IJBPAS, November, 2015, 4(11), Special Issue: 451-464
ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)
'A Bridge Between Laboratory and Reader'

www.ijbpas.com

PROPOSING AN OPTIMAL TASK SCHEDULING ALGORITHM IN CLOUD COMPUTING USING DATA-AWARE SCHEDULING AND ANT ALGORITHM

MOHAMMADREZA FARAHPOUR HAGHANI, ENSIEH AZMAYANDE

Social Security organization, Rasht, Iran; **mr.farahpour@gmail.com**

Sama Technical and Vocatinal Training College, Islamic Azad University, Rasht Branch, Rasht, Iran; **e.azmayandeh@gmail.com**

ABSTRACT

Cloud computing provides users with a virtual combination of resources via internet. Considering the dynamism of accessible resources, task scheduling is one of the most important challenges in cloud computing. Good scheduling reduces task completion time, increases processor efficiency, and ensures user satisfaction. Efficient scheduling can be obtained by various ways. Considering that ant algorithm suits dynamic environments, we employed it as an optimal solution in the scheduling algorithm proposed in this paper in order to compute optimal virtual machine for each task. To measure the pheromone, we employed resource processing capacity, load balancing, and data-aware scheduling technique. This technique selects the most suitable virtual machine based on three factors of bandwidth, origin machine load, and destination machine load. The results obtained by Cloudsim simulator demonstrated that the proposed algorithm optimized task completion time and mean imbalance degree for different cloudlets compared with FCFS algorithm.

Keywords: Cloud computing, scheduling, task, ant algorithm, data-aware technique

INTRODUCTION

Cloud computing is a model of distributed computations composed of a large number of resources and requests, which shares resources on the internet as a service [1]. These systems utilize scheduling in order to increase task completion rate, increase resource efficiency, and consequently increase computational power. [2]. Task scheduling is a key process in service infrastructure of cloud computing environment, which efficiently executes the requests entering the system considering cloud environment specifications. Task scheduling considers virtual machines as scheduling units for allocating heterogeneous physical resources to execute the tasks. Each virtual machine is an example of cloud computing and storage capacities. [3].

Due to dynamism and heterogeneity of cloud computing environment, task scheduling is a very important challenge [4]. It is difficult to find a good solution with the capability to minimize execution time of each task, to reduce execution time of a combination of tasks, and to consider the load balance between resources. Below are some of the reasons [5]:

- The number of requests entering the system is far more than the number of resources and this situation is persistent.
- Characteristics of the requests entering the environment, such as arrival time, execution time and needed memory, are in continuous change.
- Cloud computing environment is a combination of heterogeneous resources.
- The resources have dynamic hardware and software characteristics such as load volume, free memory available on system, and communication network characteristics such as bandwidth and traffic.

The aim of task scheduling is to ensure minimum execution time, maximum processor efficiency, and load balance between all processors or a combination of them. Various solutions have been proposed in this regard, including greedy algorithm, circular shift algorithm, first-come first-served algorithm most-least algorithm, and least-least algorithm. These algorithms often neglect such criteria as efficiency rate, resource use, load balance, and quick response to requests. Task scheduling is one of the important challenges in cloud computing systems, because the aim of scheduling algorithm is to optimally schedule the tasks so that the tasks are allocated to the most appropriate resources in optimal time. Moreover, due to various reasons such as heterogeneity and dynamism of resources and requests, task scheduling in cloud computing environment is an Np-Hard problem [6].

Np-Hard problems can be solved by a variety of solutions. Considering that the aim of algorithm is to optimally schedule the tasks, ant algorithm can be an optimal solution to Np-Hard problems. Ant colony optimization algorithm has been inspired by natural ants living together in big complexes. This algorithm is a very efficient algorithm in mixed optimization problems. By developing and maintaining the connection of initial routes, this algorithm optimizes the problems relating to finding and keeping optimal routes in a communication network in cloud computing environment. Ant algorithm is an optimal method for ensuring the reduction of task completion time and maintenance of load balance of virtual machines. In this algorithm, resources are selected based on the amount of pheromone left by ants. By pheromone here is meant computational capacity of resources. In addition, it is recommended to employ data-aware scheduler technique in selecting the resources [7]. Considering the heterogeneity of cloud computing environment, data-aware scheduling can offer a good solution in heterogeneous environments. This scheduler may include data transmission, data allocation, data releasing, and data deletion. It is able to control a number simultaneous requests coming from each storage system. By offering numerous patterns regarding effective use of resources, management of memory spaces, and data replacement in end-to-end work process, this scheduler can offer task scheduling in heterogeneous environments with a good efficiency rate and load balance. [7].

To maintain load balance and minimize task completion time, the proposed algorithm, as inspired by ant algorithm, selects the optimal virtual machine for each task based on the amount of left pheromone. To measure the pheromone, we used processing capacities of resources, load balancing, and data-aware scheduler technique. Data-aware scheduler technique suits dynamic and heterogeneous environments and prevents the formation of long waiting queues for task allocation to virtual machine. This technique selects the most appropriate virtual machine based on three factors of bandwidth, origin machine load, and destination machine load [8]. One of the scientific-applied objectives of this study is to improve task algorithm in cloud computing environment with an inspiration from data-aware scheduler and ant algorithm. Ant algorithm is able to find and maintain optimal routes in a communication network in cloud computing environment. This algorithm offers a variety of methods to find the optimal route. In this paper we employ data-aware scheduler technique to select appropriate resources.

Other parts of this paper are as follows: part 1 reviews the literature; parts 3 and 4 delineates ant algorithm and data-aware scheduler; and parts 5 and 6 set forth the evaluation and analysis of the proposed method using Cloudsim software.

LITERATURE REVIEW

Cloud computing includes a large number of resources and requests and shares resources via the internet as a service [10]. Cloud computing system provides user with online commercial applications via browser or other software programs. Information and applications are stored on servers and provided to users according to their requests. In this architecture, data is mostly put on available servers in the internet and both applications and cloud servers run on user's browser [2].General architecture of cloud computing consists of five layers: software as a service (SaaS), platform as a service (PaaS), infrastructure as a service (IaaS) and server layer [9]. Users might feel the presence of first and second layers more, while other layers are also very important to service providers and developers [10-11. One of the key processes in service infrastructure layer is task scheduling. The aim of task scheduling is to efficiently execute the requests entering the system with consideration to other characteristics of cloud environment [11]. In the next part we explain task scheduling.

Task Scheduling in Cloud Computing

Task scheduling in service infrastructure layer considers virtual machines as scheduling units for allocating heterogeneous physical resources in order to execute the tasks. Each virtual machine is an abstract unit of computational and storage capacities in cloud environment [3]. Generally, scheduling is the process of allocating tasks to available resources based on task specifications and demands. Available resources should be used effectively without the influence of cloud service parameters [12]. Scheduling in cloud environment can be divided into three steps: resource discovery, information collection, and task scheduling [6]. The goal of scheduling is to minimize task completion time, maximize processor efficiency, and balance load between all processors or a combination of them. Various solutions have been presented for this purpose. One of the important challenges in cloud computing systems is task scheduling, because the purpose of scheduling algorithm is to optimally schedule the tasks so that the tasks are allocated to the most appropriate resources in optimal time. Moreover, due to various reasons such as heterogeneity and dynamism of resources and requests, task scheduling in cloud computing environment is an Np-Hard problem. [7]. These problems should be solved by explorative methods. Among the explorative methods, Ants algorithm is more suitable to Np-Hard problems. This algorithm is easy to implement and is more cost effective [13]. In the next paragraphs we explain Ants algorithm.

Ant Algorithm

Ant colony algorithm is an efficient algorithm which can solve mixed optimization problems in heterogeneous and dynamic environments. The main idea of this algorithm is as follows:

- In zero time, each ant randomly selects a node to start. In this step, the amount of start pheromone $\tau_{ij}(0)$ for (i,j) is measured.

- According to probability equation 1, ants select the next node. This is repeated until all nodes are investigated.

$$\rho_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum_{k\epsilon\, allowed_k}[\tau_{ik}(t)]^{\alpha}[\eta_{ik}(t)]^{\beta}} & k\epsilon\, allowed_k \\ 0 & \text{Else} \end{cases} \quad (1)$$

$\tau_{ij}(t)$ is the value of pheromone to yal (i,j), $\eta_{ij}(t)$ is the value of explorative data which equals $\eta_{ij}(t) = \frac{1}{d_{ij}}$. d_{ij} is the distance between node *i* and *j*. α and β are two parameters for controlling relative weight between pheromone and the value of explorative data.

- Updating the pheromone relating to each group
- Comparison the amount of pheromones and selection of the most optimal and the most effective route

In the proposed algorithm, which is inspired by Ants algorithm, computational capacity of resources is considered as pheromone. However, since cloud computing environment is dynamic and faces a large volume of requests, both processor resources and storage resources must be scheduled. In dynamic and distributed environments, data-aware scheduler is a good choice for scheduling storage resources. Therefore, we use data-aware scheduling technique to compute the probability of selecting the next resource.

Data-Aware Scheduling

Data-aware scheduler has been designed for distributed and dynamic environments. This scheduler recovers the data regarding available resources and loading information for each node. Based on this data, data-aware scheduling module selects the right node for scheduling different works [14]. Furthermore, this scheduler has been designed for data replacement and data movement, so such specifications as data transfer, storage place and data releasing and separation are included in the scheduling. This scheduler enables user to schedule processor resources and storage resources at the same time and also prevents the formation of long waiting queues for allocating tasks to virtual machine [8].

In data-aware scheduling, wherever there is a large volume of tasks, tasks are separated and the copy of the required data is saved in the same node. But since the number of data copies is not sufficient, copy management mechanism is also added. In copy management mechanism, data copies must not change after scheduling. In other words, scheduler is responsible for the copies developed in scheduling process. Copy management mechanism consists of the following steps [14]:

1. Time of development of a new copy of data
2. A method for selecting a new node for the maintenance of data copy
3. Pre-scheduling the related tasks
4. Duration of expired time of data copy

The selection of a new node is influenced by three factors: total network bandwidth, origin node load, and destination node load. [14]

1. PROPOSED ALGORITHM

As mentioned earlier, scheduling means finding an optimal solution to allocate user tasks to virtual machine, with minimum time between the start of the first task and the end of the last task. Tasks are shown by T= $\{T_1, T_2, \ldots, T_m\}$ ، and vms are shown by vm= $\{vm_1, vm_2, \ldots, vm_n\}$.

With two techniques of ant algorithm for selecting resources and data-aware scheduler for measuring the pheromone, the proposed scheduling policy is as follows:

First Step: Start

In this step, start pheromone value for each virtual machine is measured. In other words, ants are randomly placed on virtual machines for the first task. By pheromone here is meant total capacity of processing resources and appropriateness of the resource. The value of pheromone VM_j is measured by equation 2:

$$T_j(0)=num\_proc_j \times MIPS\_proc_j + vm\_transmission_j + suitable\_node\_dataaware_j \quad \ldots\ldots\ldots\ldots (2)$$

Num procj is the number of processors VM_j

MIPS procj is the efficiency of processor VM_j

Vm transmission is communicative capability of VM_j

Suitable node dataawarej is the value of appropriateness of VM_j which is measured by data-aware scheduler technique according to equation 3:

$$Z = Destination\ node\ load + origin\ node\ load + network\ bandwidth \quad (3)$$

Second Step: Mechanism of decision making and selecting virtual machine

Ant K selects a task from the list of outstanding tasks and allocates the task to VM_j with probability of equation 4.

$$\rho_j^k = \begin{cases} \dfrac{[\tau_j(t)]^{\alpha}[EV_j]^{\beta}[LB_j]^{\gamma}}{\sum[\tau(t)]^{\alpha}[EV]^{\beta}\,[LB_j]^{\gamma}} & j \in 1 \ldots n \\ 0 & \text{Else} \end{cases} \quad (4)$$

$\tau_j(t)$ is the value of pheromone VM_j in time t

EV_j is computational capacity of VM_j which is measured by equation 5:

$$EV_j = num\_proc_j \times MIPS\_proc_j + vm\_transmission_j \quad (5)$$

$[LB_j]^{\gamma}$ is load balance factor of VM_j, which is measured by equation 6:

$$LB_j = 1 - \frac{\Delta}{\overline{EV}} \qquad 0 \le LB \le 1 \quad (6)$$

$\overline{EV}$ is the mean of computational capacity of all virtual machines and EV_j is computational capacity of VM_j, which is measured by equation 7:

$$\overline{EV} = \frac{1}{m}\sum_{j=1}^{m} EV_j \quad (7)$$

Standard deviation is shown by Δ and equals equation 8:

$$\Delta = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(EV_j - \overline{EV})^2} \quad (8)$$

Sometimes virtual machines have too much load, which causes the development of bottleneck in cloud and affects task completion time. Hence, load balance factor LB_j is used in algorithm to improve load balance capacity. Thus virtual machine with highest load balance value is selected.

Third Step: Updating Pheromone

When each ant finds an optimal solution for each task, the value of pheromone of the selected virtual machine is updated by the best ant with the lowest completion time.

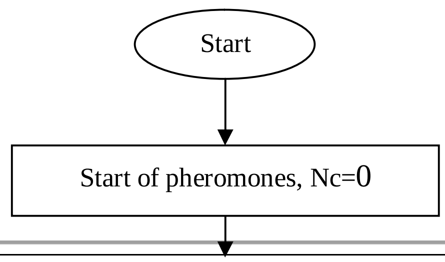
Since $\tau_j(t)$ is the intensity of pheromone VM_j is time *t*, pheromone is updated according to equation 9:

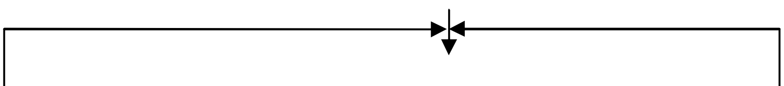
$$\tau_j(t+1) = (1-\rho) * \tau_j(t) + \Delta\tau_j$$

$\rho \epsilon (0,1]$ is pheromone reduction coefficient. The highest value equals ρ and the lowest value equals the value of pheromone of the previous solution. The value $\Delta\tau_j$ is defined as follows: When an ant completes its journey, local pheromone of the visited virtual machines is updated and the value $\Delta\tau_j$ is obtained by equation 10:

$$\Delta\tau_j = 1/T_{ik} \quad (10)$$

T_{ik} is the shortest route which has been found by *K*m ant in *i*m repetition.





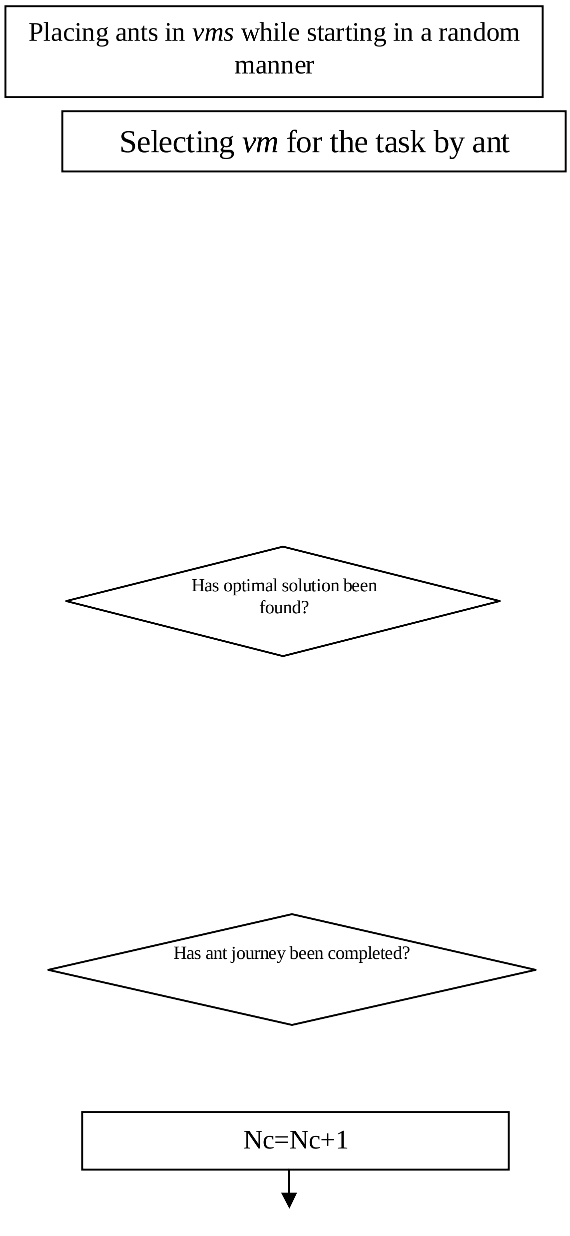


Fig1. Proposed Algorithm

Simulation of the Proposed Algorithm

There is a variety of simulation software programs which are able to simulate cloud processing and network processing. They are mostly based on Java and enable to simulate all aspects of these technologies. Cloudsim simulator is based on Java and is an open source software which can be used in Windows, Unix and Linux. Below are some important applications of Cloudsim: [1]

- Modeling and simulating data centers
- Simulating virtual machines and allocating different policies to them
- Dynamic incorporation of simulation elements; stopping, operating and resuming a simulation
- Simulating and modeling network configurations used in data centers
- Modeling federal clouds
- Showing how a cloud space is built

We simulated the proposed algorithm using Cloudsim in Linux operating system. Simulated algorithms consisted of the proposed algorithm and FCFS algorithm (scheduling policy adopted in default in Cloudsim environment). The goal of scheduling the proposed algorithm was to minimize task completion time so that load balance between virtual machines would be ensured. In the proposed algorithm, therefore, we had to investigate two factors: task completion time and the balanced distribution of tasks between virtual machines. Imbalance degree for the measurement of imbalance between virtual machines is defined according to equations 11 and 12:

$$T_i = \frac{total\_tasklenght_j}{num\_proc_j \times MIPS\_proc_j} \quad (11)$$

$total\_tasklenght_j$ is total length of tasks recorded in VM_j

num_proc_j is the number VM_j

$MIPS\_proc_j$ is the efficiency of each processor of VM_j

$$DI = \frac{T_{max} - T_{min}}{T_{avg}} \quad (12)$$

T_{max} and T_{min} show the highest and lowest values of T_i of T_{avg} of T_i mean of virtual machines.

First, a data center is established. Then, data center develops a combination of resources such as cpu, memory and bandwidth. Next, datacenter Broker is developed to simulate a Broker which interconnects the interactions between software layer as service and cloud provider. Broker operation is often in the side of this layer, which can find an appropriate cloud provider for allocating resources and services. This class sends the requests to datacenters for allocating virtual machines. Table 1 contains the parameters of cloud simulator. Table 2 contains parameters of the simulated algorithm

Table 1: Simulator Parameters

Type	Parameter	Number
DataCenterBroker	Number of datacenters	10
	Number of hosts	6-2

Virtual Machine (VM)	Total number of vms	50
	Ram memory of virtual machine	512-2048 megabyte
	Bandwidth	500-1000 byte
Tasks of cloudlet	Total number of tasks	100-500

Table 2- Parameters adjusted in simulator

Parameter	Number
Number of tasks	100-500
Number of repetitions	50
α	0.25
β	0.25
γ	0.5

α, β and γ are three parameters for controlling relative weight of pheromone and the value of explorative data and load balance, the total of which should equal 1. Since load balance is one of the goals of the proposed algorithm, we set its relative weight on 0.5.

RESULTS

In this part we show the results of the proposed algorithm and FCFS algorithm (the policy adopted in default in simulator environment). The goal of FCFS algorithm is to obtain the minimum completion time of each task, while the goal of the proposed algorithm is to select the optimal virtual machine for executing tasks, so that not only the reduction of completion time of all tasks but also load balance are ensured. To show the results, we employed a cloudsim based tool to model and analyze cloud computing environment. Unlike other simulation software which work with different programming languages, Cloud Analyst is a graphic simulation tool. In Cloud Analyst, simulation parameters can be adjusted by Wizard and simulation results can be seen in the form of diagram and tables. We executed these two algorithms for different cloudlets with the number of 100-500. Table 1 shows task completion time using Cloud Analyst software:

Figure 2 shows mean imbalance degree of two algorithms with 100-500 cloudlets. According to the Figures 1 and 2, the proposed algorithm is more efficient than FCFS. This means that the proposed algorithm has ensured system load balance and reduced general task completion time.

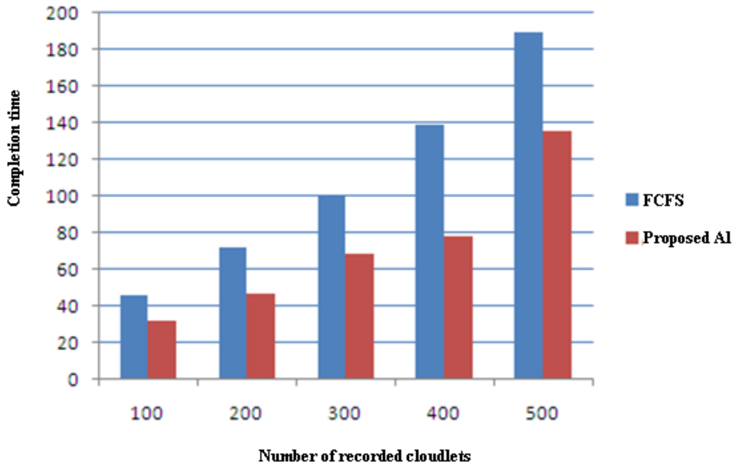


Figure 1: Mean task completion time (100-500)

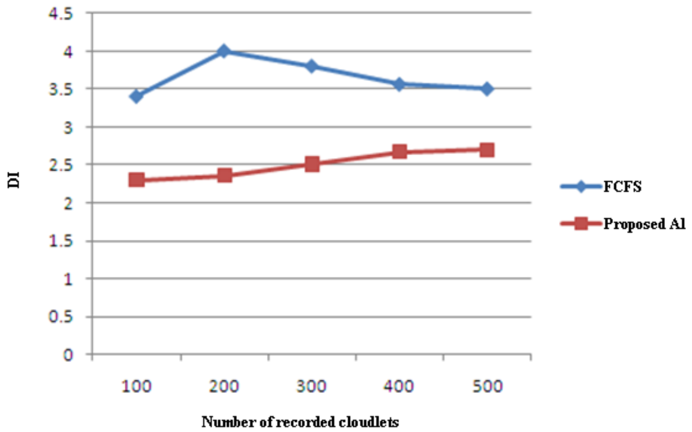


Figure 2: Mean imbalance degree of FCFS algorithm and the proposed algorithm

CONCLUSION

Cloud computing provides users with a virtual combination of resources via internet. Considering the dynamism of accessible resources, task scheduling is one of the most important challenges in cloud computing. Scheduling strategy is a computer science which has attracted much attention in the recent years. Many papers have been published to solve scheduling problem. In addition, various research fields, such as genetic algorithm and artificial intelligence, are seeking to solve scheduling problems. In this paper we addressed the issue of task scheduling in cloud computing environment. Scheduling is an Np-Hard problem. Np-Hard problems can be solved by a variety of solutions. Considering that ant algorithm suits dynamic environments and is able to solve Np-Hard problems, we used it as an optimal solution in the proposed scheduling algorithm. Ant algorithm, which is inspired by ant behavior, is able to select the optimal route between origin and destination resources. Ants can find the optimal route based on the left pheromone. The aim of the proposed algorithm, as inspired by ant algorithm, is to find optimal virtual machine for each task. The optimal virtual machine is selected based on the amount of left pheromone, so the amount of pheromone has to be measured. To measure the pheromone, we used resource processing capacity, load balance, and data-aware scheduler technique. Data-aware scheduler technique suits dynamic and heterogeneous and prevents the formation of long waiting queues for allocating tasks to virtual machine. This technique selects the most appropriate virtual machine based on three factors of bandwidth, origin machine load, and destination machine load. In the proposed algorithm, we explained how to compute optimal virtual machine based on ant algorithm and data-aware scheduler technique while taking load balance into considering, so that the algorithm can ensure load balance and reduce task completion time. Then we explained the simulation of proposed algorithm in Cloudsim environment. Finally, we compared the results of the proposed algorithm and those of FCFS algorithm and showed the results in the form of two diagrams. According to the results, the proposed algorithm reduced general task completion time and ensured load balance. The mean efficiency of the proposed algorithm is better than FCFS algorithm. This means that the proposed algorithm can ensure load balance and reduce task completion time.

Recommendations and Future Works

Considering the materials presented in this paper, a variety of topics can be recommended for study and research. Cloud computing system, especially in the field of scheduling, offers good areas and contexts for research. We hope that dear students and researchers pay special attention to the importance and application of cloud computing. Below are the related topics which can be studied in the future:

- Proposing an optimal scheduling algorithm with consideration to other explorative algorithms
- Improving task scheduling with consideration to service quality and explorative algorithms

REFERENCES

[1a] Rajkumar Buyya, chee shin yeo, Srikumar Venugopal, James Broberg, Ivona Brand,2008. cloud computing and emerging IT platform: Vision, hypr, and reality for delivering computing as 5th utility. Future Generation Computer System, ELSEVEIR, 2008.

[1b] Rodrigo N. Calheiros1, 3, Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose, and Rajkumar Buyya, 2012,. CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, The University of Melbourne, Australia : Grid Computing and Distributed Systems (GRIDS) Laboratory

Department of Computer Science and Software Engineering.

[2a] Junwei Ge, Bo Zhang, and Yiqui Fang, 2010. Research on the Resource Monitoring Model Under Cloud Computing Envirment. Springer.

[2b] Junwei Ge, Bo Zhang, and Yiqui Fang, 2010.Research on the Resource Monitoring Model Under Cloud Computing Envirment . Springer.

[3] Junwei Ma and Yueming Lu and Fangwei Zhang and Songlin Sun, 2012.Dynamic Task Scheduling in Cloud Computing Based on Greedy Strategy. Springer-Verlag Berlin Heidelberg.

[4] Zhao, L., Ren, Y., Sakurai, k., 2011. A Resource Minimizing Scheduling Algorithm with Ensuring the Deadline and Reliability in Heterogeneous Systems, International Conference on Advance Information Networking and Applications(ANIA), IEEE , 978-1-61284-313-1 , p: 275 – 282.

[5] HUANG Qi-yi, HUANG Ting-lei, 2010. An Optimistic Job Scheduling Strategy based on QoS for Cloud Computing. IEEE, 978-1-4244-6837-9/673-676

[6] Fatos Xhafa, Ajith Abraham, 2008. Metaheuristics for Scheduling in Distributed Computing Environments. Spain: Studies in Computational Intelligence,Springer-Verlag Berlin Heidelberg, 978-3-540-69260-7.

[7a] Y.J. Feng, L. Yu, G.L. Zhang , 2007. Ant colony pattern search algorithms for unconstrained and bound constrained optimization. Applied Mathematics and Computation, Published by Elsevier.

[7b] Y.J. Feng, L. Yu, G.L. Zhang, 2007.Ant colony pattern search algorithms for unconstrained and bound constrained optimization. Applied Mathematics and Computation, Published by Elsevier.

[8] Tevfik Kosar, Mehmet Balman, 2008. A new paradigm: Data-aware scheduling in grid computing. United States : Elsevier, 0167-739, 2008.

[9] Christian Baun, Jens Nimis, Stefan Tai, 2011.Cloud Computing. London : Springer , 978-3-642-20916-1.

[10] Jean-Paul Ebejer, Simone Fulle , Garrett M. Morris, Paul W. Finn, 2013. The emerging role of cloud computing in molecular modelling. Oxford Centre for Innovation, New Road : Journal of Molecular Graphics and Modelling, Elsevier Inc,Vol. 44.

[11] Paul and Sanyal G., 2012.Survey and analysis of optimal scheduling strategies in cloud environment.IEEE.

[12] Shi, Daji Ergu · Gang Kou · Yi Peng · Yong Shi · Yu., 2013. The analytic hierarchy process: task scheduling and resource allocation in cloud computing

environment. Springer Science+Business Media, DOI 10.1007/s11227-011-0625-1.

[13] Arūnas Andziulis, Dalė Dzemydienė, Raimundas Steponavičius, Sergej Jakovlev, 2011. Comparison Of Two Heuristic Approaches For Solving The Production Scheduling Problem. Information Technology And Control, Vol. 40.