A Novel Task Scheduling DSJF algorithm with K means clustering and Honey Bee Load Balancing Technique in Cloud Computing

Shanoo Aggarwal[1], Anil rao[2]
[1]*Research Scholar*, [2]*Assistant Professor, CSE Department,*
Institute of Engineering and Technology, Rajasthan Technical University, Alwar

Abstract: Cloud computing (CC) is fast-growing and frequently adopted in information technology (IT) environments due to the benefits it offers. Task scheduling and load balancing are amongst the hot topics in the realm of CC. To overcome the shortcomings of the existing task scheduling and load balancing approaches, we propose a novel approach that uses k means clustering and DSJF algorithm for task scheduling. In this paper, load balancing algorithm based on honey bee behaviour (LBA_HB) is proposed. Its main goal is distribute workload of multiple network links in the way that avoid underutilization and over utilization of the resources. Many scheduling algorithms have been proposed to schedule the tasks in cloud computing environment such as (SJF) and (FCFS) algorithms. This paper aims to improve the shortest job first scheduling algorithm in the cloud computing. In tasks scheduling (TS), the most important parameters are makes pan and response time. The proposed task scheduling technique is compared with the existing Heterogeneous Shortest Job First (DSJF) technique in terms of a) Make span time, b) Processing Time, c) Processing Cost, d) Execution Time.

I INTRODUCTION

CLOUD COMPUTING

Cloud computing provides shared processing resources and data. This can occur through the presence of a host application service provider so that the user does not need to buy a server or pay for the electricity of power and cooling. It's also convenient for communications and travels where remote workers, who can simply log in and use their applications wherever they are [1]. As increasing the number of users in cloud computing environment, the demand of shared resources is rapidly increased. Therefore, load balancing between these resources for scheduling tasks becomes a key challenge.

The cloud environment has a number of heterogeneous resources hosted in data centers in different locations. These data centers contain a large number of physical machines, which contain Virtual Machines (VM). Each VM has a specific configuration of processing power, RAM, communication bandwidth and storage associated with it. So customers do not require purchasing several hardware and software [1]. Cloud computing (CC) has two components: Provider and User. The user submits the task(s) to provider for execution. The Provider receives the users' tasks, executes them at a specific data center, then sends the results back to the user [3].

Cloud computing is a way to use the Internet in the daily life from your PC and Laptop. “Cloud Computing” came into action to know what happens when our data is moved to internet that is in “cloud”. This type of system allows employees to work remotely. Companies providing cloud services

enable users to store files and applications on remote servers, and then access all the data via the internet. Cloud security is a key concern for cloud storage providers. They not only must satisfy their customers; they also must follow certain regulatory requirements for storing sensitive data such as credit card numbers and health information. Third party audits of a cloud provider's security systems and procedures help ensure that users' data is safe.

Maintaining the security of data in the cloud extends beyond securing the cloud itself. Cloud users must protect access to the cloud that can be gained from data stored on mobile devices or carelessness with login credentials. Another cloud security issue is that data stored on a cloud hosted in another country may be subject to different regulations and privacy measures.

bs to specific assets specifically time. In environment of cloud computing, issue of scheduling of job is a greatest and testing issue. Subsequently the job scheduler ought to be rapid. The scheduling of job in cloud computing is principally centres to enhance the effective utilization of resource like lessening in completion time, memory and bandwidth. A productive strategy of job scheduling must intend for yielding less time of response in such a way that the execution of jobs submitted happens inside a conceivable least time and hence there would be a happening of event of in-time where reallocation of jobs is done. Subsequently, few dismissals of jobs happens and additional quantity of jobs could be put forward to the cloud by the customers which at last show expanding results in quickening the business execution of the cloud. [1]

Performance Metrics used for comparison of existing algorithm with the proposed approach:

a) Make span time: It is the total amount of time required to complete a group of tasks which indicates the completion time of last task. The scheduling task is to minimize the makes pan as most of the user's desire fastest execution of their applications [14]

b) Processing Time: It is time consumed by an algorithm to perform given task. It is determined using the formula given below:

$$Processing_{time} = cloudlet_{length}/(vm_{MIPS} * no_{ofPES})$$

c) Processing Cost: The cost of a task is the cost required to create an optimal scheduling based on the number of virtual machine movements divided by the total number of virtual machines on a particular physical machine. The cost function is expressed by the following equation:

$$Ci = 1/\ PM\sum (\ .\ of\ movement\ in\ VM/\ Total\ VMs\ )$$

d) Execution Time: The execution time of a complete task in the task scheduling algorithm depends on the task Length and processing capacity. This function is expressed in the following equation:

$$Ei = TL/PC$$

Where Ei denotes Execution Time, TL represents Task Length, PC denotes Processor Capacity

II RELATED WORKS

In this paper [3], the author has explained an optimized scheduling algorithm which uses K-Means Clustering Algorithm for grouping of tasks and Virtual Machines (VM). The proposed system categorizes the virtual environment based on the available applications in each machine. Experimental analysis of proposed Scheduling Algorithm based on Credits and K-means

clustering, has outperformed other job scheduling algorithms.

In this paper [4], the presenter has discussed that cloud services have heterogeneous Virtual Machines (VMs), with specified configuration and resources utilization, accommodated on various servers situated different geographical location that may lead to imbalanced among resource and task scheduling. This results in poor performance, inefficient energy consumption, violation of service level agreement (SLAs), and instability in system. To overcome these issues, Load balancing plays vital role. The aim of load balancing algorithms is to improve the resource utilization, energy saving and deduction of carbon emission. In this paper, we have surveyed the nature inspired load balancing algorithm such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), and Genetic Algorithm (GA), BAT.

In this paper [5], the author has demonstrated the proper use of load balancing techniques to increase the performance of the system and reduce the cost and energy. Enhanced Bee Colony algorithm based multi-objective task scheduling method is presented and its performance is verified and tested by the CloudSim stimulator. Also, through the experimental analysis, the best performance is shown for the proposed work.

In this paper [6], author has discussed the working of the scheduling algorithms. The algorithms studied in the paper are First Come First Serve, Shortest Job First, Opportunistic Load Balancing, and Generalized Priority Algorithm. They were put to test under the different condition and situations and were assessed dependent on parameters, for example, Cost and Make span. In light of these investigations, an intelligent tool was developed which takes in the user preferences, such as the number of jobs, their preference of time metrics and cost metrics and suggests the scheduler about the number of VMs, and the Scheduling calculation which ought to be utilized to suit the client prerequisites.

In this paper [7] improved credit based scheduling algorithm using the parameters viz. priority, task length and deadline is proposed. In the experimental result, the proposed algorithm shows better result than the existing algorithm. Credits are used to reduce the make span of the task and execute all the task in cloud.

In this paper [8], the proposed technique shows the priority tasks are removed from the overloaded virtual machine and they are allocated to under loaded virtual machine. It helps to reduce the minimum completion time, amount of waiting time of tasks in queue is minimal and achieve better resource utilization. This paper concludes that the minimum amount of time is taken to execute the tasks and better resource utilization.

In this paper [9], firstly it stores request from the user on the basis of Priority, Size and type of resource required. Apply Bee Colony Algorithm and assign resources. The performance is measured on the basis of requested load and served load, requested priority and served priority and minimizing the execution time. The proposed technique fulfils the user specific requirements such as priority execution and resource allocation.

In paper [10] they have presented a theoretical comparison of algorithms of job scheduling in environment of cloud computing. Cloud computing is a pay per-use, scalable and dynamic distributed model of computing empowering designers for

conveying applications in the midst of storage distribution and job designation. Cloud computing urges to give a pool of virtualized PC resource enabling designers to pass on applications in the midst of storage distribution and job designation. This paper expects to exhibit the execution correlation investigation of different prior algorithms of job scheduling considering different parameters. This paper talks about distributed computing and its constructs in area (i). In segment (ii) idea of job scheduling in cloud computing has been expounded. In segment (iii) existing algorithms for job scheduling are talked about, and are contrasted in a tabulated form and regard to different parameters and finally segment (iv) closes the paper giving brief synopsis of the work.

In paper [11], they have concentrated on the problem of job security scheduling problem under the cloud computing environment. The Architecture of the cloud computing platform is made up of four layers, including 1) SOA architecture, 2) Management Middleware, 3) Resource virtualization and 4) Physical Resources. Next, we formally describe the problem of the job security scheduling under cloud computing. To guarantee the security level of job scheduling, security demand and trust level are defined in our work. Afterwards, the proposed genetic algorithm-based job security scheduling is proposed. The main innovation of our proposed algorithm lies in that we utilize each chromosome to represent a schedule of a set of jobs on several computers, and then each gene is represented as a pair to describe the relationships between jobs and computers. Finally, experimental results demonstrate that our algorithm can effectively schedule jobs with higher level security and lower processing time cost.

In paper [12], they have presented survey on approaches which are related to game theory for scheduling of job in cloud computing environment. Cloud computing is one of the hopeful technologies in circumstance of today's environment. The scheduling of jobs in environment of cloud is an imperative issue in which the primary point is to plan the jobs fittingly keeping in mind the end goal to successfully use the resources furthermore meet the client's fulfilment. The provider of cloud needs to consider different viewpoints like number of cloud clients asking for an administration in the meantime, accessibility of resources around then, necessities of QoS, SLAs and so on. There are different methodologies for jobs scheduling. In this paper main concentration is on methodologies based on game theory for scheduling of jobs in cloud. Study of existing methodologies and different issues in scheduling of jobs which is based on the theory of game is the principle goal of this paper. The methodologies which are already available for scheduling of jobs centring methodologies of game theory and investigate the open issues for examination around there.

In paper [13] they have proposed an effective multi queue job scheduling for cloud computing. Cloud computing is one of the well creating field in Information Technology and Computer Science. The productive job scheduling builds the customer fulfilment and use the energy of system as far as time. An algorithm of MQS (Multi Queue Scheduling) proposed for diminishing the expenditure of both plans i.e. on demand and reservation utilizing the worldwide scheduler. Scheduling is one of the critical complicated part in cloud computing. A definitive point of global scheduler is to distribute the resources at most the greatest level. The one of the basic events in cloud

computing is job scheduling in light of the fact that the client need to pay for administrations in view of use time. The proposed technique portrays the idea of jobs clustering in light of burst time. Amid the season of scheduling the conventional strategies, for example, EASY, Shortest Job First, First Come First Serve, Improved backfill and Combinational Backfill utilizing technique of balance spiral are making fragmentation. The proposed strategy defeats this issue and lessens the starvation with all the while. This paper likewise concentrates some current algorithms of scheduling and issues identified with them in cloud computing. The proposed MQS strategy gives more significance to choose job powerfully keeping in mind the end goal to accomplish the ideal problem of cloud scheduling and thus it use the unused free space in a financial way.

In paper [14] they have discussed VM Provisioning and job scheduling in Clouds. Cloud computing has as of late developed as another worldview for facilitating and conveying administrations over the Internet. Job Scheduling assumes a critical part in Cloud Computing situations. In this paper, the effect of First Come First Served and Earliest Start time (EST) based Job scheduling are analysed. It is demonstrated that when Job Scheduling based on EST is utilized, determination of priorities can be done to execution of Cloudlets in light of their EST. After the scheduling of Cloudlets, Single Threshold based VM provisioning or Two Threshold based VM provisioning is performed at the Cloud Provider. In the proposed Two Threshold and Single Threshold dependent algorithm of VM provisioning, a savvy Data focus determination approach is utilized for selecting the Data Centre with least cost.

III PROPOSED WORK

3.1 Problem Formulation

In cloud computing, resources need to be allocated and scheduled in a way that helps both providers and users to achieve their objectives. The aim of provider is to achieve high resource utilization and user tends to meet their required application performance with minimum cost. This is termed as resource management problem. In this, research work, problem is narrowed down to attain performance optimization in terms of time and cost. To reach at optimized performance, an efficient task scheduling algorithm is proposed. The existing work is based on the heterogeneity of resources and tasks in shortest job first algorithm and attain the reduced Make span time. But using only heterogeneity concept on SJF can reduce the Make span time but due to this, starvation problem occurs. It means the task with largest length will always be scheduled at the last. Thereby, this gradually increases the processing time and make span time of scheduling the task as there is no criteria of the assigning the task to the machine also the resources utilization is missing as load is not checked before assigning the tasks.

The proposed work is based on the existing dynamic shortest job first algorithm which is integrated with the Task grouping i.e. tasks are assigned to the resources in the groups which will be generated using K-means Algorithm. Also, the load balancing concept is not used in the existing technique which will be included in the proposed. **OBJECTIVES**

1. To study and analyze the existing DSJF scheduling algorithm for cloud computing.
2. To implement proposed Task grouping based DSJF comprises of grouping using

K-means to reduce the overall processing of execution of jobs assigned to the resources.

3. To apply load balancing to the task scheduling algorithm for cloud resources.
4. To evaluate and compare the performance of existing DSJF with the proposed algorithm with respect to following parameters
   a) Make span time
   b) Processing Time
   c) Processing Cost
   d) Execution Time

Research Methodology

The proposed scheduling algorithm using the concept of grouping based on K-means then schedule the task to virtual machines using the load balancing algorithm honey bee optimization. This will reduce the Make span time and also improves the resources utilization. After sorting the tasks using length, proposing a more efficient scheduling using clustering of the sorted task and machines using K-means. Clustering at the cloudlet side is done on the basis of four parameters assigned i.e. task length, priority, deadline and cost. Clustering at virtual machine side is done using bandwidth, ram, MIPS and size. The machines and task are clustered into 3 groups namely, High, medium, low. Also, the sorting within the clustered groups is done then the sorted clustered groups are given to the load balancing Honey bee algorithm for the optimized balanced scheduling. Which will calculate the Load of the Virtual machines. Based upon overloaded and under loaded machines, tasks are assigned. Like if the load on machine is more i.e. overloaded machine, that machine will transfer the tasks to under loaded machine and vice-versa.

IV RESULTS AND DISCUSSIONS

We have used Cloud Sim as simulation tool for this work. We compared the performance of the existing DSJF algorithm with proposed algorithm using K-means clustering and Honey Bee as load balancing algorithm. The performance parameters considered are: a) Make span time, b) Processing Time, c) Processing Cost, d) Execution Time as shown in tables below. The number of tasks are varied and the performance is evaluated using existing and proposed algorithms accordingly.

No. of Tasks	Existing DSJF Algorithm	Proposed Algorithm
200	183.42 ms	104.33 ms
500	780.87 ms	615.81 ms
800	1769.08 ms	982.77 ms
1100	3126.35 ms	2750.25 ms
1400	4897.08 ms	4401.39 ms

Table 4.1

Make span Time Comparison

The above table shows that make span time is reduced by using proposed algorithm considerably. For 200 tasks, the make span time of existing algorithm is 183.42ms whereas it reduces to 104.33 ms. This is a required criteria for an efficient scheduling algorithm. It is also observed that the

performance of the proposed algorithm increases as the no. of tasks increase.

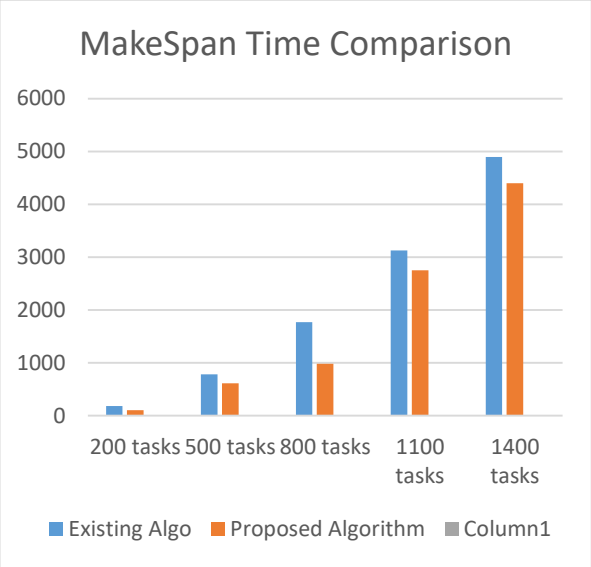

Figure 4.1

Make Span Time Comparison

II Total Processing Time Comparison

The total Processing time is measured for varied number of tasks and the results are as shown in the table 2.

No. of Tasks	Existing DSJF Algorithm	Proposed Algorithm
200	4854.88372	4509.6544
500	21100.60284	18831.5782
800	48034.1239	49372.1488
1100	85313.2752	75310.3846
1400	133276.0445	127680.5108

Table4.2 Comparison of Total Processing Time

It is clear from the Table 4.2 that the processing time decreases considerably when task scheduling is carried out using proposed algorithm as compared to the existing DSJF Algorithm. The total processing time in case of proposed algorithm is 4509.6544 ms when no. of tasks is 200, whereas it comes out to be 4854.88372 ms using existing algorithm. So, there is considerable reduction in total processing time using proposed algorithm.

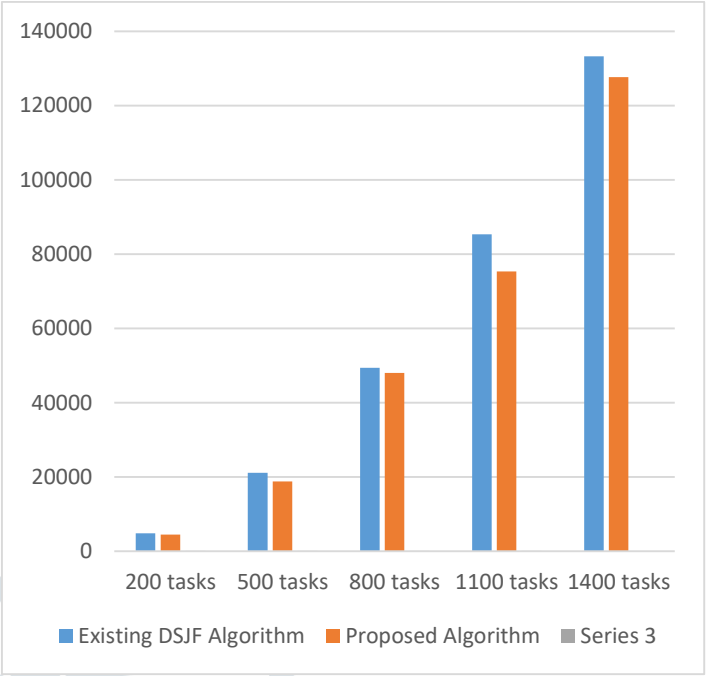

Figure 4.2: Total Processing Time Comparison

III Total Processing Cost

No. of Tasks	Existing DSJF Algorithm	Proposed Algorithm
200	22718.25	12479.438
500	57055.6979	30787.1689
800	92805.975	49412.9200
1100	128832.7259	67933.2360
1400	163910.5620	86503.8420

Table4.3 Comparison of Total processing Cost

The results after Simulation are recorded in the above table. It is showing the comparison of Total Processing Cost in case of both algorithms. It is shown that the total Processing Cost reduces significantly when we are using the proposed task scheduling Algorithm using K-means clustering and LBA_HB. When no. of tasks is 200, the total Processing Cost in case of existing Algorithm is 22718.25 and it comes out to be 12479.438 in case of Proposed Algorithm.

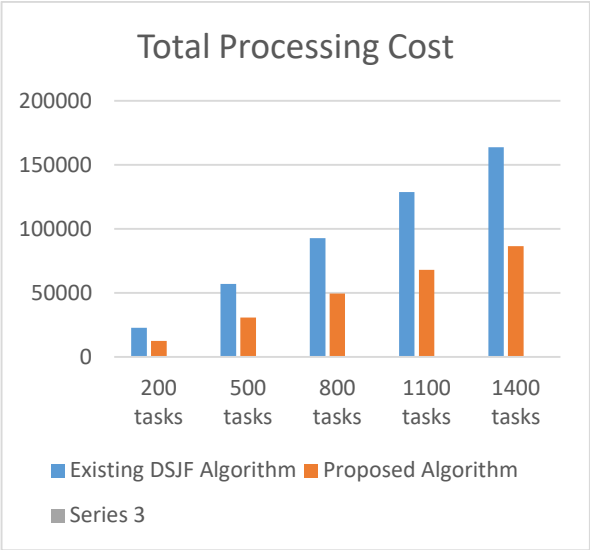

Figure 4.3
Total Processing Cost Comparison

IV Total Execution Time Comparison :

No. of Tasks	Existing DSJF Algorithm	Proposed Algorithm
200	7980	3093.0232
500	34950	13546.5116
800	79920	30976.7441
1100	142890	55383.7209
1400	223860	86767.4418

Comparison

The below Table 4.4 shows the comparison of Total Execution Time in case of Existing and Proposed Algorithm by varying the no. of Tasks. It is clear from the above table that the total execution time decreases significantly when the proposed Algorithm is being used. When no. of Tasks =200, total execution time is 7980 ms in case of existing algorithm and its value is 3093.0232 ms in case of Proposed Algorithm. So, it is depicted that total execution time decreases in case of proposed algorithm greatly.

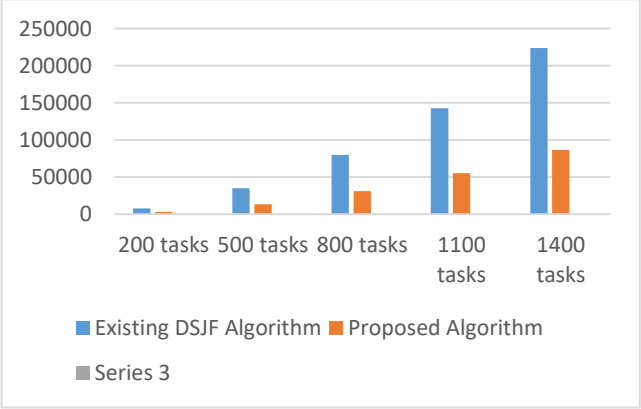

Figure 4.4
Total Execution Time Comparison

V CONCLUSION:

In this paper, existing DSJF algorithm is compared with proposed algorithm using k-means clustering scheduling and Honey Bee as Load balancing technique. The main objective of the scheduling technique is to assign users task and minimize cost and make span of the system. The multi-objective optimization approach is used to improve the scheduling performance compared to single function. The experimental outputs were taken based on 200, 500,800, 1100 and 1400 tasks. By using the Proposed Algorithm of task scheduling, a highly efficient solution is achieved. The results produced clearly show that the proposed k-means clustering with LB_HBA out performs the existing DSJF Algorithm in terms of performance and cost.

References

[1] Y. Chen, Z. Guan, X. Shao, S. K. Labs, D. M. Equipment, and H. Wuhan, "A Comparative Analysis of Job Scheduling Algorithm," *2011 IEEE International Conference on Management Science and Industrial Engineering (MSIE)* pp. 1091–1095, 2011.

[2] D. Agarwal and S. Jain, "Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment," 2014 *International Journal of Computer Trends and Technology (IJCTT)* vol. 9, no. 7, pp. 344–349, 2014.

[3] V. Sharma and M. Bala, "A Credits Based Scheduling Algorithm with K-means Clustering", *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) IEEE*, pp. 82-86, 2018.

[4] P. Jain and S. Kumar Sharma, "A systematic review of nature inspired load balancing algorithm in heterogeneous cloud computing

environment", *Conference on Information and Communication Technology (CICT) IEEE*, 2017.

[5] A P Shameer and A C Subhajini, "Throughput Maximization on Efficient Load Balancing in Cloud Task Scheduling Using Enhanced Bee Colony Algorithm*", International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) IEEE*, pp. 294-298, 2017.

[6] Y. Vijay and B. V. Ghita, "Evaluating cloud computing scheduling algorithms under different environment and scenarios", *8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) IEEE*, 2017.

[7] K. Raja, G. Sekar, "An Algorithm for Credit Based Scheduling in Cloud Computing environment depending upon Deadline Strategy", International Journal of Current Trends in Engineering and Research, Vol. 2, Issue8, pp.70-76,2016

[8] K. Zalavadiya, D. Vaghela, "Honey Bee Behavior Load Balancing of Tasks in Cloud Computing", International Journal of Computer Applications (IEEE), Vol. 139-No. 1, pp. 16-19, 2016.

[9] A. Soni, G. Vishwakarma, Y.K. Jain "A Bee Colony based Multi-Objective Load Balancing Technique for Cloud Computing", International Journal of Computer Application (IEEE), Vol. 114, No. 4, pp-19-25, 2015.

[10] A. Gupta, H. S. Bhadauria, A. Singh, and J. C. Patni, "A Theoretical Comparison of Job Scheduling Algorithms in Cloud Computing Environment," *2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015)* pp. 4–5, 2015.

[11] A. Ananth and K. C. Sekaran, "Game theoretic approaches for job scheduling in cloud computing: A survey," *Proc. - 2014 5th IEEE 5th International Conference on Computer and Communication Technology (ICCCT) 2014*, pp. 79–85, 2014.

[12] T.Mandal and S.Acharyya, "Optimal Task Scheduling in Cloud Computing Environment : Meta Heuristic Approaches," 2015 *Proceedings of International Conference on Electrical Information and Communication Technology* (EICT) pp. 24–28, 2015.

[13] AV. Karthick, E. Ramaraj, and R. G. Subramanian, "An efficient multi queue job scheduling for cloud computing," *Proc. World Congress on Computing and Communication Technologies WCCCT 2014*, pp. 164–166, 2014.

[14] A.Chinnu Edwin and A.Neela Madheswari, "Job scheduling and VM provisioning in clouds," Proc - 2013 IEEE *Third International Conference on Advances in Computing and Communications (ICACC)*, pp. 261–264, 2013.