A Novel Technique for Workflow Computation with Ant Colony Optimization in Cloud Computing Environment

Sakshi Arora[1], Amandeep Kaur[2], Dr. G.N Verma[3]
[1,2]*Department-CSE M.Tech,* [3]*Principal-SSIET, SSIET, Derabassi*

Abstract- Cloud Computing gave a new direction for the betterment of IT industry. It provides services over Internet according to pay per services use. Cloud computing environments facilitate applications by providing virtualized resources that can be provisioned dynamically. The advent of Cloud computing as a new model of service provisioning in distributed systems, encourages researchers to investigate its benefits and drawbacks in executing scientific applications such as workflows. There are a mass of researches on the issue of scheduling in cloud computing, most of them, however, are focused on workflow and job scheduling. Cloud Computing for task scheduling which guided by broker and computer by virtual machine, but cost of computation will increase if use random task localization and this problem exponential high. When use workflow which have huge number of task with dependency, so in this thesis optimize the cost and time by ACO, which give decision to broker. For this purpose, an algorithm based on both costs with user task grouping is proposed. The proposed scheduling approach in cloud employs an improved cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud.

Keywords- Cloud Computing, Scheduling, Energy Consumption, Virtualization, Bandwidth Cost, Data Centre, Hybrid

I. INTRODUCTION

Cloud Computing is the modern technique for conveying and hosting services on the internet. This model plays an important role when started any new organization from add more resources according to the need. It is an advanced computing technology that provides the quality of service, reliable for user [33]. Cloud computing organization deals with large amount of data, the demand of computing power, need to increase system investment .It is one of the popular technology now a days in IT field. Many change in the computing industry due to the cloud computing. It is an extension of Grid computing and parallel computing. Cloud computing is a model for on demand network access to shared computing resource.[33] It is a pattern for distributed computing that delivers infrastructure, software as services and platform. It helps user applications provider of dynamic services using large scalable, secure, fast, data storage and virtualized resources over internet. The main goal of cloud computing that customers only use when need, and only pay for actually use. These resources available in cloud every time and to be accessed from the cloud any time, any location via the internet [17]. It is supposed to manage the execution of tasks, operations, virtual servers, virtual infrastructure as well as the back end hardware and software resources of cloud environment. To gain the maximum benefits from cloud computing, developers must design mechanisms that optimize the use architectural and deployment [10]. The hardware and software provided support to virtualization. Virtualization is depending on many factors such as operating system, software and hardware managing them in the cloud platform, no environment to do anything without physical platform. Virtual Machine plays an important role in cloud computing because whole work of cloud related to the virtual machine. Its advantages only get when it is connect to the internet that's why user can use the powerfully computing services. It provides services according to user requirement. Virtualization is providing technical support for cloud computing applications and virtualization technology. In the past few years , the cloud computing research and development group such as IBM, Google all the well-known IT companies launched a cloud computing.

Types of Services

Cloud computing architecture having three layers for the software which require on demand services over internet. Figure1. Shows different layer cloud architecture [8],[10],[33]

- Infrastructure as a service
- Platform as a service
- Software as a service

➢ **Infrastructure as a services**

It is a first layer of Service. This layer delivers software and hardware components as a services. Sometimes IaaS is called a hosted service. Its capability provided to the user of IaaS as raw storage space, network resources and computing that's why user can run and execute applications any software that they choose. Examples of IaaS include Flexi scale, Go grid, Amazon EC2.

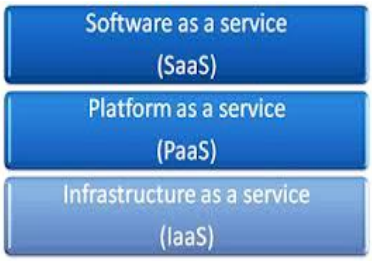

Fig.1: Types of services

➢ **Platform as a services:** It is a second layer of service. This layer uses the developer's application. The provider not only provides hardware, but also provides a toolkit and number of supported to higher level language.

- **Software as a Service(SaaS)**

These layers host the software and providing service through Internet. It reduces the maintenance cost of the customer. Cloud computing releasing their application on an entire environment that can be accessed through networks by application user. Examples of SaaS include Google mail, Google docs and so forth.

II. SCHEDULING PROCESS

Three stages of Scheduling process in cloud are as follows:

1) Resource Discovering and Filtering :-

It is a Data Centre Broker that Discover the resources i.e., present in the network system and collecting status information relating to them.

2) Decision stage:-

Target resource is selected based on certain parameters of task and resources.

3) Submission Task:-

Task is submitted to resource selected.

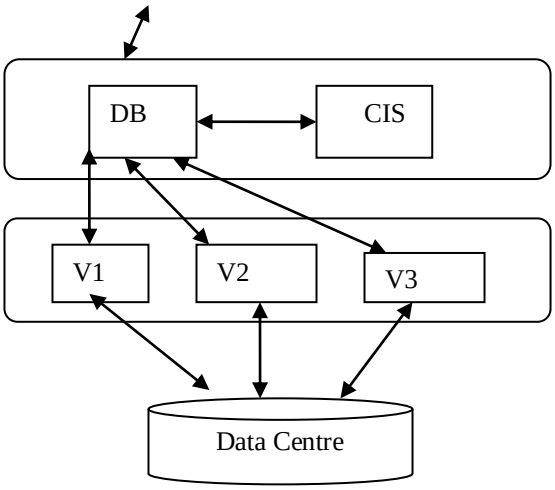

Fig.2: Scheduling Process

Ant Colony Optimization

ACO is the first algorithm to find the optimal path in a graph, based on the behaviours of ant. Most of the researches use ACO to solve NP hard problems e.g., traveling salesman problem, graph colouring problem and many other problems. These tasks include shortest path traversal to find food source information sharing with other ants by generating pheromone. The ant takes most of the time to travel down the path and back again, the more time the chemical have to evaporate .The shortest path selected then they get marched over the more frequently , and thus the chemical density becomes higher on short path than longer ones [3][18][27].

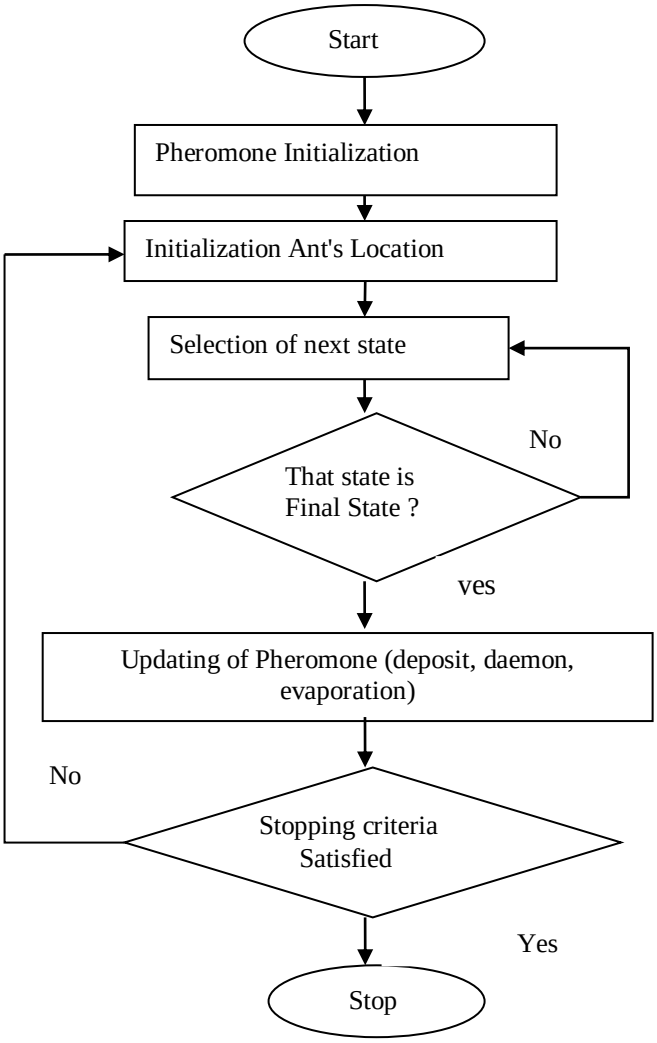

Fig.3: Flowchart of Ant Colony Optimization Algorithm

Particle swarm optimization

PSO is a sell adaptive global search based optimization technique. PSO has become popular simplicity and effective in wide range of application with law computation cost .The PSO algorithm is same other population based algorithm like genetic , but there is no direct link of individuals of the population .PSO algorithm assign task to the virtual machine in best but manner the task check the virtual machine and task is being assigned to the proper virtual machine that minimum wastage of memory .The objective is the minimize the cost of the execution application on cloud and also include time of processing transfeDrring ,transfer and process cost[10] [22].

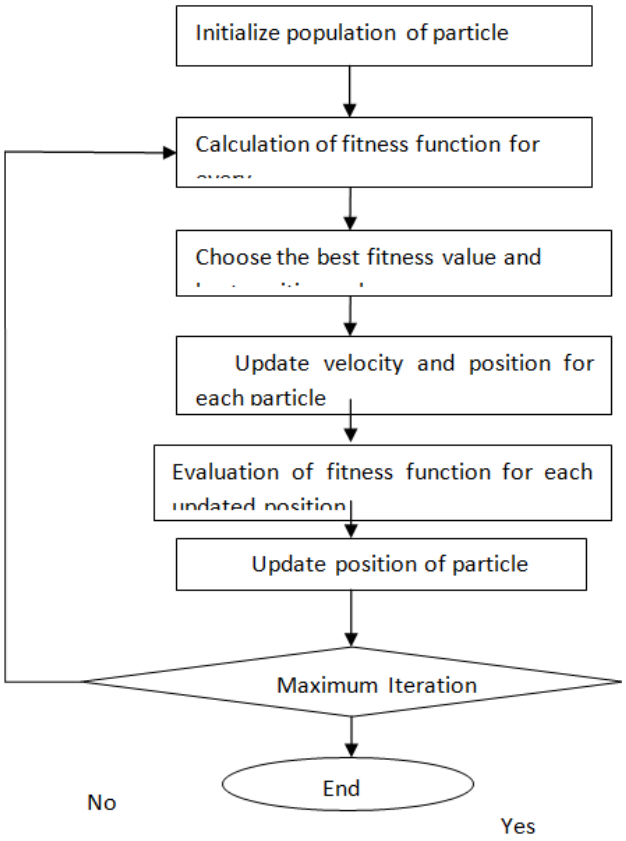

Fig.4: Flowchart of Particle Swarm Optimization

III. PROBLEM STATEMENT

The main problem which is identified through survey is that all algorithms focus on generating approximate or near-optimal solutions, it is impossible to generate an optimal solution within polynomial time and therefore workflow scheduling is NP hard problem. The solutions provided in literature survey are not optimal for utility like environments such as IaaS Clouds however; they focus into the challenges and potential for workflow scheduling. When planning the execution of a workflow in an IaaS Cloud environment. When planning the execution of a workflow in an IaaS Cloud environment there are two main opinions: (a) Resource Provisioning (b) Scheduling. In the resource provisioning phase the computing resources that will be used to run the tasks are selected and provisioned and in Scheduling, a schedule is generated and each task is plotted onto the best-suited resource. The selection of the resource and mapping of tasks is done in such a way that requirements of different user in terms of Quality of Service are met. Numbers of algorithms are proposed till now like PSO, ACO, GA and many more, to meet QoS parameter by scheduling large scale workflows. There are different aims of Resource provisioning and scheduling heuristics. Our effort focuses on scheduling the workflow on IaaS computing resources and finding an optimal solution by minimizing cost and execution time.

Proposed Solution

The principal idea of this research work is that in order to integrate some basic philosophies of Cloud computing such as the elasticity and heterogeneity of the computing resources and to meet the user's Quality of Service requirements on Infrastructure as a Service (IaaS), there should be resource provisioning and scheduling strategy for scientific workflows. In order to minimize cost and execution time, PSO algorithm was proposed but in some cases it fails to produce schedules with worse make span and greater costs as long as the deadline is met [14]. Hybrid (ACO) was used in Mobile ad-hoc network and gives better presentation in terms of delay, power, consumption, and communication cost as compare to standard ACO [25]. Therefore for better results if we use ACO optimization algorithm in an IaaS Cloud an efficient execution plan can be designed in order to reduce make span and pay a lower price.

IV. METHODOLOGY

1. Define the workflow Different workflows from different scientific areas are chosen and simulated on cloud environment.

2. Apply Ant Colony Optimization Algorithm: The algorithm Ant colony optimization (ACO) is a heuristic algorithm which is based on the behaviour of the ants seeking the shortest path between anthill and the location of food source. With the mechanism of positive feedback and distributed cooperation, it is proved to be a useful heuristic algorithm for solving NP hard problems.

3. Schedule Generation: Construct a schedule by converting a particle's position.

4. Evaluate Make span: Make span is evaluated to meet the given deadline.

5. Evaluate Cost: For each Workflow cost in terms of money is calculated.

6. Compare the result: Simulate and compare the result of proposed algorithm with PSO and ACO.

Proposed Algorithm

Input: Workflows like Cyber Shake, Ligo, Genome, and Montage

Output: Optimal value of Scheduling the task

1. Input Workflow X
2. Topological Sort (X)
3. For i ϵ X
   {
   Provide VM randomly
   i ++ ;
   }
4. Initialize ACO parameter
   X_t^{best} Time taken by VM
5. For i ϵ X
   {
   Determine the Value C_{xt}^{best}
   Save X_t^{best}
   }

If $f(X_t^{best})^{new} \leq f(X_t^{best})^{old}$ then $X = (X_t^{best})^{new}$

Else $X^{min} = (X_t^{best})^{old}$ ……eq. 1

6. If X_t^{best} is not converge
7. Initialize

For i ϵ X
{
Initialize local velocity X^t
Initialize global velocity Y^t
}
For each task
{
Update velocity
Update VM
For i = 1, 2, 3……X
If ($X^t < X_t^{best}$)
$X_1^t = X_t^{best}$
}
For each task
{
If ($Y^t < Y_t^{best}$)
$Y_1^t = Y_t^{best}$
}

8. Compare local and global best to eq. 1
9. If $X_t^{ACO} < Y_t^{PSO}$

{
Then set threshold X_t^{ACO}
else
Y_t^{PSO}

10. VM update or migration according to Y_t^{PSO}
11. Terminate the process.

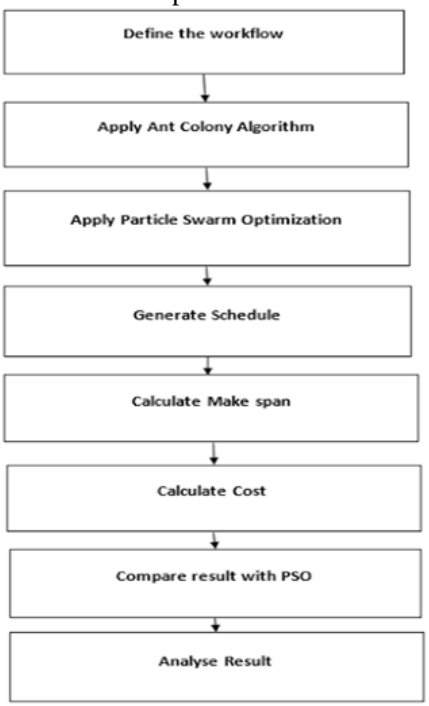

Fig.5: Workflow Diagram

V. RESULTS AND CONCLUSIONS

TYPE	PARAMETERS	VALUE
DATACENTER	Total no. of Data Centres	1
	Total Hosts	2
	Type of Manager	Time-shared
	Cost of Data Centre	1 to 10
VM	Type of Manager	Time-shared
	No. of Virtual Machines	1 to 5
	MIPS of PE	650
	No. of PE per VM	4
	RAM(MB)	512
CLOUDLETS/ TASKS	Total no. of Tasks	1-300
	Length of Tasks	5000-9000

Table 4.1 Parameter

Setting for Cloud simulator

Parameter Setting

To analyze and compare the proposed algorithm with PSO algorithm for scientific workflows on cloud [14] few parameters have been set to a pre-defined value. Various parameters of data Centre, virtual machines and cloudlets have been set to the different values. Number of iterations of different scientific workflows is analyzed and results are compared on two different parameters

(i) Cost (ii) time delay with PSO algorithm on IAAS cloud computing environment. Table 4.1 shows the parameter setting for cloud sim simulation toolkit.

modelled as an optimization problem which aims to minimize the overall execution cost while reducing the make span and the problem was solved using the hybrid of ACO and PSO. The experiments were conducted by simulating four well known workflows (Cyber shake, Ligo, Genome, Montage) on CloudSim , which shows that our solution has an overall healthier performance than other state-of-the-art algorithms. The worthy results are achieved because PSO (particle swarm optimization) play important role in global optimization and ACO(ant colony optimization) optimize locally and we have merge the two algorithms by taking the best out of them. With the proposed approach in most of the workflows we are able to produce lower cost efficient schedule meanwhile also reducing the time delay.

VI. RESULTS

				CYBERSHAKE		
VM	Cost(BASED)	Cost(Proposed)	Energy(BASED)	Energy(Proposed)	Time Delay(ACO)	Time Delay(Proposed)
two	0.45	0.29	112	100.2756	0.567	0.435
four	3.45	2.49	400	351.88	4.34	3.735
six	0.32	0.31	102	99.3966	0.578	0.465
eight	3.23	2.15	378.45	357.92	3.567	3.225
ten	2.45	1.73	345.34	336.3398	2.976	2.595

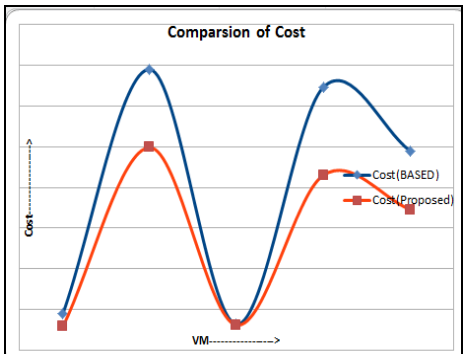

Fig. 4.2.1 Comparison of Based Cost and Proposed Cost

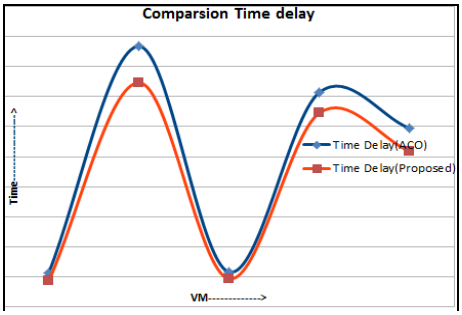

Fig. 4.2.2 Comparison of Time Delay(ACO) and Proposed Time Delay

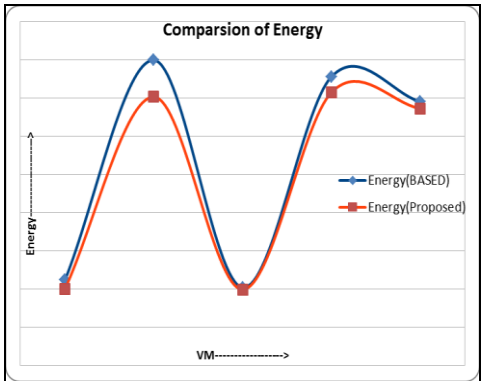

Fig. 4.2.3 Comparison of Energy(Based) and Energy(Proposed)

VII. FUTURE SCOPE

In this thesis we presented a scheduling strategy for executing scientific workflows on IaaS Clouds. The scenario was As future work, we would like to explore various options for the selection of the preliminary resource pool as it has a major impact on the performance of the algorithm. We would also like to research with different optimization approaches such as genetic algorithms and compare their performance with ACO. Another future work is to study the data transfer cost between data Centres so that VMs can be deployed on different locations. Finally, we wish to implement our approach in a workflow engine so that it can be utilized for deploying applications in real life environments.

VIII. REFERENCES

[1]. Alyaseri, Sana, and Alaa Aljanaby. "Population based Heuristic Approaches for Grid Job Scheduling." *International Journal of Computer Applications* 91.5 (2014).

[2]. Bardsiri, Amid Khatibi, and Seyyed Mohsen Hashemi. "A review of workflow scheduling in cloud computing environment." *International Journal of Computer Science and Management Research* 1.3 (2012): 348-351.

[3]. Bilgaiyan, Saurabh, et al. "Study of Task Scheduling in Cloud Computing Environment Using Soft Computing Algorithms." *International Journal of Modern Education and Computer Science (IJMECS)* 7.3 (2015): 32.

[4]. Binitha, S., and S. Siva Sathya. "A survey of bio inspired optimization algorithms." *International Journal of Soft Computing and Engineering* 2.2 (2012): 137-151.

[5]. Chawla, Yogita, and Mansi Bhonsle. "A Study on Scheduling Methods in Cloud Computing." *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 1.3 (2012): 12-17.

[6]. COMPUTING, FIREFLY ALGORITHM IN CLOUD. "A Load Balancing Model Using Firefly Algorithm in Cloud Computing." *Journal of Computer Science* 10.7 (2014): 1156-1165.

[7]. Devi, Poonam. "Implementation of Cloud Computing By Using Short Job scheduling."*International Journal of Advanced Research in Computer Science and Software Engineering* (2013).

[8]. Dillon, Tharam, Chen Wu, and Elizabeth Chang. "Cloud computing: issues and challenges." *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. Ieee, 2010.

[9]. Ernemann, Carsten, Volker Hamscher, and Ramin Yahyapour. "Benefits of global grid computing for job scheduling." *Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on*. IEEE, 2004.

[10]. Fazel Mohammadi, Dr, Shahram Jamali, and Masoud Bekravi. "Survey on Job Scheduling algorithms in Cloud Computing."2014

[11]. Goyal, Akhil, and Navdeep S. Chahal. "A Proposed Approach for Efficient Energy Utilization in Cloud Data Center." *International Journal of Computer Applications* 115.11 (2015).

[12]. Jafarzadeh-Shirazi, Omid, Gholamhosein Dastghaibyfard, and Mohammad Mahdi Raja. "TASK SCHEDULING WITH FIREFLY ALGORITHM IN CLOUD COMPUTING." *Science International* 27.1 (2015).

[13]. Karaboga, Dervis, and Bahriye Akay. "A comparative study of artificial bee colony algorithm." Applied Mathematics and Computation 214.1 (2009): 108-132.

[14]. Khadwilard, Aphirak, et al. "Application of firefly algorithm and its parameter setting for job shop scheduling." First Symposium on Hands-On Research and Development. Vol. 1. 2011.

[15]. Kumar, Santosh, and R. H. Goudar. "Cloud Computing–Research Issues, Challenges, Architecture, Platforms and Applications: A Survey." International Journal of Future Computer and Communication 1.4 (2012): 356-360.

[16]. Liu, Jing, et al. "Job scheduling model for cloud computing based on multi-objective genetic algorithm." IJCSI International Journal of Computer Science Issues 10.1 (2013): 134-139.
[17]. Liu, Xiaocheng, et al. "Priority-based consolidation of parallel workloads in the cloud." Parallel and Distributed Systems, IEEE Transactions on 24.9 (2013): 1874-1883.
[18]. Merkle, Daniel, Martin Middendorf, and Hartmut Schmeck. "Ant colony optimization for resource-constrained project scheduling." Evolutionary Computation, IEEE Transactions on 6.4 (2002): 333-346.
[19]. Miao, Yue. "Resource Scheduling Simulation Design of Firefly Algorithm Based on Chaos Optimization in Cloud Computing." International Journal of Grid and Distributed Computing 7.6 (2014): 221-228.
[20]. Mingxin, Wu. "Research on Improvement of Task Scheduling Algorithm in Cloud Computing." Appl. Math 9.1 (2015): 507-516.
[21]. Naik, Pradeep, Surbhi Agrawal, and Srikanta Murthy. "A survey on various task scheduling algorithms toward load balancing in public cloud." American Journal of Applied Mathematics 3.1-2 (2015): 14-17.
[22]. Pandey, Suraj, et al. "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments." Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on. IEEE, 2010.
[23]. Parsa, Saeed, and Reza Entezari-Maleki. "RASA-A New Grid Task Scheduling Algorithm." JDCTA 3.4 (2009): 91-99.
[24]. Radack, Shirley. "Cloud Computing: A Review of features, benefits, and risks, and recommendations for secure, efficient implementations." NIST, ITL Bulletin (2012).
[25]. Salot, Pinal. "A survey of various scheduling algorithm in cloud computing environment." IJRET: International Journal of Research in Engineering and Technology, ISSN (2013): 2319-1163.
[26]. Schopf, Jennifer M. "A general architecture for scheduling on the grid." (2002)
[27]. Shimpy, Er, and Mr. Jagandeep Sidhu. "DIFFERENT SCHEDULING ALGORITHMS IN DIFFERENT CLOUD ENVIRONMENT." algorithms 3.9 (2014).
[28]. Singh, Ashutosh Kumar, et al. "A Review of Cloud Computing Open Architecture and Its Security Issues." INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH 1.6 (2012).
[29]. SO, Kuyoro "Cloud computing security issues and challenges." International Journal of Computer Networks 3.5 (2011).
[30]. Susila, N., S. Chandramathi, and Rohit Kishore. "A Fuzzy-based Firefly Algorithm for Dynamic Load Balancing in Cloud Computing Environment."Journal of Emerging *Technologies in Web Intelligence* 6.4 (2014): 435-440.

Sakshi Arora, I have done my b.tech in CSE from Gurukul Vidyapeeth Institute of Engineering & Technology. I have done my M.Tech in CSE from SSIET, Derabassi. My Research work which I have done is about the workflow with ACO technique in Cloud Computing Enviornment which I have done is about the workflow with ACO technique in Cloud Enviornment.