ISSN 2229-5518

Deadline and Cost Based Ant Colony Optimization Algorithm for Scheduling Workflow Applications in Hybrid Cloud

Lovejit Singh, Sarbjeet Singh

Abstract—Hybrid cloud is the best solution for organizations to scale their resources when demand increases. They store their confidential data on private cloud and can hire the instances of public cloud if there is brisk need of resources to handle the overload of organization operations. In this paper, Ant colony optimization algorithm has been proposed that schedule workflow applications in hybrid cloud. ACO finds the best schedule which results into reduction of execution cost while meeting the deadlines of workflow applications in hybrid environment. The performance of Ant colony optimization algorithm has been compared with other workflow scheduling algorithms such as genetic algorithm, min-min algorithm and max-min algorithm.

Index Terms— Ant Colony Optimization, Workflow Application, Scheduling, Private, Public, Hybrid Cloud.

---

1 Introduction

Cloud computing involves interconnection of data centres over the Internet. These data centres provide software, infrastructure and platform resources as services to customer on pay-as-you use basis. Customers do not need to buy physical infrastructure, rather they rent the usage from third party providers. Cloud computing allow users to access cloud services from anywhere in the world through web interfaces. The servers in the cloud are shared by multiple users and this provides benefits to customer as well as third party provider. The multiple users access these servers at different time instances and improve the utilization of data centres. Through centralization of resources the upfront expenditure of hardware as well as software is also reduced. The user can sign the contract with third party and policies of contracts are covered in Service Level Agreement (SLA) document. Cloud computing deployment model describes where the physical data centres are located and who manages them. There are three types of deployment models: private, public and hybrid. In private clouds the data centre's computing resources are hosted within the organization. The computing resources are managed by organization itself. In public clouds, the computing resources are provided to general public over the Internet. The customer does not invest in physical infrastructure and get the access of resources on rent by employing pay-as-you go model. Public clouds offer access to large pool of scalable resources to customers based on SLA (Service Level Agreement) documents. Hybrid clouds combine private clouds with public clouds through technology that enables the migration of applications from private clouds to public clouds. They share common API's (Application Programming Interface) that enable the hybrid operations.

- *Lovejit Singh is currently pursuing doctor of philosophy in department of computer science and engineering at Panjab University, Chandigarh, India. E-mail: pu.lovejitjhajj@gmail.com*
- *Sarbjeet Singh is currently working as Associate Professor in department of computer science and engineering at Panjab University, Chandigarh, India. E-mail: sarbjeet@pu.ac.in.*

Scheduling is a process through which tasks are assigned to computational resources at data centres. The data centre employs broker or scheduler to make scheduling decisions. The scheduler uses scheduling algorithms such as ACO (Ant Colony Optimization) to bind tasks with virtual resources. There are various types of cloud applications that are executed by cloud users at data centres. One of the cloud applications is workflow application in which tasks are linked as parent child relationships and are represented with directed acyclic graph (DAG). The DAG consists of nodes and edges. The node stands for the tasks and edges characterize relation of one task with other task. In some workflow applications tasks are executed level by level as shown in figure 1. The tasks at level 0 are executed before the execution of tasks at level 1.

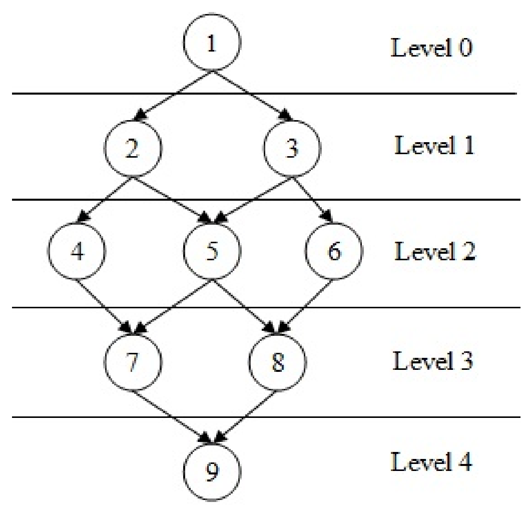

Fig. 1. Workflow Graph

Scheduling workflow applications in hybrid clouds is a challenging task. It needs the migration of dependent tasks from private cloud to public cloud in away that dependencies among tasks are preserved. In this paper, ant colony optimization algorithm is employed to handle workflow scheduling problem on hybrid clouds. The rest of the paper is structured as follow: The related work is presented in section 2. The problem overview is presented in section 3. Ant Colony Optimization algorithm at private cloud is presented in section 4. Ant Colony Optimization algorithm at public cloud is presented in section 5. Experimental results and comparison are presented in section 6 and Section 7 concludes the work carried out.

2 Related Work

In [1], Heterogeneous Earliest Finish Time (HEFT) algorithm was proposed to schedule workflow applications in heterogeneous computing environment. HEFT uses objective function to minimize the makespan of workflow application over diverse computing systems. The Deadline-Markov Decision Process algorithm in [2] partition the Directed Acyclic Graph (DAG) of workflow application and assign sub deadlines to each partition according to user defined deadline. It minimizes the execution cost of workflow application while meeting the user set deadlines. The MDP also support on-demand resources through single level SLA (Service Level Agreement). The hybrid cloud optimized cost (HCOC) algorithm was proposed in [3] which schedule workflow applications on hybrid clouds. HCOC schedules tasks on public cloud if the tasks fail to spot on private cloud. It considers sub-deadline of task, performance of resources and cost of resources of public cloud while migrating the tasks on public cloud from private cloud. The deadline and cost based workflow scheduling algorithm [4] was proposed in hybrid cloud, in which sub deadline is assigned to each task in workflow application and migrate the tasks to public cloud whose deadline is not met at private cloud.

The above scheduling algorithms share common characteristics that they all belong to heuristic technique of scheduling which may result into finding the solution which is locally optimal. None of them uses meta-heuristic such as ant colony scheduling algorithm to schedule workflow application in hybrid Clouds. Ant colony algorithm is an artificial intelligence based algorithm, where we simulate the ants moving from one food source to other. In [5] Ant Colony Optimization (ACO) was used to schedule workflow application in Grid which calculate pheromone value on multiple QoS parameters. In [6] the level based workflow application is scheduled using genetic algorithm. In [9] authentication mechanism is proposed that uses username and password to verify the authentication of cloud user. In [10] presents a comparative study of job scheduling strategies in grid computing. In [7] according to survey, there is a need to test the performance of meta-heuristic algorithm such as ant colony optimization in hybrid clouds. So, in this paper ACO algorithm is proposed to schedule level based workflow applications in hybrid clouds.

3 Problem Overview

Hybrid clouds combine the features of private clouds as well as public clouds. The user submits its workflow application for execution to scheduler/broker. The scheduler uses ant colony optimization algorithm to find the schedule for workflow application on private cloud. The ACO calculates the pheromone value based on MIPS rate of virtual machines at private cloud. If scheduler finds the schedule for workflow application within the user defined deadlines at private cloud then it binds the tasks with resources according to schedule otherwise it assign sub deadlines to workflow tasks and migrate them to public cloud. At public cloud, the ACO calculates the pheromone value of each virtual machine based on cost of execution. It selects those machines from public clouds which are cost effective and executes the workflow application within sub deadlines. Figure 2 shows the hybrid cloud architecture i.e. the combination of private and public cloud. These clouds are interconnected through Internet. Broker/Scheduler is responsible for allocating resources to workflow tasks. It uses ant colony optimization algorithm to achieve the objective of cloud service provider as well as cloud user.

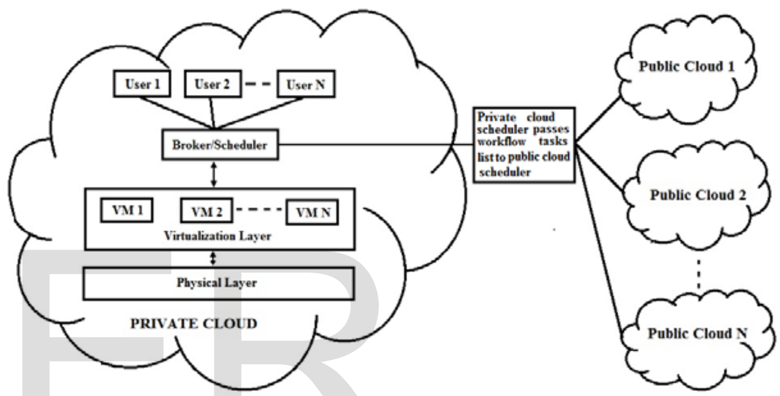

Fig. 2. Hybrid Cloud

4 Aco At Private Cloud

The proposed ant colony optimization algorithm finds the schedule at private cloud which results into achieving user defined deadlines. If the deadlines are missed at private cloud then sub deadlines are assigned to workflow tasks and migrated to public cloud.

Following steps are followed by scheduler to schedule the tasks at private cloud.

1. Let VM be the list of all the virtual machines on private cloud and T be the list of tasks that are to be assigned on the cloud.
2. Add the tasks in the list T according to DAG, which means tasks at level i in the DAG are added before tasks at level j in the list.
3. Tasks at the same level of the DAG are assigned according to their size; bigger tasks are arranged first.
4. At private cloud cost is not considered so machines are picked on the basis of MIPS.
5. Assume a colony has N ants. Start Ant Colony with first tour.
6. While there are ants in the colony repeat the following steps.
7. Start tour of ant A (k), where k=0, 1, 2....N.
8. Loop while there are more tasks in the list T.

ISSN 2229-5518

9. Pick task t (i) from the list T, where i=0, 1, 2.....n.
10. If this is the first ant in the colony, initialize pheromones value $\text{þ}_i\ (t)$ on the basis of MIPS of virtual machines.

$\text{þ}_i\ (t) = pe\_mips_i \times pe\_num_i$

Where pe_mips_i is MIPS rate and pe_num_i is number of processing elements in virtual machine i.

Choose a virtual machine V (j) randomly, while giving the preference to virtual machine with higher pheromone value.

11. Update local pheromone value on virtual machine.

$\text{þ}_i\ (t+1) = (1 - e) \times \text{þ}_i\ (t) + (1 / T_{et})$

Where e is evaporation rate $0<e<1$, $\text{þ}_i\ (t)$ is initial pheromone value on virtual machine i and T_{et} is execution time of task t on virtual machine i. T_{et} is calculated by using following formula.

T_{et} = *(Instruction length of task) / (MIPS rate of virtual machine).*

12. After all tasks are assigned and the tour of ant A(k) is ended. Update pheromones globally.
13. If all ants finished their tour, prepare a final schedule by picking the virtual machine with highest pheromone value for the task.
14. Compare the total finish time TF of the task with workflow application deadline D

If (TF<D) make this final schedule

Else

Assign sub deadline to all the tasks in the list by dividing workflow application deadline D on the basis of task's length.

15. Compare sub deadline of the task with finish time of the task.

If tf (i) > d (i) put task t(i) to the list that will be sent to public cloud, where tf (i) is the total finishing time and d (i) is the sub deadline of tasks T (i).

All the tasks not executed within the sub deadline are sent to public cloud.

5 ACO AT PUBLIC CLOUD

Receive the list of tasks for which sub deadlines are missed at private cloud. The scheduler allocates resources to these tasks at public cloud. It uses ant colony optimization algorithm for allocating virtual machines to these tasks. It finds the schedule that results into less cost and executes the workflow application within the sub-deadlines.

Following steps are followed by scheduler to schedule the tasks at public cloud.

1. Assume there are M ants on the public cloud. Start ant tour.
2. While there are more ants in the colony repeat the following steps.
3. Start tour of ant A (k), where k=0, 1, 2.....M.
4. Loop while there are more tasks in the list.
5. Pick task t (i), where i=0, 1, 2....N.
6. If this is the first ant in the list, initialize pheromone on the virtual machines. On public cloud the pheromone values are assigned on the basis of virtual machine cost. So, lesser will be cost of virtual machine higher will be the pheromone value.

$\text{þ}_i\ (t) = 1 / c_i$

Where c_i is the cost of virtual machine i.

7. Pick vm (j) randomly on the basis by giving preference to virtual machine with high pheromone value, where j= 1, 2.......K.
8. Find estimated finish time of the task on the virtual machine.

EFT = (Instruction length of task) / (MIPS rate of virtual machine)

If (EFT (i) < d (i)) assign task to this virtual machine.

Else

Find other virtual machine for the task that will satisfy the above condition. Assign task to the selected virtual machine.

Where, EFT (i) is estimated finish time of the task (i) on selected virtual machine and d (i) is the sub deadline of the task.

9. After assigning task on the virtual machine update local pheromone on the selected virtual machine.

$\text{þ}_i\ (t+1) = (1 - e) \times \text{þ}_i\ (t) + (1 / E_c)$

Where e is evaporation rate $0<e<1$, þ_i (t) is initial pheromone value and E_c is execution cost of task t on virtual machine i.

10. At the end of the ant tour, update pheromone globally.
11. After finishing ant's tour, pick the virtual machine with highest pheromone value for the tasks and make this final schedule.

6 EXPERIMENTAL RESULTS

The experiments are conducted using CloudSim [8]. Following configuration values were set when performance evaluation was carried out in hybrid cloud. Table I provides information about simulated resources which were used to perform experiments i.e. number of data centres, host configuration and virtual machine parameters etc.

TABLE 1
SIMULATION PARAMETERS

Number of Data Centres	5
Number of Hosts on each Data Centre	25
MIPS Rate of Host	10000
RAM of Host	65536 MB
Number of VMs on each Host	2
RAM of VM	512 MB
MIPS of VM	Varying from 200 to 4000
Mode of VMs on Host	Time-Shared
Total number of VMs	250

During experiments, five Data Centres were interconnected and one of them was private cloud and others were public cloud. The collaboration of private and public clouds make hybrid cloud network.

6.1 Comparison of Ant Colony Optimization with Other Scheduling Algorithms on the Basis of Cost in Hybrid Cloud

The performance of Ant Colony was compared with genetic algorithm, min-min Algorithm and max-min algorithm in hybrid cloud. Figure 3 shows that the ACO optimizes execution cost of workflow application better than genetic algorithm, min-min and max-min algorithm when it schedules the tasks to public cloud for which deadlines were missed at private cloud.

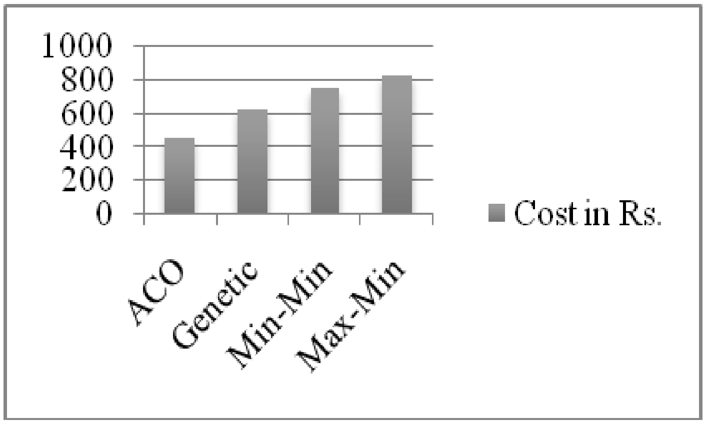

Fig. 3. Comparison of ACO with Other Scheduling Algorithms in Hybrid Cloud

6.2 Performance Evaluation of Ant Colony Optimization on Public Cloud and Hybrid Cloud

Ant Colony Optimization algorithm calculates the pheromone value of virtual machine in private cloud on the basis of MIPS rate. According to [3], the execution cost of private cloud is neglected because it is owned and maintained by organisation itself. Hybrid cloud organisations only pay for that tasks that skip deadlines on private cloud and need migration and execution on public cloud.

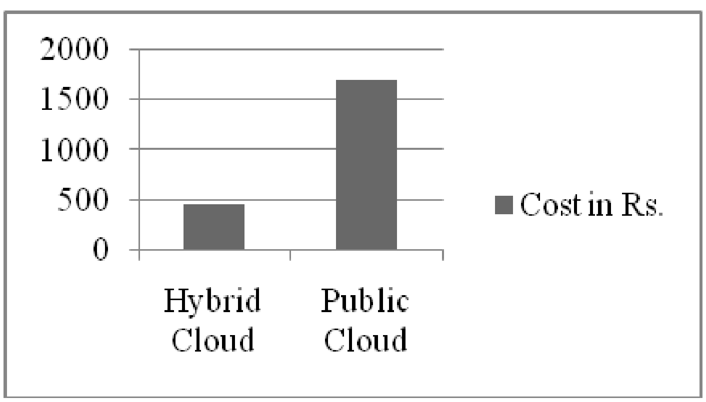

Fig. 4. Performance Evaluation of ACO on Public and Hybrid Cloud

In figure 4, the execution cost of workflow application using ant colony optimization algorithm was tested on both public cloud and hybrid cloud. There were 50 interdependent tasks in a workflow application which were executed on public as well as hybrid cloud.

7 Conclusion

In this paper meta-heuristic based ant colony optimization algorithm has been proposed to schedule workflow applications in hybrid clouds. User defines the deadlines of workflow application. The Ant Colony Optimization algorithm has been used to optimize the execution cost of those tasks whose sub deadlines were missed at private cloud and are shifted to public cloud to complete their execution within sub deadlines. The performance of ACO was compared with other meta-heuristic scheduling algorithms viz. genetic algorithm and heuristic scheduling algorithm i.e. min-min and max-min algorithm. The performance of ACO was also evaluated on the basis of execution cost in public as well as in hybrid cloud. In future, we can add other parameters while selecting resources such as reliability of virtual machine or other QoS parameter while assigning the tasks in hybrid clouds.

References

[1] H. Topcuoglu, S. Hariri, and M.-Y. Wu, "Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing", IEEE Trans. Parallel and Distrib. Sys. vol. 13, no. 3, pp. 260–74, 2002.

[2] J. Yu, R. Buyya, and C. K. Tham,"Cost-based Scheduling of Scientific Workflow Applications on Utility Grids", Int'l. Conf. e-Science and Grid Computing, pp. 140–47, July 2005.

[3] L. F. Bittencourt and E. R. M. Madeira, "HCOC: A Cost Optimization Algorithm for Workflow Scheduling in Hybrid Clouds", J. Internet Svcs. and Apps. vol. 2, no. 3, pp. 207–27, Dec 2011.

[4] Nitish Chopra and Sarbjeet Singh, "Deadline and Cost based Workflow Scheduling in Hybrid Cloud," International Conference on Advances in Computing Communications and Informatics (ICACCI), pp. 840-846, 2013.

[5] Wei-Neng Chen, "An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem with various QoS Requirements", IEEE Transactions On Systems, Man and Cybernetics, vol. 30, no. 1,, pp. 29-43, 2009.

[6] Lovejit Singh and Sarbjeet Singh, "A Genetic Algorithm for Scheduling Workflow Applications in Unreliable Cloud Environment", Recent Trends in Computer Networks and Distributed Systems Security Communications in Computer and Information Science Volume 420, pp 139-150, 2014.

[7] Lovejit Singh and Sarbjeet Singh, "A Survey of Workflow Scheduling Algorithms and Research Issues", International Journal of Computer Applications (IJCA), Vol. 74, No. 15, pp. 21-28, July 2013.

[8] Rodrigo N. Calheiros, Rajiv Ranjan, Cesar A. F. De Rose and Raj Kumar Buyya, "CloudSim: A Novel Framework for Modelling and Simulation of Cloud Computing Infrastructures and Services", GRIDS Laboratory, The University of Melbourne, Australia, 2009.

[9] Maninder Singh and Sarbjeet Singh, "Design and Implementation of Multi-tier Authentication Scheme in Cloud", International Journal of Computer Science Issues, Vol.9, Issue 5, September 2012.

[10] Simrat Kaur and Sarbjeet Singh, "Comparative Analysis of Job Grouping Based Scheduling Strategies in Grid Computing", International Journal of Computer Application, Vol. 43, No. 15, April 2012.