# Study on Load Balancing Techniques in Ant colony Optimization for Cloud Computing

Ayushi Sharma
M.Tech (C.S)
Banasthali Vidyapeeth University
Jaipur, Rajasthan

Umang Singh, PhD
Assistant Professor
Institute of Technology &Science,
Mohan Nagar, Ghaziabad

## ABSTRACT
Cloud Computing has become interesting area among the IT industries. It is a collection of interconnected virtual machines as to serve pay per use model as per the demand and requirements of the users. The main aim of cloud computing is to provide efficient resources to each cloud user who are participating in the network. Cloud computing have various challenges like data sharing, load balancing, information retrieval, security etc. This attempt focuses on load balancing factor for cloud computing by using Ant Colony Optimization. Load Balancing is used to distribute the heavy workload among nodes to all other nodes in cloud environment. It is very helpful for cloud users to achieve proper resource utilization and also maintain fault tolerance using different algorithms like Min Max algorithm, ACO algorithm. In this paper, brief review an introduction of ACO algorithm for cloud computing is presented. Further this paper focuses on load balancing techniques algorithms and also discusses its architecture in depth. This paper would be beneficial for researchers of the domain.

## Keywords
ACO algorithm, Load Balancing, Cloud computing.

## 1. INTRODUCTION
Cloud computing is technology that are rapidly increase in terms of both academia and industry. Many researchers can research on this popular technology. But this technology is encouraged by industry than academic which determines its focus on user's applications [1].Cloud computing allows everyone to use software and computing services on-demand at anytime, anywhere and anyplace using the internet. This paper describe about the cloud computing challenges. A cloud is collection of various nodes which perform computation according to the client's request. The nodes of cloud can randomly provide to the client request so they can vary in quantity, thus the load of each node can vary [2]. Completion times of nodes in same cloud are different in which the underloaded nodes can take minimum completion time comparison than an overloaded nodes. This problem not only confirmed to particular cloud but it is related to every network like grid etc. In cloud computing there is a big issue of Load balancing when accessing the resources from cloud provider because multiple users can access a single resource at a time, the load will increase of particular node and performance will decrease. Various algorithms like Round Robin, Min Max Algorithm and Ant Colony Optimization algorithm are used for balancing the nodes. In this paper focus is on ACO algorithm for load balancing in cloud network.

## 2. LITERATURE REVIEW
In this section of paper the comparison of different Ant Colony Optimization Algorithms which are researched by different researcher [3] are discussed. By Shu-Ching Wang, Kuo-Qin Yan in their research paper "Towards a Load Balancing in a Three-level Cloud Computing Network", they suggested algorithms i.e. OLB (Opportunistic Load Balancing) that is used for utilize the better executing efficiency and LBMM (Load Balance Min-Min) that is used for maintain the load balancing of system. There are 3 phase i.e. Request Manager, Service Manager and service node. In the phase OLB algorithm is used to assign task to the service manager through the manager. In second phase LBMM scheduling algorithm is used to choose the suitable service node to execute subtask by service manager. In this research paper, using both algorithm it shows the available size of memory and transmission rate, the available CPU capacity, are the three factors for the duration of execution. In Cloud task Scheduling based on Load balancing ACO [4][5] the author proposed the LBACO(Load Balancing Ant Colony Optimization) algorithm, dynamically this algorithm is used to find the optimal resource allocation for each task. They also shown comparison with Basic ACO algorithm and FCFS algorithm and the result of nodes are balanced dynamically and tasks are mutually independent i.e. no precedence constraints between task and computationally intensive which is not realistic for systems in cloud. Main disadvantage of this algorithm is the heterogeneity of system. Heterogeneity of system issue is removed in that paper [6] using Load Balancing mechanism based on Ant colony & complex network theory. Using this algorithm it provides the good scalability of nodes and better fault tolerant system. But practically during run time environment the overhead increases and it gives poor response time, due to lack of synchronization of Ants, an ant encounters dead state at the end. Removing these issues in cloud network by this paper modified ACO framework algorithm.

It minimizes the throughput of heterogeneous computing system. In this algorithm it involved the basic pheromone updating formula. Using this pheromone table it gives the better utilization of resource but does not give the good fault tolerant system and ant move only one direction when updating the pheromone table. By this paper [7] "Load Balancing of Nodes in Cloud Using Ant Colony Optimization" the author proposed both direction i.e. Forward Direction and Backward direction. This algorithm can work fast because ant can move both directions for balancing the nodes in network. Main advantage of this algorithm is better utilization of resources. But this algorithm can run on limited parameter like speed, fault tolerance, network overhead and it take more power consuming , not consider the energy related issues.

# 3. PRELIMINARY WORKING CONCEPT

## 3.1 About Cloud

According to NIST, "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[1]. Cloud Computing provide basically 4 types of models and 3 types of services i.e. given below:

### *3.1.1 Cloud Computing Services:*

1. *Software as a services:*

Software is delivered as a service to the customer, which makes it unnecessary to have a physical copy of the software on all the devices of an organization. A SaaS provider deploys the software to the user on demand and this software is purchased by the license model. Customers can have access the software at all the times without worrying about deployment and maintenance of software [2]. Example: - Gmail, salesforce.com etc.

2. *Platform as a services:*

In this model, PaaS provides its services as a platform which means collection of set of software's and development tools hosted on the provider's servers.Tools, components and a development environment are delivered as a service to the customers, which help the customers to test and develop, operate web-based applications over the Internet. Example:- force.com, azure etc.

3. *Infrastructure as a services:*

Infrastructure as services (IaaS) is a single tenant cloud layer where the Cloud computing provider's dedicated resources are only shared with contracted clients at a pay-per-use fee. Infrastructure is delivered to the user by the provider over the network. Using this model, different software's including system or application software run by the user [1]. The user has the ability to provision computing power, storage, networks. Example:-rackspace.com, go grid etc.

### *3.1.2 Cloud Models:*

1. *Private Cloud:*

Privatecloud is only for particular organization it means the services of private cloud can accessible by only member of the organization [1].

2. *Public Cloud:*

Public cloud are provided its services both member of organization and other common public [2]. The users can be charged for the time duration they use the services. The services on public cloud are provided by proper authentication.

3. *Community Cloud:*

In Community Cloud, the cloud infrastructure is by several organizations and limited number of individual supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations) [1].

4. *Hybrid Cloud:*

Hybrid Cloud is combination of the previous three cloud deployments models.

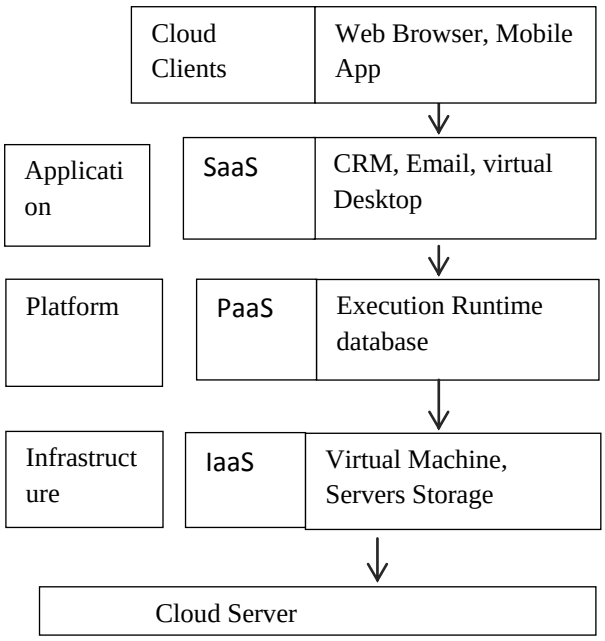


**Figure 1: Services of Cloud Computing [1]**

### *3.1.3 Working of Cloud Computing Architecture*

Cloud computing can classified into ends i.e. Front end and back end. Both are connected to each other by network, usually Internet. Front end is the computer user or any client which want to access the cloud services [1]**.**Front end is collection of applications and interfaces that are required to access the cloud computing platforms. Example: - Web Browser. The Back End of this architecture refers to the cloud itself. It consists of all the resources which are required to provide cloud computing services [2]. It is the collection of computer termed as servers, host a list of applications which are together called the back end and it comprises of huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc. Clients need to have the computer network and the application to access the cloud computing system. On the back end side, the various computers, servers and data storage systems that makes the "cloud" of computing services. Cloud includes every type of applications, from email and file storage to music and video games in most cases each app will have its own dedicated servers [1].The front end and back end of a cloud computing system use computer software, called middleware to communicate to each other A central server administers the cloud systems and follows a set of rules, called protocols,to ensure everything run smoothly.

.

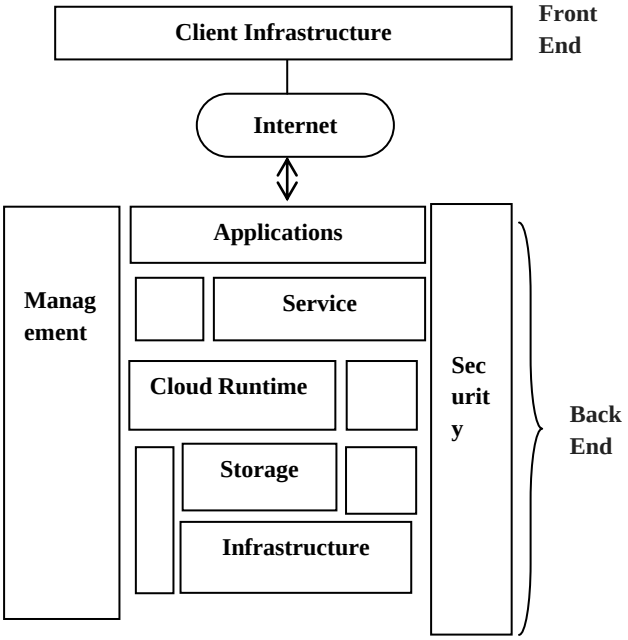


**Figure 2:Shows the graphical representation**

### *3.1.4 Issues of Cloud Computing*

There are various types of cloud computing issues that are given below [9]:

1. *Security and Privacy:*

It is the main issue of cloud computing because users can put his own data running his software on someone else's hard disk through someone else's CPU daunting to many. When moving data and application in cloud network different security issues arise like loss of control on data, heterogeneous nature of resources and various security policies [1].

2. *Performance:*

Performance is the second biggest issue in cloud network. When any client moves the any data and cloud infrastructure, cloud must be providing higher performance [9]. By capabilities of applications running on the cloud system can measured the performance of cloud network.

3. *Reliability and Availability:*

Resources which are provided to customer should be reliable; no disruptions (loss of data, code reset during execution) are arising. The best way to achieve reliability is redundant resource utilization. Availability can be understood as the client obtaining the resources whenever they are needed cloud should be providing these services at that time [1,9].

4. *Scalability:*

Scalability can be defined as the ability of the system to perform well even when the resources have been scaled up. Scalability can be achieved in two ways- horizontally and vertically whereby horizontal (scale out) scalability refers to addition of more nodes to the system like as adding a new system to an existing service provider system[10]. Vertical scalability (scale up) refers to addition of resources to a single node in the cloud system.

5. *Interoperability and Portability:*

Interoperability is the capability for using the same tools or application across different cloud service providers platforms. Cloud portability confirmed that one cloud solution will be able to work either other platforms, applications or with other clouds [10].

6. *Resource Management and Scheduling:*

Resources management is big challenge of cloud computing because multiple users can access same resources at a time. Resource Management is shows at various levels i.e. hardware, virtualization level with performance, software security and other parameters being dependent on the management [9]. It includes the management of memory, disk space etc. Job scheduling is a type of resource provisioning where order of jobs execution is established. Finishing of job execution is to optimize some parameters through response time, waiting time, turnaround time and resource utilization.

## 3.2 About Load Balancing

Load Balancing is the main issues of cloud network. Load Balancing is a technique that are used to distribute the load on different node in distributed server for better utilization of resources and avoiding the situation where few nodes are overloaded and few nodes are idle or under loaded [11]. Suppose that any node is overloaded and other node is idle or less loaded in cloud environment then Load Balancing are used to manage load and traffic by different technique like min-max algorithm, Round robin, Active clustering Ant colony optimization algorithm. Load balancing are two types that are below:

### *3.2.1 Types of Load Balancing:*

### *1. Static Load Balancing Algorithm:*

In this load balancing, load is distributed evenly across all the nodes and traffic are divided into different servers. In static load balancing prior knowledge of the system is needed so that decision of shifting the load is not depend on the current state of the system. Static load balancing is used for low load in cloud environment [11].

### *2. Dynamic Load Balancing:*

In dynamic load balancing, firstly the lightest server in the whole network or system is searched and preferred for balancing a load either node or group of nodes do the task of load balancing. It can take two forms: centralized and semi-distributed [12]. The major advantage of dynamic load balancing is that if someone node fails, it will not stop the system; it will only affect the performance of the system.

### *3.2.2 Working of Load balancing and its execution*

This fig shows how load balancing algorithm can manage the load balancing problem. First of all the job request and resources are come from different locations from different clients to cloudlet. Cloudlet is small cloud which are used to store these coming request [12]. Cloudlet can connect with Data center controller (DCC). DCC are used to processes the request and find the virtual machine to assign the jobs. After then Load balancer is used. Load balancer can distribute the workload from nodes to among different nodes in cloud network. DCC are proposed the algorithm for load balancing in cloud computing i.e. named as Load Balancing algorithm [13]. Load Balancing algorithm for managing the work load in cloud environment. Load Balancer make a table that contains (VMid) i.e. the id of the virtual machine, this table shows the priority of the virtual machine (VM) and state of the virtual machine either AVAILABLE or BUSY [14 ].

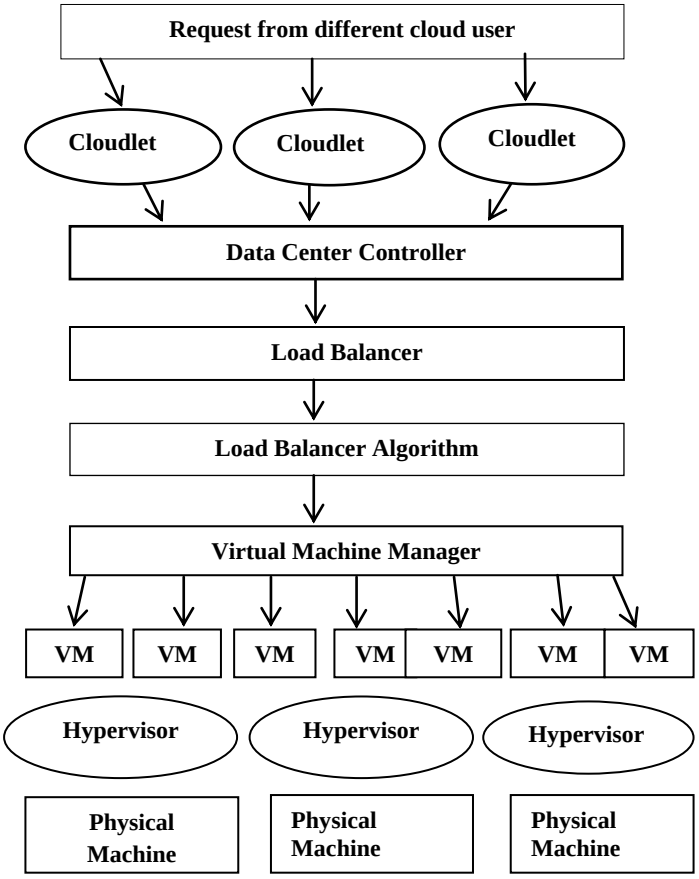


**Figure 3: Load Balancing and its Execution [21]**

If node is found:

- The id of the virtual machine (VMid) is returned back to the Data Center Controller. With the VMid, Data Center Controller assigns the job to the virtual machine. The Load Balancer can maintained table and also be updated about the new allocation.

If not found:

- The Data Center Controller will wait for a signal from the Virtual Machine which is sent if the Virtual Machine is available. These signals are obtained from the observer pattern used by VMs. On receiving the signal, the job will be assigned to that Virtual Machine. Virtual Machine will process the request and reply back with the response cloudlet to the Data Center controller [21].

DCC will reply back to the users who has sent the request also it will informed to the Modified Central Load Balancer about the de-allocation of the Virtual Machine after then it will updated the table. If there are any request will waiting to be served repeat this process again until all the request are served.

### *3.2.3 Challenges of Load balancing in Cloud*

Cloud Computing is widely used in today's environment but the research of cloud computing is in initial stage [15]. At present there are various scientific challenges yet to be resolved by scientific community some challenges are arise when resolved the load balancing issues that are given below:

1. *Automated service provisioning:*
   Elasticity is a key feature of cloud computing i.e. automatically releasing or allocating the resources in cloud environment. The challenge is using the optimal resources how the clouds elasticity can be used and still with the same traditional systems performance? [13]
2. *Virtual Machines Migration:*
   The concept of virtual machine migration is to visualize a machine as a set of file. The objective of virtual machine migration is to distribute the load of node in a datacenter and unloading a heavily loaded machine or system is possible done by moving virtual machine between them. When the virtual machine dynamically distributes the load bottleneck problem is arise in cloud computing, so the challenge is to remove the bottlenecks in cloud computing systems [11].
3. *Energy Management:*
   This is also key feature of cloud computing that allows everyone to use the resources from a global center rather than using their own resource. Energy Management may be termed as the economy of scale. [12].
4. *Stored Data Management:*
   Another major challenge is storage of datawhether that being individual data storage or company outsourcing its data storage. So now challenge is that, how can the data are distributed in the cloud network with minimum storage and fast accessible of stored data [13].

## 3.3 About Ant Colony Optimization

Macro Dorigo can introduce this concept i.e. Ant Colony in his Ph.D. in 1992. Ant have capability to find the path between nest and food. When ants find the path, they lay some chemical substances i.e. Pheromone and it is on ground. All the ants can follow this pheromone while finding the path [15]. The chemical can attract by the ants so all ants can follow the same path in which the path have high probability of the pheromone on ground for find the food and return to nest. By following the pheromone trails, the ant subsequently came to the food sources. The intensity of the pheromone can vary on different factors like the quality of food sources, distance of the food, etc. Ant's traversals from one node to another node using pheromone intensity and also update the pheromone trail of that path, so this path becomes more feasible if more ants follow this path [16]. Paths that have the highest pheromone intensity have the shortest distance between the point and the best food source. The movements of these ants independently update a solution set.

1. *Forward movement-*
   In this type of movement the ants move for extracting the food, or searching for the foodsources.
2. *Backward movement-*
   In this type of movements the ants after picking up food from the food sources traverse back to the nest for storing their food.

### *3.3.1 Working of ACO Algorithm*

ACO algorithm are used for efficiently distribution of load among nodes and also used for efficient utilization of resources in cloud network. By Kumar Nishantet all, in this paper first of all choose the RLBN node (Regional Load Balancing Node) by CCSP (Cloud Computing Service

provider) which will act as a head node. The selection of head node is not permanent if head node is not worked properly due to inevitable circumstances the new head node can be elected [7]. The head node is chosen in that way it has maximum number of neighboring nodes, as that can help our ants to traverse in different possible direction of the network of CCSP.

These ant traverses all over the network in that way they know about the location of both nodes i.e. underloaded nodes and overloaded nodes in network [19]. When these ants can traverse in network all ants can update the Pheromone table and this pheromone table can be used to store information about resource utilization for each node. There are two types of movements are performed by ants that are given below:

1. **Forward Movement**
   The ants continuously move in the forward direction in the cloud facing overloaded node or under loaded node.
2. **Backward movement**
   When an ant faces an overloaded node in its movement when it has previously faces an under loaded node then ant will go backward to the under loaded node and check if the loaded node is still under loaded or not [18]. If it finds this node is under loaded then it will redistribute the work to the under loaded node in network. The vice-versa is also possible.

On the basis of these movements ant can update the pheromone table of nodes. Each node has specification like number of resources, memory constraint, processing elements etc. All the information is stored in this pheromone table by ants. Updation of pheromone table aretwo types i.e. Foraging Pheromone and Trailing Pheromone.

### *3.2.1. Pheromone Updation*

Basically there are two types of pheromone updation are available for ants movement. Ant can update this table after performing each task by ant. The type of pheromone being updated by the ant would identify the type of movements of the ant and also tell about the kind of node the ant is searching for [7]. Two types of pheromone are given below:

#### *1. Foraging Pheromone:*

In ACO algorithm the ant would lay down FP after facing underloaded nodes for searching overloaded nodes. So after an ant going to anunder loaded node it will try to find the next path through foraging pheromone [7].

#### *2. Trailing Pheromone:*

In this algorithm the ant would lay down FP after facing under loaded nodes for searching overloaded nodes. So, after an ant going to an under loaded node it will try to find the next path through foraging pheromone [7].

## 4. SUMMARY

Cloud Computing provide its services and resources to the user. These services can access by different user form anywhere in the world. But providing these services the different type of issues are arise like load balancing, security issue etc. Load balancer are used to distribute the workload from one node to another node. So Ant Colony Optimization algorithm is used for solving this issue. This paper presents brief introduction of cloud computing and load balancing and also explain the working of Ant Colony Optimization algorithm for solving the load balancing issue

## 5. REFERENCES

[1] Book**:** Cloud Computing: Concepts, Technology and Archiecture by Thomas Eri.

[2] Klaithem Al Nuaimi, NaderMohamed, Mariam Al Nuaimi,1andJameela Al-Jaroodi, **"**Partial Storage Optimization and Load Control Strategy of Cloud Data Centers", Hindawi Publishing Corporation e Scientific World Journal Volume 2015.

[3] Wang S., Yan K., Liao W. and Wang S, "Towards a load balancing in a three level cloudcomputing network" 3rdInternational Conference on Computer Science and Information Technology, 108-113, 2010.

[4] Arabi E. keshk, Ashraf B. El-Sisi, Medhat A. Tawfeek, "Cloud Task Scheduling for Load Balancing based on Intelligent Strategy", I.J. Intelligent Systems and Applications, 2014, 05, 25 36 Published Online April 2014 in MECS.

[5] Kun Li, GaochaoXu, Guangyu Zhao, Yushuang Dong, Dan Wang, "Cloud Task scheduling based on Load Balancing Ant Colony Optimization", 2011 Sixth Annual ChinaGrid Conference by IEEE.

[6] Zehua Zhang and Xuejie Zhang, "A Load Balancing Mechanism Based on Ant Colony and Complex Network Theory in Open Cloud Computing Federation" by research gate.

[7] R. Rastogi Kumar Nishant, Pratik Sharma,, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization." Proceedings of the 14th International Conference on Computer Modelling and Simulation (UKSim), March 2012, IEEE, pp: 3-8.

[8] Qi Zhang · Lu Cheng · Raouf Boutaba, "Cloud computing: state-of-the-art and research challenges", 25 February 2010 / Published online: 20 April 2010 by Springer.

[9] Mohammad Sajid, ZahidRaza "Cloud Computing: Issues & Challenges", International Conference on Cloud, Big Data and Trust 2013, Nov 13-15.

[10] Rashmi 1, Dr.G.Sahoo2, Dr.S.Mehfuz3, "Securing Software as a Service Model of Cloud Computing: Issues and Solutions", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol.3, No.4, August 2013

[11] Shagufta Khan , Niresh Sharma , "Effective Scheduling Algorithm for Load balancing (SALB) using Ant Colony Optimization in Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014.

[12] PreetiKushwah ,"A Survey on Load Balancing Techniques Using ACO Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6310-6314.

[13] Aloksingh, Vikas Kumar Tiwari, Dr. BhupeshGour, "A Survey on Load Balancing in Cloud Computing Using Soft Computing Technique's", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 9, September 2014,

[14] Soumya Ray and Ajanta De Sarkar, "EXECUTION ANALYSIS OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING ENVIRONMENT", International Journal on Cloud Computing: Services and Architecture (IJCCSA),Vol.2, No.5, October 2012.

[15] DharmeshKashyap, JaydeepViradiya, "A Survey Of Various Load BalancingAlgorithms In Cloud Computing", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 11, NOVEMBER 2014.

[16] Book: Ant colony optimization by Macro Dorigo and Thomas Stutzle.

[17] Ant colony optimization theory: A survey Marco Dorigoa,∗,1, Christian Blumb.

[18] Hamid Mehdi Department of Computer, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran, "Optimization Task Scheduling Algorithm in Cloud Computing to Minimize Cost – ACO", International journal of Computer Science & Network Solutions Nov.2014-Volume 2.No.11

[19] Namrata Goswami1, Kunjal Garala2, PrashantMaheta, "Cloud Load Balancing Based on Ant Colony Optimization Algorithm", CONFERENCE PAPER · APRIL 2015