Optimization of Task Scheduler in Cloud Computing

Sharada Ramesh Geete
PG Student,
SSVPS BS Deore College of Engineering, Dhule, 424005, India

Mandre B. R.
Associate Professor,
SSVPS BS Deore College of Engineering, Dhule, 424005, India

ABSTRACT
Cloud computing is a rising technology in distributed computing. Cloud computing is one of the latest technology. Cloud is developing day by day and focuses on many challenges one of the most important is scheduling. Main objective of scheduling algorithm is to make proper utilization of the resources. The Goal of this project is to apply the scheduling algorithm in public cloud to execute the given no of tasks within deadline and budget cost. Here PBACO algorithm is used to solve optimization problem and to achieve the global optimal path and avoid the local optima. The simulation result shows the algorithm reduces the execution time and user budget cost. The algorithm completes all task execution with Minimum makespan and minimum cost using the Cloudsim.

General Terms
Cloud computing, scheduling algorithm, task Scheduling.

Keywords
Cloud computing, Scheduling algorithm, Task scheduling, deadline, budget cost.

1. INTRODUCTION
Cloud computing is one of the popular and latest leading technology. Cloud computing is a large scale distributed computing paradigm in which a pool of abstracted, virtualized, dynamically scalable and services are delivered on demand to external customers over the internet. Here cloud consists of set virtual machine which include both computational and storage facility [1].

Cloud computing provides three important services. They are infrastructure as a service, platform as a service and software as a service. These services are availablein pay per use on demand model.Scheduling is one of the most famous activity in cloud computing environment to increase the efficiency of work and performance of task. Task scheduling is valuable concept which is greatly affects the behavior of the performance of tasks [10].

Task scheduling is valuable concept which is greatly affecting the behavior of the performance of cloud service provider. Task scheduling algorithm is a method by which tasks are matched, or allocated to data center resources. The task scheduling algorithm solves optimization problem, such as minimize the makespan. Task scheduling makes suitable for execution of deadline because dynamics environment are affect the system that is load and performance. This method has a great advantage in terms of makespan they minimize them within deadline and user budget cost [10].

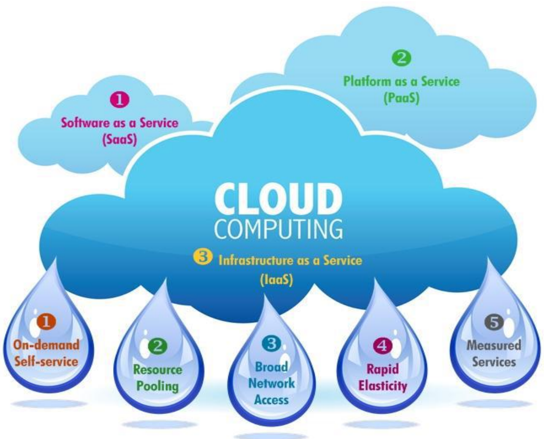

Fig 1:Cloud services

2. RELATED WORK
In 2011, P. Agrawal et Al focused on scheduling for major factor is makespan by considering the characteristics of cloud computing. The algorithm is energy aware scheduling for distributed system. The purpose of this algorithm is to minimize the execution time that means minimize the makespan and most important increase the energy level [5].

In 2014, M. Alejandra Rodriguez and R. Buyya proposed the PSO, here this algorithm consider factor is cost and deadline of task. The purpose of this algorithm is to minimize the execution cost within deadline. The particle swarm optimization to minimize the execution cost. Proposed algorithm takes replication of task in ideal time of virtual machine such that tasks are complete its execution before deadline [7].

Cloud computing used the rule based scheduling algorithm. In 2014, C. W. Tsai, W. Cheng Huang et Al proposed a hyper heuristic scheduling algorithm. This algorithm tries to improve the performance of the system. Also reduce the makespan. So here improving heuristic scheduling algorithm called as hyper heuristic algorithm. Heuristic technique is used to optimize the makespan and reliability of workflow application [3].

In 2015, Baoxian Zhang and senior member et Al proposed the nearly optimal packet scheduling algorithm. It is historically difficult NP hard problem. Proposed algorithm achieves the best optimal solution. This scheduling algorithm is reducing the execution time within deadline [2].

3. SYSTEM MODEL
In cloud computing, a scheduling optimization model is assumed that there is K tasks t $=\{t_1, t_2 \dots t_i, \dots t_k\}$and N resources r $=\{r_1, r_2 \dots r_j, \dots r_n\}$. When task$t_i$ $=(C_i, M_i, D_i, B_i)$.ThenC_i represents CPU usage, M_i represents

memory usage, D_i represents the deadline of task and B_i represents budget cost of user.

Each virtual resource in cloud is represented by two important parameters such as CPU and memory. Therefore, $r_j = (C_i,\ M_i)$ represents CPU utilization and memory usage of that resource. When scheduling these K no of tasks to N resource, the most important achievement is global optimal span. These all tasks should get complete before deadline and budget cost.

3.1 Costing of Resources

In this optimization scheduling algorithm calculate the cost of execution of tasks. When different task to different resource then cost is more reflect by using resource cost model and user budget costs. This resource cost model divided into two main part that is CPU and memory cost. Then the cost of CPU calculate by using following formula 1,$C_{cost}(j) = C_{base} * C_j * T_{ij} + C_{trans}$(1)

Here C_{base} is the base cost when a resource is used by the lowest utilization. $T_{ij}$is the duration time the task $T_i$runs in resourceR_j. C_{trans} is cost associated with CPU transmission. There are value for C_{base} = 0.17/ hour and C_{trans} = 0.005. Then next is cost of memory calculate by using following formula 2,

$$M_{cost}(j) = M_{base} * M_j * T_{ij} + M_{trans} \quad (2)$$

Similarly, $M_{base}$is the base cost when memory is 1 GB.$T_{ij}$is duration time of the task $T_i$runs in resource R_j. there are value of $M_{base} = 0.05$ and $M_{trans} = 0.50$. Cost function can be obtained as formulas 3 and 4,

$$C(j) = \sum_{j=1}^{n} C_{cost}(j), \quad (3)$$

$$M(j) = \sum_{j=1}^{n} M_{cost}(j) \quad (4)$$

Task is performing that time calculate the execution time and cost both are minimum. So minimize makespan and minimize the cost before time and cost by using following formulas,

This Multiobjective optimization problem can be described in formulas through 5-8.

$$\text{Minimize} \sum_x H(x) = F(x), B(x) \quad (5)$$

$$\text{s.t. } B(x) = C(x) + M(x) \quad (6)$$

$$B(x) \leq \sum_{i=1}^{k} Bi \quad (7)$$

$$F(x) \leq \sum_{i=1}^{k} Di \quad (8)$$

Here F(x) is a function of the performance objective referred as makespan; B(x) is the objective function of user budget costs.

3.2 System architecture

Figure 2 shows the system architecture of the PBACO algorithm. When user submit task for task manager then task manager accept the task and manage them. Scheduler is allocating tasks to resources using optimization scheduling. They judge the resource $r_j$allocating to requirement of the taskt_j. The global resource manager updates the information. Here calculate the duration time of tasks and also calculate the resource cost using this information from the resource model. Local resource manager manage the local virtual resource to CPU and memory load. When task is optimize and achieve global optimal path [1].

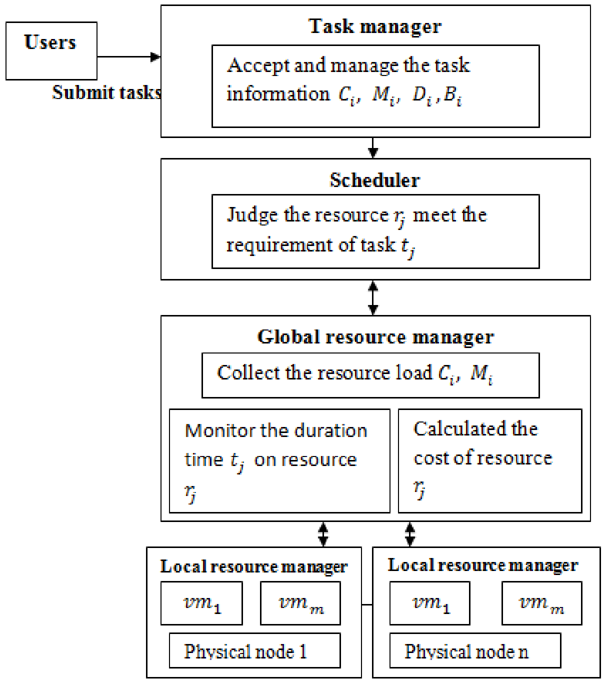

Fig 2: System architecture

3.3 PBACO scheduling algorithm

Algorithm completes the scheduling process by using the foraging process of real ants. The algorithm helps to complete the execution within deadline and user budget cost in public cloud. Multi objective optimization scheduling model is performing scheduling for task and achieves the optimal span, deadline, resource utilization and execution cost [1].

Ant colony optimization scheduling method

This scheduling method has following three main steps:

1. The transition probabilities of behavior choice.

The transition probabilities of behavior choice when tasks are scheduled using this optimization scheduling algorithm. Here number of task is successfully allocated to resource. This process is continuing to next task until all scheduling is completed. Then this step is farming path that means they finding the optimal path. So this choice formula assigning tasks to resources is important with pheromone and heuristic information. That time achieving the optimal solution. Task t_i to resource $r_j$is as formula 9,

$$P_k(T_i, R_j) = \left\{ \frac{[\tau(T_i,R_j)]\alpha\ [\eta(T_i,R_j)]\beta}{\sum_{h\epsilon gk(T_i,R_j)}[\tau(T_i,h)]\alpha\ [\eta(T_i,h)]\beta'} \right\}$$

Otherwise 0 (9)

Here $\tau(T_i, R_j)$ is pheromone of the task. $\eta(T_i, h)$ is the heuristic information of task. Then α and β is the weight factor of the heuristic information and pheromone [1].

2. The fitness function.

This Fitness function is used to describe the quality of feasible solution. This function is used to optimize the task. In this scheduling algorithm achieve the makespan and costs minimized. So fitness function of the formula 10,

$$\gamma e^{(-F(x))} + \delta e^{-B(x)} \quad (10)$$

Here γ and δ is weight factors of performance and cost. F(x) and B(x) are the performance and cost objective functions [1].

3. Updating pheromone.

Updating pheromone method is used to update the information. If fitness is high, the pheromone of the path should strongly allow more ants to find path. This process is executed until maximum number of iteration. Equation 11 is used to increase amount of pheromone [1].

$$\tau(T_iR_j) = (1-\rho) * \tau(T_iR_j) + \Delta\tau(T_iR_j) \quad (11)$$

$$\Delta\tau(T_iR_j) = \begin{cases} Q(\gamma e^{(-f(x))} + \delta e^{(-b(x))}), & (T_i, R_j) \in \text{path} \\ 0 & otherwise \end{cases} \quad (12)$$

Where ρ is the pheromone evaporation factor and $\Delta\tau(T_iR_j)$is the increment amount of pheromone. Q is constant. Multi objective optimization scheduling algorithm achieves good optimal path within deadline and budget cost. The PBACO algorithm is given below:

Algorithm: multi objective ant colony optimization scheduling method
Input:

$t_1, t_2, \dots\dots\ t_i, \dots\dots\ t_k, r_1, r_2, \dots..r_j\ \dots..., r_n, iter_{max}$

Output:

Map($t_{i,}r_j$)
BEGIN
Initialize ants' distribution amongr_j;
DO
FOR each ant do
FOR each *Ti* do
Select next route;
END FOR
Evaluate fitness of individual path by formula 10;
IF $r_j$meets the optimization problem Then
Output the map $(t_{i,}r_j)$;
Update pheromone along its path by formula
END IF
END FOR
UNTIL $iter_{max}$
END

Where, $iter_{max}$ is represent maximum number of iteration, which has been set to 100.

4. SIMULATION OF CLOUDSIM

Cloudsim is technique where a program models behavior of the system. Cloudsim is a new open source toolkit developed using java that generalized and advanced simulation framework allows simulation of cloud computing. Cloudsim toolkit supports both system and behavior modeling of cloud system components are supported by Cloudsim toolkit, such as data center, virtual machine and resource. So these simulation tool for creating cloud component environment and used as simulator solving task scheduling problem. When create datacenter with set of hosts and number of virtual machine as resource. That time optimize task and execute them.

4.1 Parameter Description

In this cloud environment is used the single data center. There were 100 hosts and 10 virtual machines. The parameter setup of VMs in the data center is range of 1860 MIPS to 2660 MIPS for CPU computing ability. The Virtual machine configuration is given below table 1,

Table1. The parameter setup of VMs in the data center

Parameter	Value
CPU computing ability	1860 MIPS, 2660 MIPS
RAM	4096 MB
Bandwidth	100 M/s
Storage	10 G

The cloudlet is input job or set of tasks to be executed in cloud environment. The parameter setup of tasks in data center is given below table 2.

Table 2. The parameter setup of task in the data center

Parameter	Value
Length	[400,1000] MIPS
File size	[200,1000] MB
Output size	[20,40] MB
Number of task	[100,600]

4.2 Simulation Description

Result analysis was conducted on Dell PC with Intel i3 CPU and 4 GB of memory running window 7 and Cloudsim 3.0. Cloudsim is used to 10 virtual machines in single data center. This task scheduling contains the number of tasks and these tasks are provided for scheduling. When we create the taskthat time every task find there execution time and calculate the total estimated time. Then after applying PBACO algorithm achieve actual time of this scheduling algorithm that type of execution is show the graph.

Figure 3 is shows the comparison of PBACO and FCFS scheduling algorithm. Then compare execution time of PBACO scheduling algorithm and execution time of FCFS scheduling algorithm. PBACO algorithm minimizes the makespan.

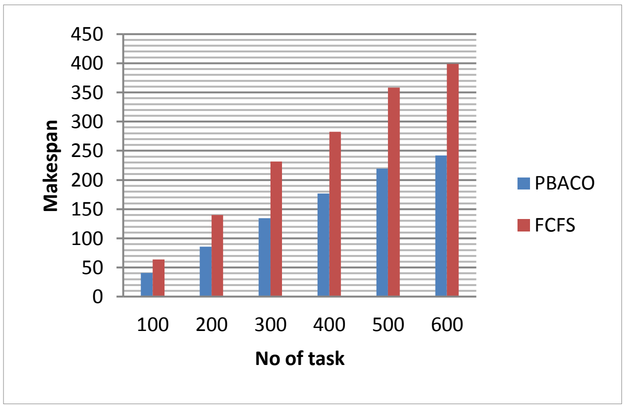

Fig 3: Actual time for PBACO and FCFS

Next graph shows the execution time of each task after which point makespan is calculated mean value. That means arrival Rate is 10 for all tasks i.e. 100 to 600.

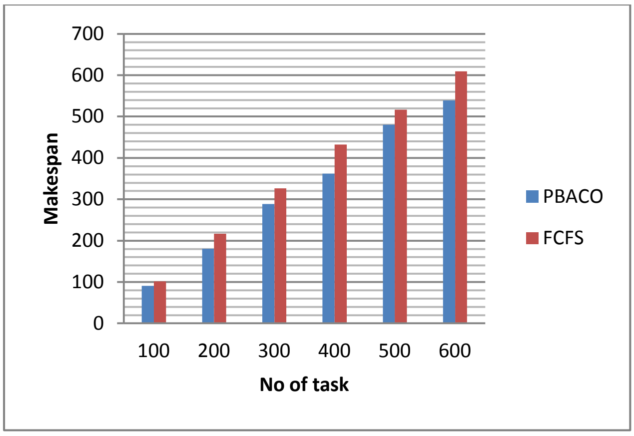

Fig 4: Arrival rates 10 for 100 to 600 tasks

Then same as next graph the arrival rate 80 for all tasks that is 100to 600. The execution time of each task after which point Makespan is calculated mean value.

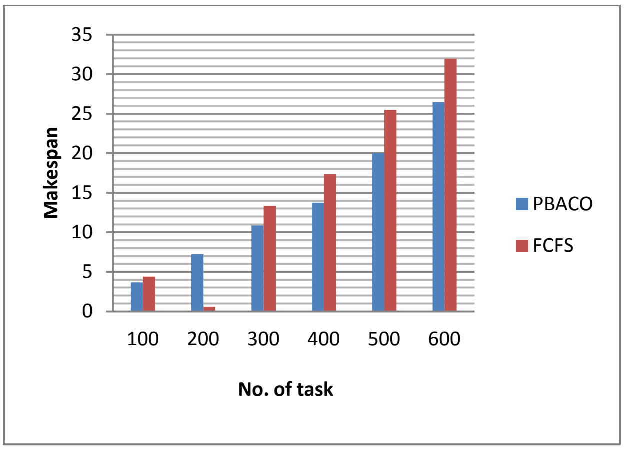

Fig 5 Arrival rates 80 for 100 to 600 tasks

Figure 6 is shows the comparison of PBACO and FCFS scheduling algorithm. Therefore compare execution total cost of PBACO scheduling algorithm and execution total cost of FCFS scheduling algorithm. PBACO algorithm minimizes the cost using resource cost model.

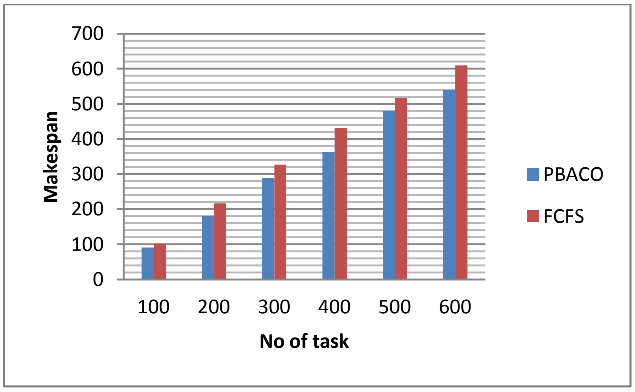

Fig 6: Actual cost of PBACO and FCFS

This graph show just load for resource that means how many task are assign to resource. This graph shows the task 200 and 600 respectively for resource utilization both algorithms. So here FCFS resources R4, R6, R9 are overloaded as compared to PBACO resources.

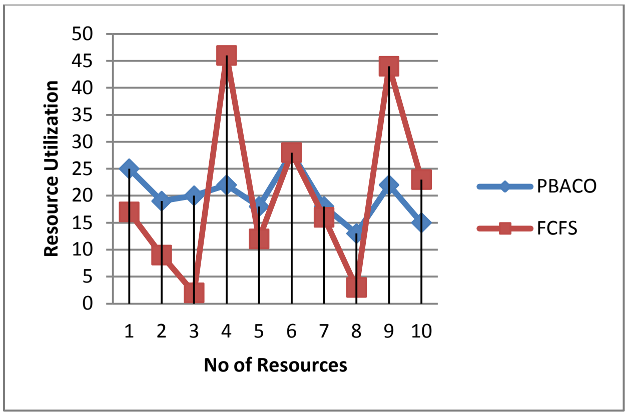

Fig 7: Resources utilization for having task 200

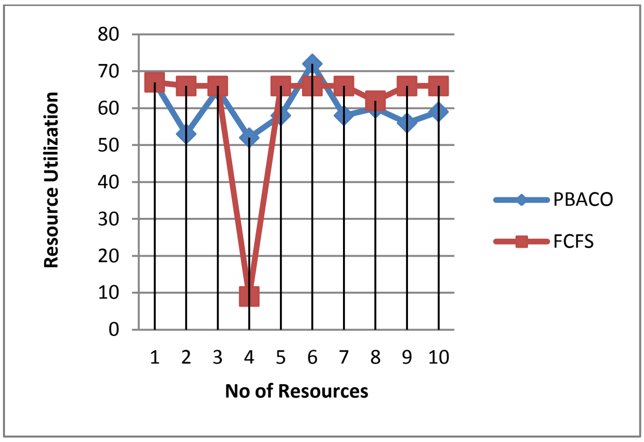

Fig 8: Resources utilization for having task 600

Deadline violation rate is scheduling the performance which obtains by response time and complete time of task. (Response time + Completion time). The deadline violation rate D_v is calculated as formula 13,

$$v = \frac{n_d}{K} * 100\% \qquad (13)$$

So, here n_d is number violating the deadline time in K task. This graph shows deadline violation rate of PBACO algorithm and FCFS scheduling algorithm. The deadline violation rates have been lower with a lower number of tasks. Deadline violation rates are tasks 200 and 600 respectively compare with both algorithms.

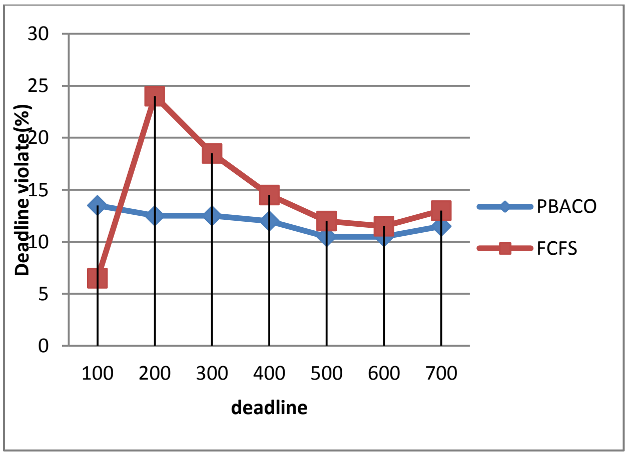

Fig 9: The violation rate of task 200

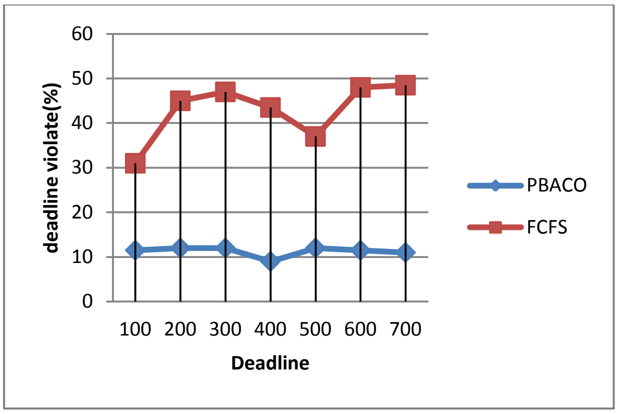

Fig 10: The violation rate of task 600

5. CONCLUSION

Multi objective scheduling method PBACO uses the Resource cost model. The PBACO algorithm is used to solve optimization problem. This algorithm Uses the performance and budget constraint to evaluate cost and provides the quality solution. The algorithm completes all task execution with minimum Makespan and minimum cost. This algorithm is to optimize the performance and user budget cost of task scheduling. Multi objective scheduling algorithm achieves the best optimal span, resource utilization, and completes task execution within deadline and budget cost.

In future work, by developing other than ACO algorithm compare with them. And also performance analysis is calculating using ACO algorithm.

6. REFERENCES

[1] L. S. S. D. Z. H. LIYUN ZUO, "A Multi-Objective Optimization Scheduling Method Based on the Ant Colony algorithm in cloud computing," Ieee Access, pp. 2687-2699, 2015.

[2] X. Wan, J. Luo, and X. Shen B. Zhang, "A nearly optimal packet scheduling algorithm for input queued switches with deadline guarantees," IEEE Trans. Comput., vol. 64, 2015.

[3] Wei-Cheng Huang, Meng-Hsiu Chiang Chun-Wei Tsai, "A Hyper-Heuristic Scheduling Algorithm for Cloud," IEEE TRANSACTIONS ON CLOUD COMPUTING, vol. 2, pp. 236-250, 2014.

[4] S. Hariri, and M.-Y. Wu H. Topcuoglu, "Performance-effective and low complexity task scheduling for heterogeneous computing," IEEE Trans. Parallel Distrib. Syst , vol. 13, pp. 260-274, 2002.

[5] P. Agrawal and S. Rao, "Energy-Aware Scheduling of Distributed Systems," IEEE Trans. Autom. Sci. Eng., vol. 11, 2011.

[6] G. Zhang, and W. Tan X. Zuo, "Self-adaptive learning PSO-based deadline constrained task scheduling for hybrid IaaS cloud," IEEE Trans.Autom. Sci. Eng, vol. 11, 2014.

[7] Maria Alejandra Rodriguez and Rajkumar Buyya, "Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds," IEEE TRANSACTIONS ON CLOUD COMPUTING, vol. 2, pp. 222-235, april- june 2014.

[8] Kurt Vanmechelen and Jan Broeckhove Ruben Van den Bossche, "Cost-Optimal Scheduling in Hybrid IaaS Clouds for Deadline Constrained worloads," in IEEE 3rd International Conference on Cloud Computing, 2010, pp. 228-235.

[9] Shahina K.K Sharfan P.S, "Adaptive Algorithm and a Genetic Algorithm for Minimizing Cloud Task Length with Prediction Error," International journal of modern trends in enginering and research, pp. 305-312, 2015.

[10] Xiaoyong Tang, and Keqin Li, Kenli Li, "Energy-Efficient Stochastic Task Scheduling on," IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS Heterogeneous Computing Systems, vol. 25, pp. 2867-2876, NOVEMBER 2014.

[11] M., and RajkumarBuyya Rodriguez Sossa, "Deadline based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds," IEEE Transactions on Cloud Computing, vol. 2, pp. 222-235, 2014.