A. Paper Organization

The rest of the paper is organized as follows. In Section II, we briefly present few relevant related works reported in the literature. In Section III, we discuss some important concepts of factors influencing the proposed system model required for understanding the solution. In Section IV, we theoretically prove the algorithm with different simulations and its comparisons with previous works. Finally, we conclude the paper in Section V by discussing, how this work can be extended in the future.

2. Related Work

Cloud computing is a rapidly growing area. Cloud Computing offers utility-oriented IT services to the users worldwide over the internet. Amandeep Verma and Sakshi Kausha, says that users submit their workflows along with some Quality of Service (QoS), constraints like deadline, budget, trust, reliability etc. for computation. In their concept, they consider the two constraints: deadline and budget. Deadline and Budget Distribution-Based Cost-Time Optimization (DBD-CTO) algorithm proposed the workflow scheduling algorithm that minimizes execution cost while meeting time frame for delivering results and analyze the behavior of the algorithm[1].

Alexandru Iosup & Simon Ostermann & Nezih Yigitbasi perform an empirical evaluation of the performance of four commercial cloud computing services including Amazon EC2, which is currently the largest commercial cloud that compares through trace-based simulation of the performance characteristics and cost models of clouds and other scientific computing platforms[12].

Qi Zhang, Lu Cheng and Raouf Boutaba suggests about the cloud computing infrastructure, service suppliers and users formation in resource market. They present a cost-based resource scheduling paradigm in cloud computing by leveraging market theory to schedule and compute the resources to meet user's requirements. They design an algorithm and protocol for cost-based cloud resource scheduling [10].

Ioannis A. Moschakis & Helen D. Karatza proposed a technique to use gang scheduling where a set of tasks are scheduled to execute simultaneously on a set of processors. Usually tasks are scheduled by user requirements. So, taking account of existing model, new scheduling strategies are proposed to overcome the problems of the performance unpredictability with the help of scheduling technique between user and resources [13].

Ayman g. Fayoumi develop a discrete event simulation to evaluate the performance with respect to the different load points. The performance metrics were the average waiting time inside the balance as well as the number of tasks. The performance study includes evaluating the chance of immediate serving or rejecting incoming tasks [11].

3. Simulation and Performance Analysis

3.1 Scheduling

Scheduling, whether it is located at Gateway1, Gateway2, Gateway3, is a core activity in mobile cloud computing that impacts the overall system performance and utilization. Due to the inherent dependencies between computation and data, scheduling workflow tasks is generally more difficult than scheduling for embarrassingly-parallel jobs. As stated before, most cloud scheduling approaches for workflows aim at single-shot workflow executions and only takes the account of simple constraints on time and costs. In each scheduled gateway cost can be calculated based on scheduling for every user. The model proposed, targets data-intensive workflows for continuous and incremental processing, also enforces constraints over the data communicated between tasks, while still fitting the utility paradigm. Our model implies that data must be shared via NoSQL database, which achieves better performance, scalability, and availability.

3.2 Execution Time

It specifies the maximum time a task can be on hold (without being triggered) since its last execution occurred. Considering $\theta(o)$ provides the time (e.g., seconds) passed since the last execution of a task, that is dependent on the availability of data in the object container o, this time constraint $\kappa\theta$ enforces that $\theta(o) < \kappa\theta$ at any given time. Scheduling agent will calculate the start time and end time from the user, for accessing the cloud. Each and every seconds of timing should be calculated for execution of data. When content size increases, there is a shifting from one gateway to another with higher bandwidth to secure the data loss during the transmission.

3.3 Cost Model

The cost model is based on considering task complexity and dynamic price definition. Assessing task complexity regarding processing and memory requirements has been explored with various works. Regardless of the approach employed, we can determine an estimation on how long each task will take to complete, with a given capacity awarded in the node (i.e., time = task complexity /worker capacity). In the general case where the infrastructure is shared by many users and workflows, the price of executing each task is calculated depending on the resources required. Then it is pondered with the overall system load.

Usually, the cost of executing a workflow for the first time will be the sum of the cost of executing its tasks. Additionally, the interval between consecutive executions of a given task can be significant, there is no point of paying (regardless of real money or some form of credits)in according to the common cloud cost model of Virtual Machine(VM) hours of execution, as these may be idle for the majority of time.

Therefore, we implement a service where task executions are incurred only for the time of execution, plus a tax of 10% to account for the overhead of reusing resources by switching among guest VM instances that execute different tasks, possibly from different workflows.

For every time while logging into the cloud, start time can be calculated for every seconds while using the cloud data instead of user data. Users data can be reduced based on the gateway fixed rate in which gateway the user logging in. If changing occurs to another gateway, for example, if the user logging into gateway1, the data cost reduction will be 0.001 rate per second. If data loading time exceeds more than the fixed time limit, automatically the user location changes to gateway2.

The data cost reduction time will be high than gateway1 but it reduces the loading time function. Each and every time when changing the gateway, it will take 120 kbps to transfer from one gateway to another gateway, at that time data cost reduction will happen.

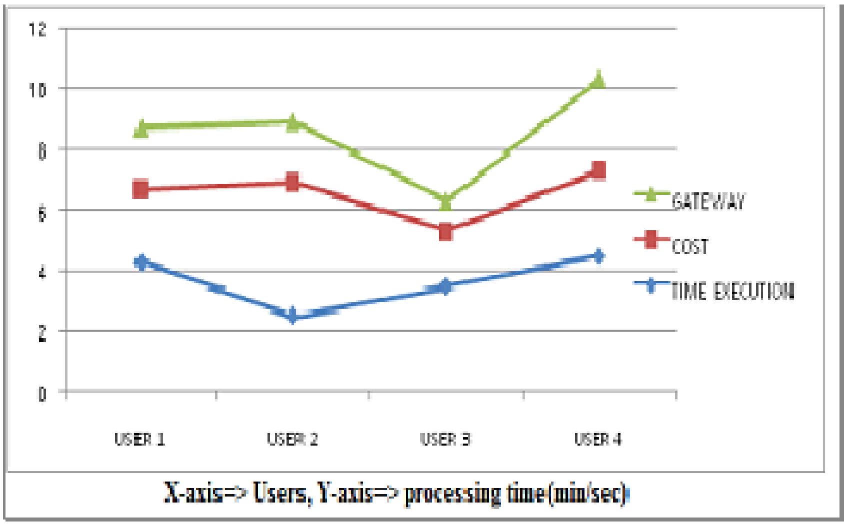

Figure 3.1: USER COST MODEL

Figure 3.1 shows the cost model of each user. The cost analyzer will calculate the total reduction of data cost (in terms of MB) based on under which gateway the user has logged in and how much time has been used for processing the cloud.

4. Greedy Algorithm

Based on greedy algorithm, we have proposed an algorithm which is suitable for dynamic heterogeneous resource environment connected to the scheduler through homogeneous communication environment. Greedy approach is one of the best approaches used to solve the job scheduling problem. According to the greedy's approach - "An algorithm always makes the choice that looks best at that moment. That is, it makes a locally optimal choice in the hope that this choice will lead to a globally optimal solution". Deadline Constrained Based algorithm is used to improve the completion time of tasks. The greedy algorithm is used with the aim of minimizing the turnaround task of individual tasks, resulting in an overall improvement of completion time.

We compared three different approaches with our algorithm: Greedy-time, Greedy-cost, and Random. Greedy-time selects the worker for each task that offers the minimum processing time at that moment. Similarly, Greedy-cost selects the worker at each step that offers the minimum processing cost. And greedy-random selects a random worker for each task.

Table 4.1: Time Factor Analysis

Type	*Usage of Data*
Internet browsing/surfing	5MB per hour
Emails	500KB per email (estimate includes attachments & spam)
File Uploading and downloading	Content size

Algorithm

```
Content size cs;
Fileupload st;
Fileupload ET;
Datacost Ds;
Foreach (upload file)
{
If (cs>1)
{
Ds= (et-st)*rate/persec;
}
}
```

Depending upon the content size (Cs) the loading time differs in the process. As soon as the user logged in and start to uploading the file , execution time starts (st). Ending time(et) of the process also counted which denotes the end execution time for every usage of cloud. Accessed start time and end time will be calculated by the cost analyzer. From that user accessing time for every second at particular gateway is estimated (rate /sec).

Table 4.2: Gateway Usage Data (cost /sec)

GATEWAY (kbps)	Rate /sec
Gateway 1 (1500)	0.001
Gateway 2 (2000)	0.002
Gateway 3 (3000)	0.003

The table 4.2 describes the gateway speed in kbps. Based on each gateways rate, data reduction for that corresponding user may be reduced in case of shifting to another gateway and also the data cost will be deducted from user on cost-time based analysis.

Table 4.3: Cost-Time Based Comparative Analysis

No.of Simulations	Gateway	Execution Time	Execution cost	Execution Cost in proposed model
1	G1	5.15	1.03	0.0515
2	G2	10.15	2.03	0.203
3	G3	12.45	3.05	1.375

Table 4.3 shows the difference between greedy algorithm and proposed changes in greedy algorithm. Time calculation and cost deduction factors are estimated. Comparatively, cost factor has been minimized using the proposed algorithm.

5. Conclusion

The mobile cloud computing is internet based computing in which resources are provided to users on demand, based on users dynamic location change. Workflow scheduling focus on mapping and managing the execution of interdependent tasks on shared resources that are not directly under its control. Cost and time reduction is one of the very important concerns to build up a cloud system for mobile users. Based on the file content size, gateway shifting, file uploading time and ending time factors, the cost can be calculated. We compared three different approaches with our algorithm: Greedy-time, Greedy-cost, and Random. Using proposed greedy algorithm data cost reduction can be minimized compared to previous model. In future, analysis based on the performance in terms of server usage can be extended based on individual user mobile devices. Also various factors and simulations can be estimated for upcoming heterogeneous network.

References

[1] Amandeep Verma and Sakshi Kaushal, "Deadline and Budget Distribution Based Cost-Time Optimization Workflow Scheduling Algorithms for Cloud", in the proceedings of International Journal of Computer Applications(IJCA), pp.1-4,2012.

[2] G. Malathy, Rm. Somasundaram and G.Vidhya "A Survey on Workflow Scheduling Algorithms and Map Reduce Method in Cloud Environment", International Journal of Communication and Engineering (IJCE), vol.6, Issue 1, pp. 60-65, March 2012.

[3] Gajendra Singh Thakur, Ravindra Gupta, Shubhra Mukharjee "A Survey on Cloud Computing and its Services", International Journal of Science, Engineering and Technology Research(IJSETR), vol.1, Issue 1, pp.17-20, July 2012.

[4] Mladen.A.Vouk, "Cloud Computing-Issues , Research and Implementation", Journal of Computing and Information Technology(JCIT), pp. 235-246, 16 April 2008.

[5] Navjot Kaur,Tarnjit Singh Aulakh and Rajbir Singh Cheema "Comparision of Workflow Scheduling Algorithms in Cloud Computing", International Journal of Advanced Computer Science and Applications (IJACSA), vol.2, pp.81-86, 2011.

[6] O.M. Elzeki ,M.Z.Rashad and M.A. Elsoud "Overview of Scheduling Tasks in Distributed Computing Systems", International Journal of Soft Computing and Engineering (IJSCE), vol.2, Issue 2, pp. 36-39, Sep 2012.

[7] P.K.Srinivasan, "Time-Cost Scheduling Algorithm", International Conference on Computing and Control Engineering(ICCCE), April 2012.

[8] Weiwei Chen and Ewa Deelman, "WorkflowSim: A Toolkit for Simulating Scientific Workflows in Distributed Environments", The 8th IEEE International Conference on eScience, Oct 2012.

[9] Yogita Chawla and Mansi Bhonsle, "A Study on Scheduling Methods in Cloud Computing", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol.1, Issue 3, pp. 12-17, Sep-Oct 2012.

[10] Qi Zhang ,Lu Cheng and Raouf Boutaba, "A Cost-based Resource Scheduling Paradigm in Cloud Computing", 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 417-422, 2011.

[11] AYMAN G. FAYOUMI, (2011) "PERFORMANCE EVALUATION OF A CLOUD BASED LOAD BALANCER SEVERING PARETO TRAFFIC" Journal of Theoretical and Applied Information Technology, Vol. 32 No.1

[12] Alexandru Iosup & Simon Ostermann & Nezih Yigitbasi (2010) "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE TPDS, MANY-TASK COMPUTING,

[13] Ioannis A. Moschakis & Helen D. Karatza , (2011) "Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System with Job Migrations and Starvation Handling", IEEE