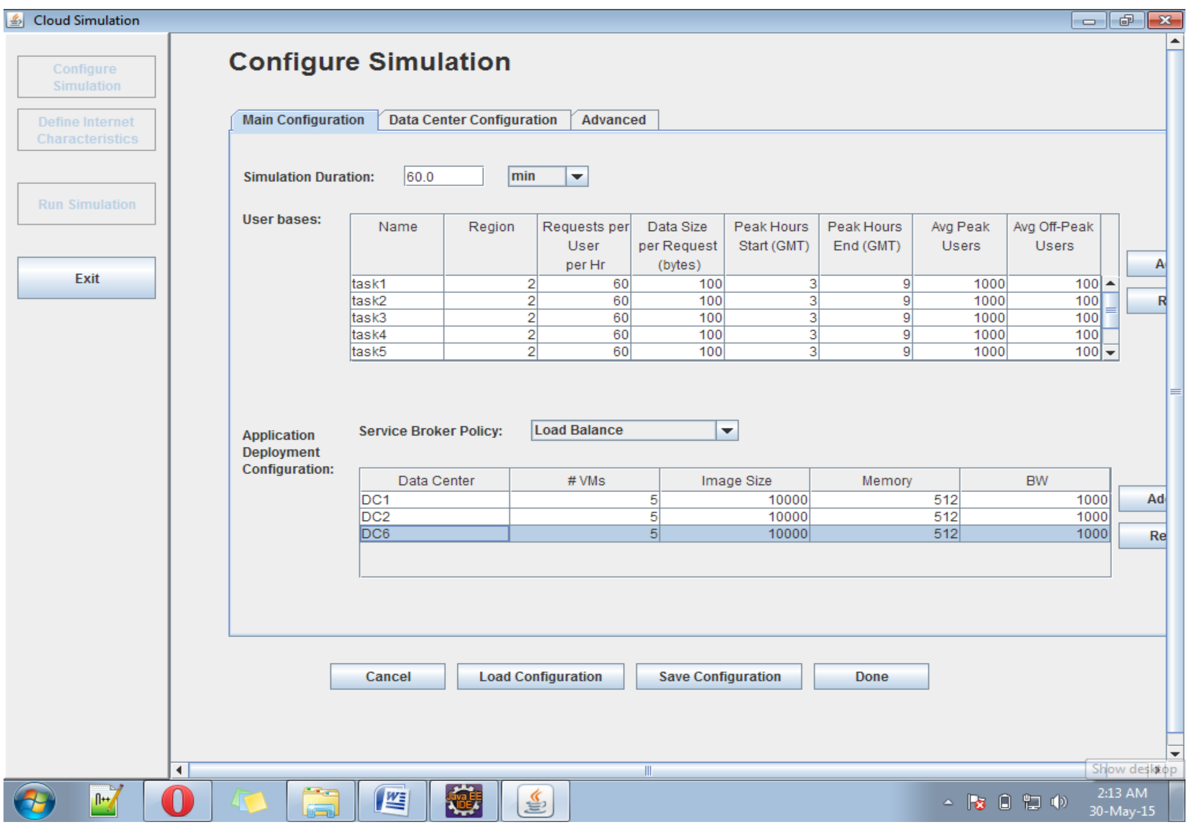

Figure 4.21 – Main Configuration Panel and Task Allocation

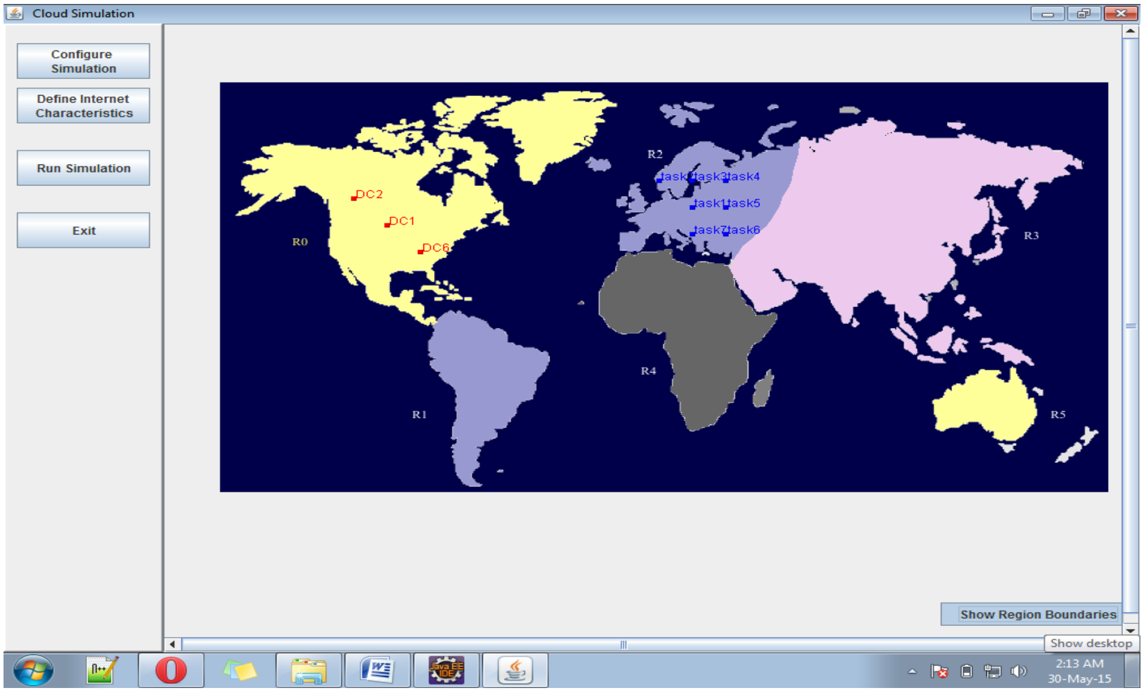

Figure 4.22 – Show Region Boundaries

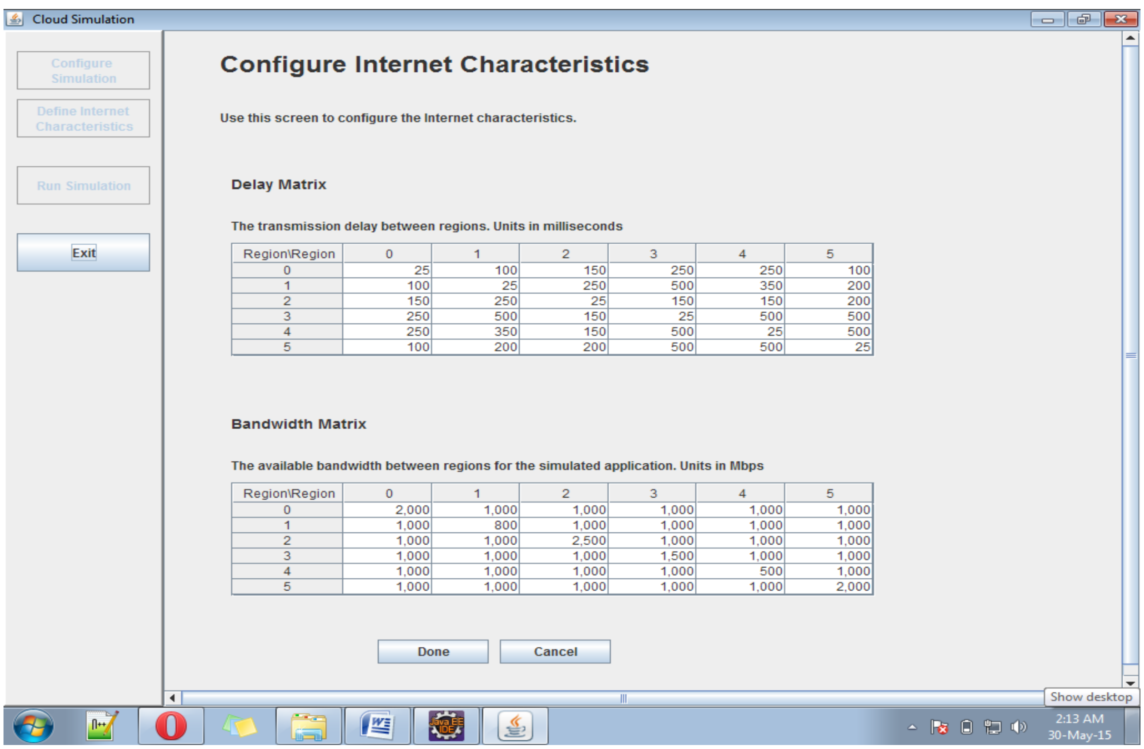

Figure 4.23 – Delay Matrix and Internet Characteristics

Figure 4.24 – Execute Simulation Scenario - 2

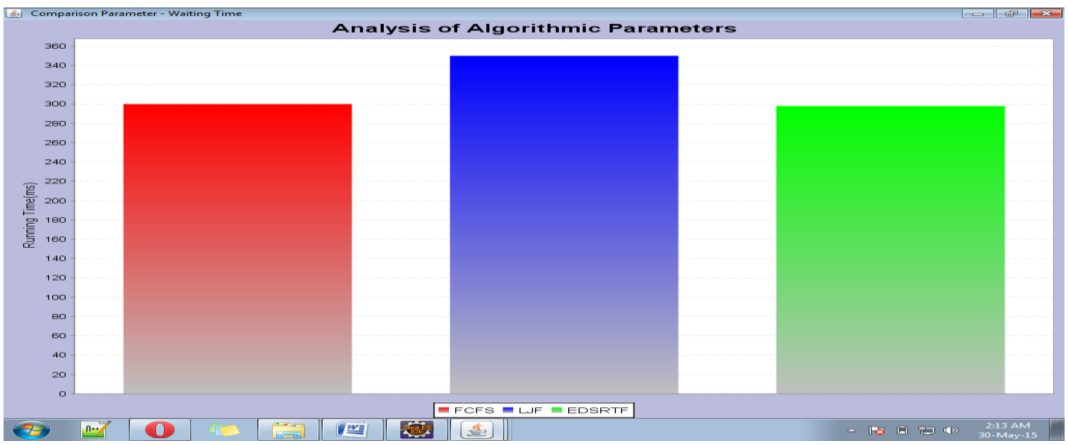

Figure 4.25 – Analysis on the Parameter Waiting Time

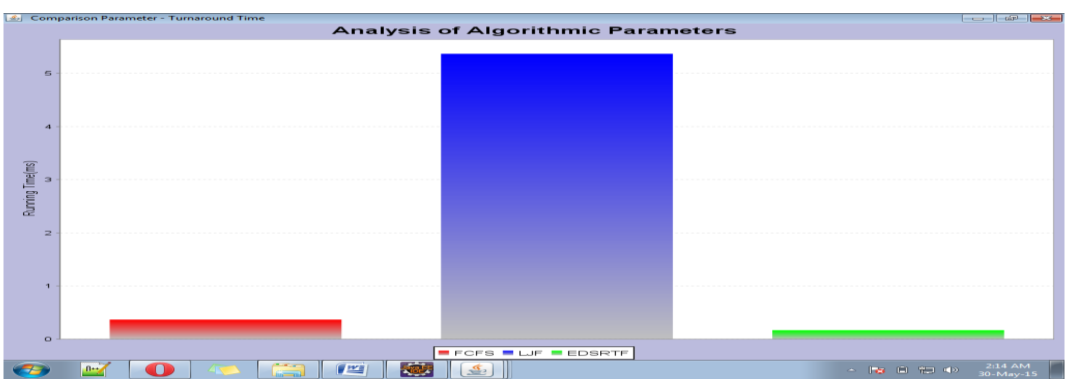

Figure 4.26 – Analysis on the Parameter Turnaround Time

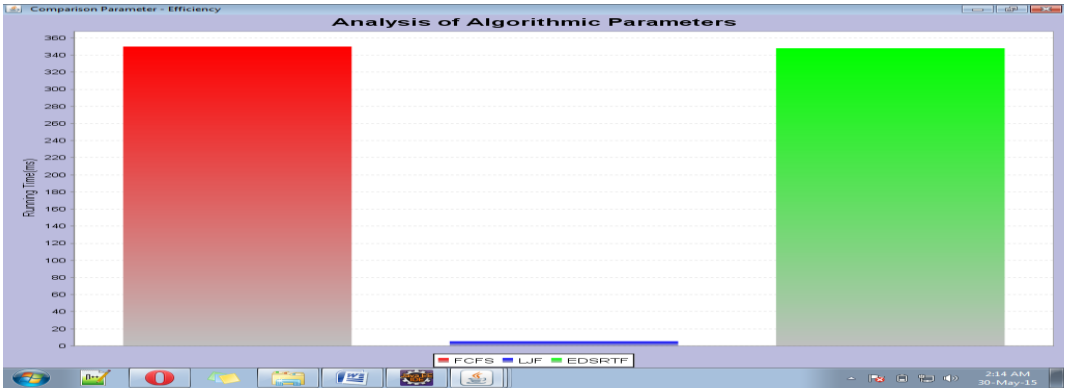

Figure 4.27 – Analysis on the Parameter Efficiency

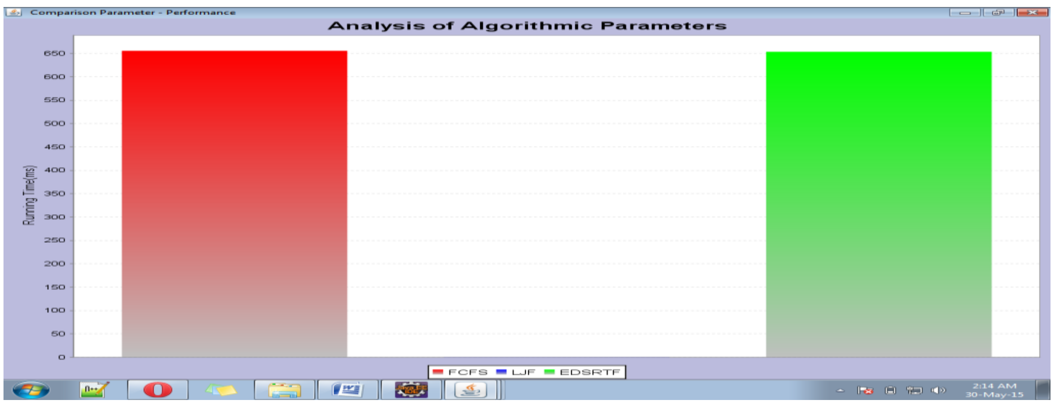

Figure 4.28 – Analysis on the Parameter Performance

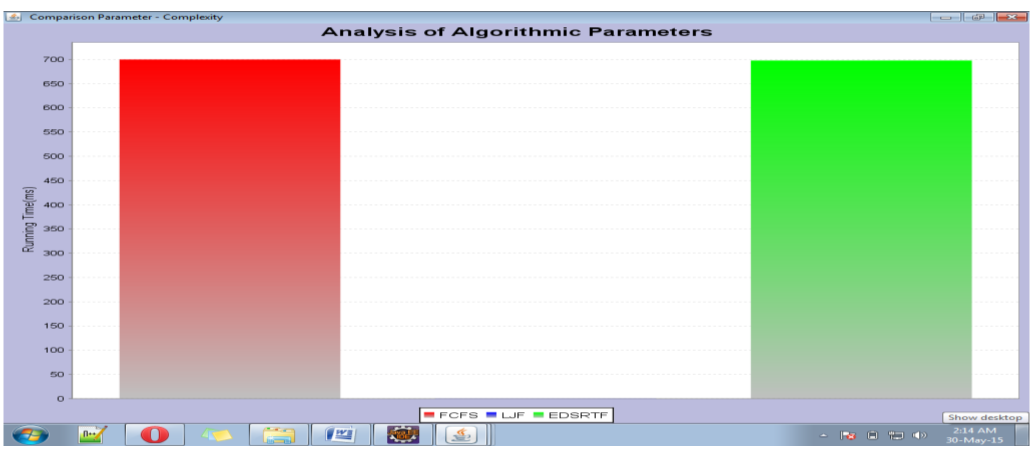

Figure 4.29 – Analysis on the Parameter Complexity

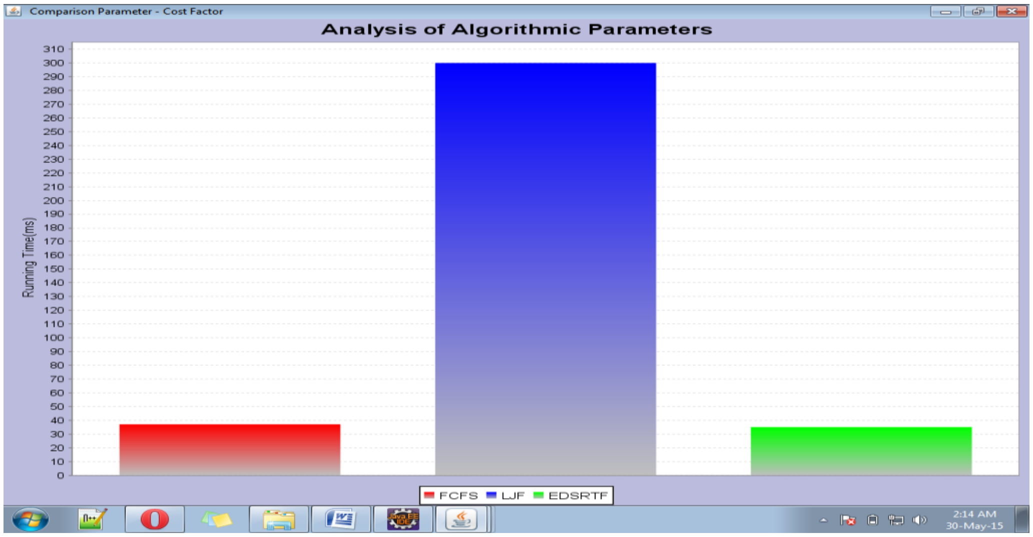

Figure 4.30 – Analysis on the Parameter Cost Factor

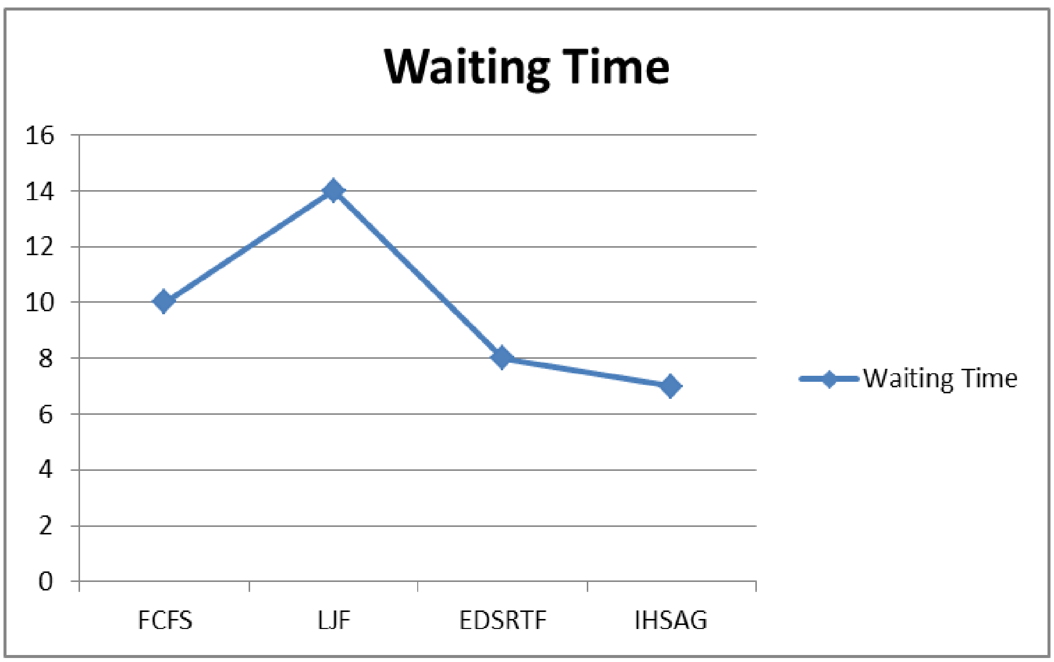

Figure:4.31 Waiting Time of All Algorithms

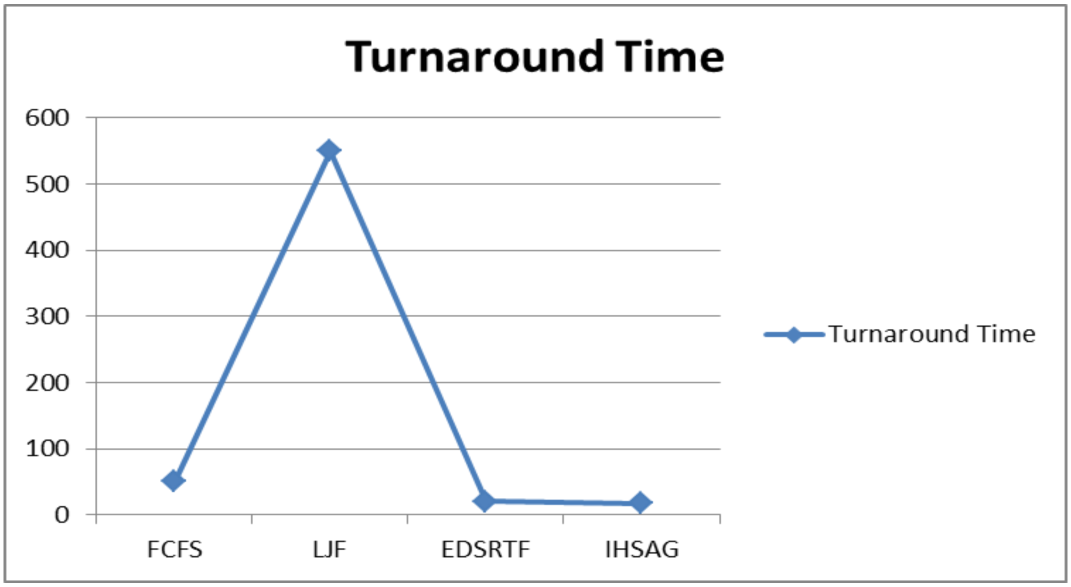

Figure:4.32 Turnaround Time of All Algorithms

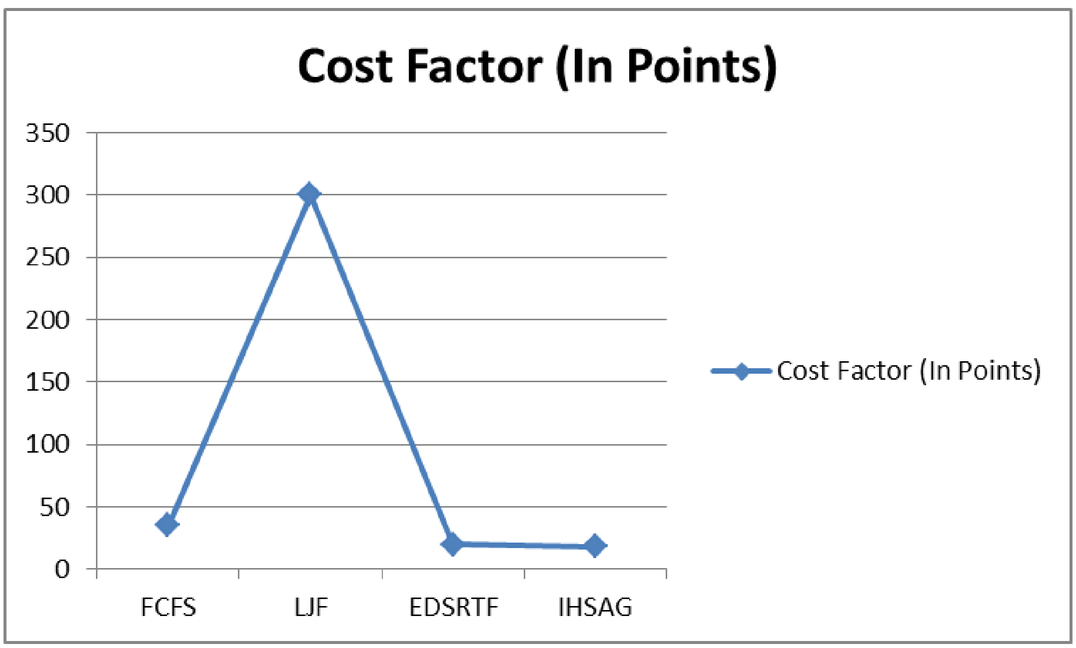

Figure: 4.33 Costs Factor of All Algorithms

It is evident from the simulation results and Table 1 that the cumulative result based on all the parameters are effective and better in the proposed approach name IHSAG.

Implemented Scheduling Algorithm	Waiting Time	Turnaround Time	Efficiency (In Percentage)	Performance (In Percentage)	Complexity	Cost Factor (In Points)
FCFS	10	50	75	85	690	35
LJF	14	550	20	40	20	300
EDSRTF	8	20	70	87	680	20
IHSAG (Improved Hybrid Scheduling Algorithm for Grid Infrastructure)	7	18	95	96	670	18

Table 1 - Tabular Comparison of the Results Obtained

Chapter 5

CONCLUSION AND FUTURE SCOPE

There are lot of scheduling algorithms already existing in grid computing scheduling environments. The research work aims to produce a more effective scheduling algorithm to reduce the effort by the grid in the procedure to execution of jobs.

In existing algorithms comparison is done in terms of waiting time and turnaround time with two algorithms. The proposed algorithms is implement the grid computing environment and process the 100% of the jobs with in comparison to the previous implemented algorithms.

This research work focuses on job scheduling algorithms and their performance on multiple parameters in the grid environment.

For future scope of the work, following techniques can be used in hybrid approach to better and efficient results –

- Particle Swarm Optimization
- HoneyBee Algorithm
- Simulated Annealing
- Genetic Algorithmic Approaches

Chapter 6

REFERENCES

---

[1] Abba, H. A., Shah, S. N. M., Zakaria, N. B., & Pal, A. J. (2012, October). Deadline Based Performance Evaluation of Job Scheduling Algorithms. In *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2012 International Conference on* (pp. 106-110). IEEE.

[2] Abba, H. A., Zakaria, N. B., & Pal, A. J. (2012). Comparative Analysis of Some Grid Scheduling Algorithms on Real Computational Grid Environment.

[3] Adil Yousif, A. (2011). Scheduling jobs on grid computing using firefly algorithm. *Journal of Theoretical and Applied Information Technology*, *33*(2), 155-164.

[4] Amudha, T., & Dhivyaprabha, T. T. (2011, June).Qos priority based scheduling

[5] Arora, S., Singh, S., Singh, S., & Sharma, B. (2014, December). Mutated firefly algorithm. In *Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference on* (pp. 33-38). IEEE.

[6] Arsuaga-Rios, M., & Vega-Rodriguez, M. (2014, August). Cost optimization based on brain storming for grid scheduling. In *Innovative Computing Technology (INTECH), 2014 Fourth International Conference on* (pp. 31-36). IEEE.

[7] Bi, J., Zhu, Z., Tian, R., & Wang, Q. (2010, July). Dynamic provisioning modeling for virtualized multi-tier applications in cloud data center. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on* (pp. 370-377). IEEE.

[8] Chang, R. S., Chang, J. S., & Lin, P. S. (2009). An ant algorithm for balanced job scheduling in grids. *Future Generation Computer Systems*, *25*(1), 20-27.

[9] Dong, F., & Akl, S. G. (2006). Scheduling algorithms for grid computing: State of the art and open problems (Vol. 504). Technical report.

[10] Dwivedi, S. K., & Gupta, R. (2014, September). A simulator based performance analysis of multilevel feedback queue scheduling. In *Computer and Communication Technology (ICCCT), 2014 International Conference on* (pp. 341-346). IEEE.

[11] Dipti Sharma, D., & Mittal, P. (2013). Job Scheduling Algorithm for Computational Grid in Grid Computing Environment. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(5).

[12] Gomathi, S., & Manimegalai, D. (2011, July). An adaptive grouping based job scheduling in grid computing. In *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on* (pp. 192-196). IEEE.

[13] Kannan, G., & Selvi, S. T. (2010, August). Nonpreemptive Priority (NPRP) based Job Scheduling model for virtualized grid environment. In *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on* (Vol. 4, pp. V4-377). IEEE.

[14] Murugesan, G., & Chellappan, C. (2010). An economic-based resource management and scheduling for grid computing applications. *arXiv preprint arXiv:1004.3566*.

[15] Ku-Mahamud, K. R., & Nasir, H. J. A. (2010, May). Ant colony algorithm for job scheduling in grid computing. In *Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on* (pp. 40-45). IEEE.

[16] Manimala, R., & Suresh, P. (2013, February). Load balanced job scheduling approach for grid environment. In *Information Communication and Embedded Systems (ICICES), 2013 International Conference on* (pp. 336-339). IEEE.

[17] Murugesan, G., & Chellappan, C. (2010). An economic-based resource management and scheduling for grid computing applications. *arXiv preprint arXiv:1004.3566*.

[18] Nithyapriya, D., & Krishnamoorthy. (2010). Hybrid adaptive meta-scheduling system for grid computing. *IEEE*, 356-360.

[19] Nithya, G., & Engels, R. M. S. (2014, February). Multi-agent brokering approach and firefly algorithm for job scheduling in grid environment. In *Electronics and Communication Systems (ICECS), 2014 International Conference on* (pp. 1-5). IEEE.

[20] Prajapati, H. B., & Shah, V. A. (2014, February). Scheduling in grid computing environment. In *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on* (pp. 315-324). IEEE.

[21] Sharma, R., Soni, V. K., Mishra, M. K., & Bhuyan, P. (2010). A survey of job scheduling and resource management in grid computing. *world academy of science, engineering and technology*, *64*, 461-466.

[22] Sharma, R., Soni, V. K., & Mishra, M. K. (2010, June). Notice of Retraction An improved resource scheduling approach using Job Grouping strategy in grid computing. In *Educational and Network Technology (ICENT), 2010 International Conference on* (pp. 94-96). IEEE.

[23] R. Z., Wu, M. S., Mi, X., & An, Q. (2012, September). Task Scheduling Algorithm Based on Triangle Module in Grid Computing. In *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on* (pp. 1-4). IEEE.

[24] Sharma, R., & Verma, S. K. (2014). A Deadline based Task Scheduling Algorithm for Heterogeneous Grid Environments. *International Journal of Computer Applications, 89*(5).

[25] Srinivasan, S., Kettimuthu, R., Subramani, V., & Sadayappan, P. (2002). Characterization of backfilling strategies for parallel job scheduling. In *Parallel Processing Workshops, 2002. Proceedings. International Conference on* (pp. 514-519). IEEE.

[26] Xhafa, F., Kołodziej, J., Barolli, L., & Fundo, A. (2011, September). A GA+ TS hybrid algorithm for independent batch scheduling in computational grids. In *Network-Based Information Systems (NBiS), 2011 14th International Conference on* (pp. 229-235). IEEE.

[27] Wang, J., Wu, G., Zhang, B., & Hu, X. (2012, September). A heuristic algorithm for scheduling on grid computing environment. In *ChinaGrid Annual Conference (ChinaGrid), 2012 Seventh* (pp. 36-42). IEEE.

Websites

[1] http://www.gridcomputing.com/

[2]http://www.ibmpressbooks.com/articles/article.asp?p=169508&seq Num=2

[3] http://www.redbooks.ibm.com/redbooks/pdfs/sg246778.pdf

[4] http://www.softcomputing.net/meta_mhs.pdf

[5] http://www.prudens.com/patens/compswinfra/gridcomp.html