

# Hybrid Workflow Management in Cloud Broker System

Dongsik Yoon<sup>(✉)</sup>, Seong-Hwan Kim, Dong-Ki Kang,  
and Chan-Hyun Youn

School of Electrical Engineering, KAIST, Daejeon 305-701, Korea  
{dongsik.yoon, s.h.kim, dkkang, chyoun}@kaist.ac.kr

**Abstract.** In Cloud broker system, workflow application requests from different users are managed through workflow scheduling and resource provisioning. In workflow scheduling phase, most existing algorithms allocate each task on certain VM in serial. In general, single task does not fully utilize allocated resource such as CPU, memory, and so on. When multiple tasks are processed with same resource in parallel, the resource utilization is improved that leads to saving the cost. In order to solve this problem, the Parallel Task Merging scheme in the same VM is proposed, which saves the cost of execution while satisfying SLA deadline. After workflow scheduling, VM resource provisioning is required. Auto-scaling VM resources approach is proposed, which adjusts the number of VMs while the number of requests varies. In this paper, we do experiment the parallel task merging and auto-scaling approaches on different environments to observe on which conditions these two approaches are working well or not.

**Keywords:** Workflow scheduling · Virtual machine allocation · Cloud resource provisioning

## 1 Introduction

In Cloud system, various scientific workflow applications can be processed with different goals. In general, lowering the total execution time of workflow requires higher cost, while saving the total execution cost of workflow results in longer execution time. Moreover, there are many other things to consider such as selection of resource (including CPU cores, memory, storage, and so on) types, ordering tasks in workflow to execute and Virtual Machine (VM) allocation of each task. To support these whole procedures, Cloud broker system was proposed to mediate between service users and resource providers. In Cloud broker system, guaranteeing the fairness between many users is additional challenge [1].

Workflow management by Cloud broker system can be categorized into two main parts; workflow scheduling and resource provisioning [1]. When users submit workflow applications with their Service Level Agreement (SLA) such as deadline or budget to the broker system, the workflow application is parsed into tasks connected with dependencies, which is represented as Directed Acyclic Graph (DAG). In workflow scheduling phase, the information of each task such as earliest start time, latest finish

time, historically earliest execution time or estimated execution time on each VM type should be obtained. Furthermore, each task should be ordered and VM type should be decided to be allocated to the task, while SLA is not violated. When the scheduling of each workflow is executed, available VM resource with proper type should be allocated to each task. However, in heterogeneous distributed computing system environment, such workflow task scheduling is generally NP-complete problem [2]. Since the arrival of each task is random and the status of resource is changing, dynamic scheduling is required. Furthermore, in resource allocation process, task affinity obtained by profiling should be used, which refers to the degree of suitability to certain computing resource. Such task ordering and allocating resource problems constitute the workflow scheduling, which is difficult to solve.

Scheduling of single workflow application has been studied by many researches. The main goal of the workflow scheduling is to reduce total execution time using limited resource or budget. One of the workflow scheduling algorithms is Heterogeneous Earliest-Finish Time (HEFT) algorithm [3]. It considers critical path to minimize makespan which is actual execution time on the distributed computing environment. The critical path means the longest path in the workflow. However, HEFT algorithm is a single objective scheduling considering only makespan, not the execution cost. Furthermore, this algorithm is a static scheduling that is hard to cope with unexpected results. Another workflow scheduling algorithm is IaaS Cloud Partial Critical Path (IC-PCP) algorithm [4]. It is multiple objective scheduling which considers both cost and time, but static scheduling algorithm. Additionally, Time Distribution (TD) Heuristics [5] is dynamic scheduling and multiple objective scheduling. It has a goal to minimize the cost while satisfying the deadline given by user. On task division phase, it classifies simple tasks and synchronization tasks which have multiple parent nodes or child nodes. On planning phase, all workflow tasks are allocated to proper resources based on Markov Decision Process (MDP) [6].

In addition, GAIN/LOSS approach [7] is the heuristic based workflow scheduling. All tasks are allocated on resources that minimize the makespan, then re-allocated by using GAIN/LOSS weight value. This approach does not guarantee the best optimization, however it has low time complexity and easiness to implement.

With the existing workflow scheduling algorithms, each request from users is isolated and allocated to separate VMs. In result, each task in a certain VM is executed in serial, not in parallel. However, in general, single task does not use full allocated resources such as CPU, memory and so on, which means that the resource on processing each task is not maximally utilized [8].

In this paper, we propose the Parallel Task Merging approach to improve the resource utilization in processing workflow applications, for saving the total execution cost. In our proposed scheme, multiple tasks are simultaneously processed on the same VM instance when certain conditions are satisfied. With this parallel task merging scheme, the utilization of resources can be increased, which results in saving the cost of VM resources.

When the workflow scheduling is completed, Cloud broker system should allocate each task to VM type decided in scheduling phase. To support this, Cloud broker system should monitor resource utilization and adjust the number of VMs. In this paper, we propose and utilize dynamic Auto-scaling cloud resource scheme. This scheme adjusts the number of VM resources according to workflow application requests.

The rest of the paper is organized as follows. Section 2 describes the details of the Parallel Task Merging approach in workflow scheduling phase and Auto-scaling approach in VM resource provisioning phase. Section 3 presents and evaluates the experimental results. Section 4 concludes the paper.

## 2 A Hybrid Model of the Parallel Task Merging and Auto-Scaling Algorithms in Cloud Broker System

### 2.1 Problem Description

When the workflow applications are executed in Cloud environment, the workflow scheduling and resource provisioning constitutes two main parts. Since the whole procedure of workflow management is considerably complicated for ordinary users, Cloud broker system was suggested as an intermediary [1]. The relationships between workflow service users, Cloud broker system and resource providers are presented as shown in Fig. 1.

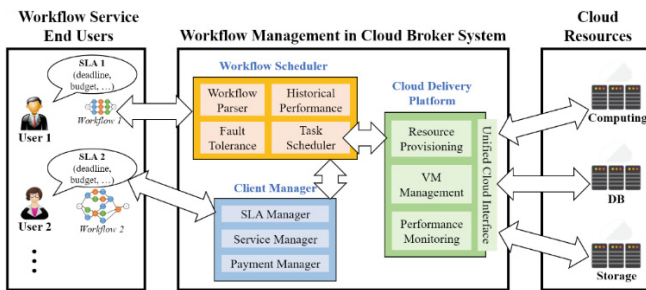


Fig. 1. An architecture of workflow management system in Cloud service [1]

At the beginning, a user submits tasks to be scheduled to Cloud resource broker with SLA constraints. The received workflow is parsed to individual tasks and dependencies between tasks in Workflow Management Module. The parsed workflow is then scheduled by VM allocator using workflow scheduling algorithm. When the workflow scheduling process is executed, the actual allocation of VMs to the given workflow is conducted by Cloud computing resource adaptor.

However, the existing workflow scheduling algorithms isolate each request from users and allocate tasks to separate VMs, while each single task usually does not maximally utilize allocated resources such as CPU, memory, and so on. Moreover, the resource pool management could be more efficient if VM requests are predicted and the number of VMs is adjusted, while keeping backup idle VMs to save the startup time.

### 2.2 The Parallel Task Merging and Auto-Scaling Scheme

To increase the utilization of VMs in a compact way, we propose the Parallel Task Merging algorithm. This approach merges multiple tasks on the same VM instance for

running simultaneously, if the VM cost is reduced and possibly increased execution time does not violate SLA deadline. To adopt this approach, it is essential to acquire the execution time of merged tasks on the same VM instance prior to the workflow scheduling processes. Despite such parallel task merging would result in increasing execution time of each task, it is worthy only if SLA deadline is still not violated because the VM cost would be saved. In reality, many tasks are not fully utilizing given resources such as CPU and memory, which makes parallel-task-merging possible.

A workflow application is represented as a directed acyclic graph  $G(Q, E)$ , where  $Q$  is a set of  $n$  tasks  $(q_1, q_2, \dots, q_n)$  and  $E$  is a set of  $m$  edges  $(e_1, e_2, \dots, e_m)$  [1, 4, 5, 8]. Each edge represents dependency between two tasks. The task which does not have any parent task is  $q_{start}$  and the task which does not have any child task is  $q_{end}$ . To adapt our approach, SLA deadline should be submitted with workflow  $G(Q, E)$  to be scheduled. When the scheduling is executed, the VM type among the set of VM flavor types  $FlavorSet$  to be allocated on each task  $q_i$  is decided, and each task  $q_i$  is allocated on proper VM instance, which is represented as  $VM_{q_i}$ . Such task scheduling and resource allocation is a hard problem, because of data dependencies between tasks according to constraint of the workflow topology.

To adapt parallel task merging approach, the workflow should be initially scheduled by some existing approach. In this paper, we use GAIN/LOSS which is heuristic based resource allocating algorithm [7]. In GAIN approach, all tasks are allocated to resources that minimize the execution cost, then re-allocated to the machine where the largest makespan benefit is obtained by the smallest cost. This is repeated until the whole budgets are exceed. On the other hand, in LOSS approach, tasks that are initially scheduled by existing scheduling algorithm are re-allocated to cheaper machine, until the cost becomes equal or less than budget.

Once the workflow is scheduled, the parallel task merging approach shown in Fig. 2 is executed. If two tasks are overlapped and satisfies all conditions to be merged, then two tasks are allocated to the same VM and executed in parallel. Although the execution time of each task could be increased, it does not affect the total execution time of workflow much, because the tasks on the critical path, which represents the longest path in workflow and decides the total execution time, is not considered to be merged.

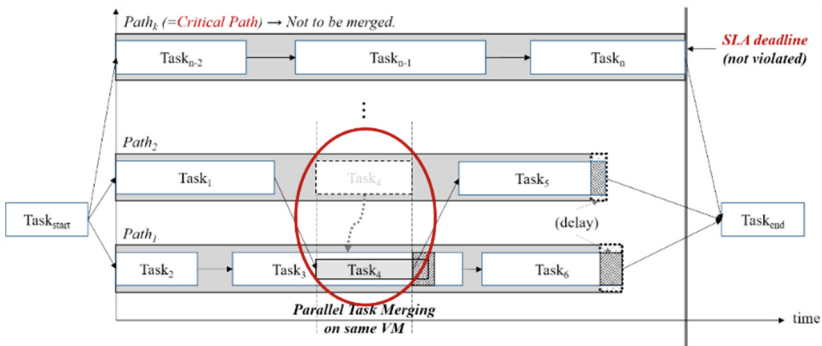


Fig. 2. Parallel task merging approach

The pseudo code of the parallel task merging is shown in Fig. 3. Input is the initially scheduled workflow  $G_1(Q, E)$  and SLA deadline from the user.  $Q_{merged}$  is the set which contains the task pairs to be merged, and initially set to null. For all task pairs  $(q_i, q_j)$  in the workflow which are not in  $Q_{merged}$  and  $CriticalPath$ , merge  $q_i$  into  $q_j$  in  $VM_{q_j}$  if they are overlapped.  $ST(q_i)$  is the start time of task  $q_i$  and  $ET(q_i)$  is the end time of task  $q_i$ . Tasks on  $CriticalPath$  are not considered to be merged, because parallel task merging increases the execution time of each task and  $CriticalPath$  decides the total execution time of whole workflow. If the total execution time of workflow is increased, then SLA deadline could be violated. The execution time of  $q_i$  and  $q_j$  after merging is calculated as follows,

$$XT(q_i)_{aftermerging.VM_{q_j}} = (XT(q_i)_{alone.VM_{q_j}} - OT(q_i, q_j)_{alone.VM_{q_j}}) + \frac{OT(q_i, q_j)_{alone.VM_{q_j}}}{XT(q_i)_{alone.VM_{q_j}}} \cdot XT(q_i)_{merged.VM_{q_j}} \quad (1)$$

$$XT(q_j)_{aftermerging.VM_{q_j}} = (XT(q_j)_{alone.VM_{q_j}} - OT(q_i, q_j)_{alone.VM_{q_j}}) + \frac{OT(q_i, q_j)_{alone.VM_{q_j}}}{XT(q_j)_{alone.VM_{q_j}}} \cdot XT(q_j)_{merged.VM_{q_j}} \quad (2)$$

where  $XT(q_i)_{aftermerging.VM_{q_j}}$  is the execution time of  $q_i$  when  $(q_i, q_j)$  are merged on  $VM_{q_j}$ . Additionally,  $XT(q_i)_{alone.VM_{q_j}}$  is the execution time of  $q_i$  on  $VM_{q_j}$  without merging and  $OT(q_i, q_j)_{alone.VM_{q_j}}$  is the overlapped execution time of  $(q_i, q_j)$  on  $VM_{q_j}$  without merging. Using the execution time after merging, check whether  $XT(q_i)_{aftermerging.VM_{q_j}} + XT(q_j)_{aftermerging.VM_{q_j}} - OT(q_i, q_j)_{aftermerging.VM_{q_j}}$  is shorter than  $XT(q_i) + XT(q_j)$ , which means execution time before merging of  $q_i$  and  $q_j$ , respectively. With the changed information of VMs, the total cost of the workflow  $TotalCost(G_2(Q, E))$  is also calculated. If the first condition is satisfied and the total cost of the workflow after merging is less than the original total cost of the workflow, then put  $(q_i, q_j)$  task pair into  $Q_{merged}$  set. After all these recursive procedures, the re-scheduled workflow  $G_2(Q, E)$  is output.

When the workflow scheduling is executed, the Cloud broker system should manage resource provisioning. In this paper, we propose Auto-scaling scheme which adjusts VM resource provision according to varying workflow application requests. The pseudo code of the auto-scaling scheme is shown in Fig. 4. This scheme predicts VM requests in each VM type using the autoregressive integrated moving average (ARIMA) model [9–11] and is conducted every period  $T$ . The number of predicted VM requests in each VM type is calculated as follows,

**Algorithm 1. Parallel Task Merging****Input:** Initially scheduled workflow  $G_1(Q, E)$ , SLA deadline**Output:** Workflow  $G_2(Q, E)$  with parallel task merging

---

```

 $Q_{merged} \leftarrow \text{null}$ 
for all ( $q_i \in Q$  and  $q_i \notin Q_{merged}$  and  $q_i \in \text{CriticalPath}$ ) do
  for all ( $q_j \neq q_i$  and  $q_j \in Q$  and  $q_j \notin Q_{merged}$  and  $q_j \in \text{CriticalPath}$ ) do
    if ( $ST(q_j) \leq ST(q_i)$  and  $ET(q_j) > ST(q_i)$ ) then
       $G_2(Q, E) \leftarrow$  apply task merging of ( $q_i, q_j$ ) to  $VM_{q_j}$  for  $G_1(Q, E)$ 
      calculate  $XT(q_i)_{\text{aftermerging}, VM_{q_j}}$  and  $XT(q_j)_{\text{aftermerging}, VM_{q_j}}$  (using
      Eqs. (1) and (2))
      calculate  $OT(q_i, q_j)_{\text{aftermerging}, VM_{q_j}}$ 
      if ( $XT(q_i)_{\text{aftermerging}, VM_{q_j}} + XT(q_j)_{\text{aftermerging}, VM_{q_j}} - OT(q_i, q_j)_{\text{aftermerging}, VM_{q_j}} \leq$ 
 $XT(q_i) + XT(q_j)$  and  $TotalCost(G_2(Q, E)) < TotalCost(G_1(Q, E))$ ) then
         $Q_{merged} \leftarrow$  add ( $q_i, q_j$ )
      end if
    end if
  end for
end for
 $G_2(Q, E) \leftarrow G_1(Q, E)$ 
for all ( $(q_i, q_j) \in Q_{merged}$ ) do
   $G_2(Q, E) \leftarrow$  apply task merging of ( $q_i, q_j$ ) to  $VM_{q_j}$  for  $G_1(Q, E)$ 
end for

```

---

**Fig. 3.** The algorithm of parallel task merging scheme

$$N_{instance,k}^{predicted}(t) = \left[ \frac{1}{T} \sum_{i=t}^{t+T} R_{p,k}(i) - N_{instance,k}^{processing}(i) \right] \quad (3)$$

where  $N_{instance,k}^{predicted}(t)$  is the number of predicted VM requests in VM type k,  $R_{p,k}(t)$  is the predicted requests in VM type k during  $[t, t + 1]$  and  $N_{instance,k}^{processing}$  is the number of VM instances in VM type k which is processing some jobs. If obtained  $N_{instance,k}^{predicted}(t)$  is greater than zero, then add  $N_{instance,k}^{predicted}(t) + N_{instance,k}^{backup}$  VM instances in VM type k. Otherwise, remove  $N_{instance,k}^{predicted}(t) - N_{instance,k}^{backup}$  VM instances in VM type k.  $N_{instance,k}^{backup}$  is the number of backup idle VM in VM type k. With this scheme, there are always  $N_{instance,k}^{backup}$  backup VMs in each VM type which process nothing to save some time in generating new VM instance. The goal of this approach is to keep low number of VMs leading to better cost efficiency, while keeping backup idle VMs in each VM type to save VM instance startup time.

**Algorithm 2. Auto-scaling VM resources****Input:**

historical VM requests in each VM type during  $[t - T, t - 1]$ , where  $T$  is a VM requests prediction period

set VMTYPE: each component of set is VM flavor type

**for all** ( $k \in \text{VMTYPE}$ ) **do**

predict VM requests during  $[t, t + T]$

calculate  $N_{instance,k}^{predicted}(t)$  (using Eq. (3))

**if** ( $N_{instance,k}^{predicted}(t) > 0$ ) **then**

add  $N_{instance,k}^{predicted}(t) + N_{instance,k}^{backup}$  VM instances in VM type k

**else**

remove  $N_{instance,k}^{predicted}(t) - N_{instance,k}^{backup}$  VM instances in VM type k

**end if****end for**

**Fig. 4.** The algorithm of auto-scaling VM resources scheme

In Sect. 3, we test the workflow management process to evaluate our two proposed schemes. The parallel task merging approach is adopted in workflow scheduling phase, and the auto-scaling scheme is adopted in resource provisioning phase. Through the experiment, we discuss how much cost could be saved through increasing the utilization of resources with the proposed approaches, while satisfying SLA deadlines.

### 3 Experimental Environment and Performance Evaluation

In order to evaluate the performance of the parallel task merging scheme with auto-scaling resource provisioning, we establish and execute Cloud resource broker system on the Cloud environment using OpenStack [12], which is the open source Cloud platform. In addition, we utilize one of the OpenStack component called Nova, which support VM resource managements such as VM instance allocation, VM image enrollment, VM flavor type management, and so on. The details of the experimental configuration is shown in Fig. 5. In our experiment, we use 5 HP Xeon (2.4 GHz) machines consisting of 4 Nova compute nodes for actual computing works, and one Nova controller node which manages the entire operations between compute nodes. Each machine has 8 CPU cores, 16 GB Memory, 1 TB storage and two wired Network Interface Cards (NIC) which generate private and public network. OpenStack Cloud platform in our experiment is based on Ubuntu 15.04.

In our experiment, we adopt open-source based scientific workflow application called Montage [13]. Montage is a toolkit, which is designed to assemble Flexible Image Transport System (FITS) images into custom mosaics. In our experiment, we use M105, M106 and M108 FITS images to be processed by Montage application. During the Montage workflow process, each FITS image is processed by several tasks in workflow, which are called mImgtbl, mMakeHdr, mProjExec, mAdd and MJPEG.





In addition, in order to mimic the real VM type model of Cloud provider, we adopt GoGrid [14] provider model. The configuration of each VM flavor type and cost is shown in Table 1. In this paper, we suppose that the cost is charged based on the exact resource using time, unlike commercial pricing model that charges based on fixed time unit, such as hour, month or year.

In our experiment, we utilize three workflow types shown in Fig. 6. Each workflow type with its own deadline and budget forms each service user type. Accordingly, there are three user types. These user types have deadlines of 300, 700 and 800 s and budgets of 33.33, 60 and 66.67 dollars, respectively. The inter arrival time of requests from users conforms to exponential distribution with mean 500 s and request type is randomly selected each time. Since our parallel task merging scheme requires initial workflow scheduling with resource allocation, we adopt GAIN and LOSS algorithms [7] as an initial scheduling. Prior to adopt our approach on workflow scheduling process, we acquire the execution time of merged tasks on same VM instance from many experimental results.

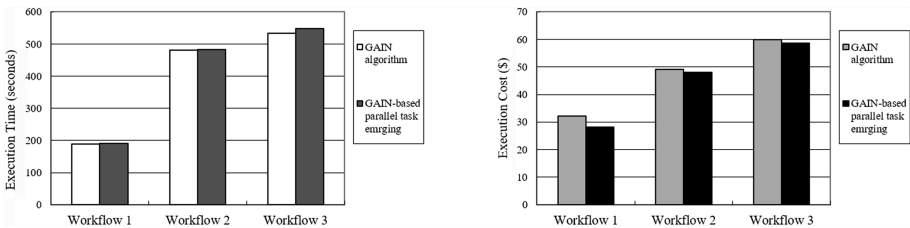
Figure 7 shows the execution time and cost of each workflow scheduled by GAIN algorithm without and with our parallel task merging scheme. When parallel task merging scheme is adapted on GAIN scheduling algorithm, the total execution cost of

**Table 1.** VM pricing model corresponding to each flavor type [14]

VM type	CPU core(s)	RAM	Storage	Cost per hour
Small	1	1 GB	50 GB	\$0.08
Medium	2	2 GB	100 GB	\$0.16
Large	4	4 GB	200 GB	\$0.32
X-large	8	8 GB	400 GB	\$0.64

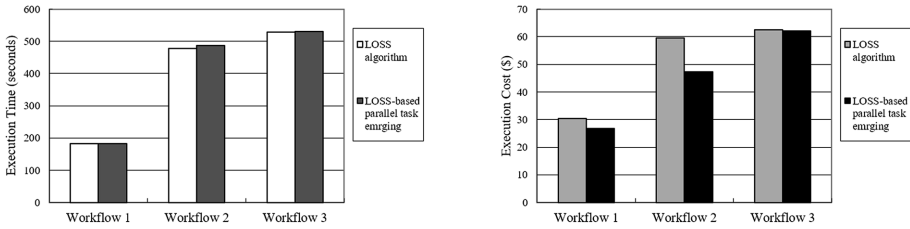
each workflow is decreased by 12.7 %, 2.5 %, 2.2 % respectively, compared to original GAIN algorithm. On the other hand, the total execution time of each workflow is increased by 1.3 %, 0.6 %, 2.8 % respectively, however, it does not violate each SLA deadline.

Similarly, Fig. 8 shows the execution results of each workflow scheduled by LOSS algorithm without and with our proposed scheme. When parallel task merging scheme



**Fig. 7.** Execution time and cost of GAIN and the proposed algorithms

is adapted on LOSS scheduling algorithm, the total execution cost of each workflow is decreased by 11.5 %, 20.4 %, 0.9 % respectively, compared to LOSS algorithm



**Fig. 8.** Execution time and cost of LOSS and the proposed algorithms

without our approach. The total execution time of each workflow is increased by 0.1 %, 1.9 %, 0.5 % respectively, however, it still satisfies each SLA deadline.

Since the goal of our parallel task merging approach is to increase the resource utilization, the execution cost is decreased while the total execution time is increased. However, the increased execution time is not huge to violate SLA deadline, because when we choose the tasks to be executed in parallel, tasks in the critical path are not selected. Therefore, the increased time caused by parallel execution of tasks has a few influence to affect the whole execution time of workflow. Moreover, if total increased time of tasks does not make the each corresponding path to reach the time length of critical path, then the execution time of workflow would be totally unrelated to our parallel task merging scheme. Additionally, the auto-scaling approach also contributes to saving the cost of resources, since this approach allocates many requests on each VM as long as the number of requests does not exceed upper limit.

## 4 Conclusion

In this paper, we proposed two algorithms in Cloud broker system that consists of workflow scheduling phase and resource provisioning phase. In workflow scheduling part, most traditional approaches have allocated each task on certain resource in serial, not in parallel. However, generally single task does not maximize the utilization of given resource, which led us to propose the parallel task merging approach that allocate overlapped multiple tasks on the same resource, simultaneously. In resource provisioning part, we proposed auto-scaling approach, which adjust the number of VMs according to changing the number of requests.

Through the experimental performance evaluation, we showed that our proposed approaches decreased the total execution cost of each workflow compared to conventional GAIN and LOSS algorithms [7], while the SLA deadline was still satisfied. Generally, single task does not maximally utilize allocated resource such as CPU, memory, and so on. With our parallel task merging scheme, the multiple tasks were processed on the same resource simultaneously, which led to the improved resource utilization and the decreased cost. Additionally, in the resource provisioning phase, the

proposed auto-scaling algorithm made the efficient resource pool management through adjusting the number of VMs and maximizing the utilization of VMs.

**Acknowledgments.** This work was supported by ‘The Cross-Ministry Giga KOREA Project’ grant from the Ministry of Science, ICT and Future Planning, Korea.

## References

1. Ren, Y.: A cloud collaboration system with active application control scheme and its experimental performance analysis. Master’s thesis, KAIST (2012)
2. Gary, M.R., Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Co., New York (1979)
3. Topcuoglu, H., Hariri, S., Wu, M.Y.: Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Trans. Parallel Distrib. Syst.* (2002)
4. Abrishami, S., Naghibzadeh, M., Epema, D.H.J.: Deadline-constrained workflow scheduling algorithms for infrastructure as a service clouds. *Future Gener. Comput. Syst.* (2013). (Elsevier)
5. Yu, J., Buyya, R., Tham, C.K.: Cost-based scheduling of scientific workflow applications on utility grids. In: *International Conference on e-Science and Grid Computing* (2005)
6. Howard, R.A.: *Dynamic Programming and Markov Processes*. The Massachusetts Institute of Technology Press, Cambridge (1960)
7. Sakellariou, R., Zhao, H.: Scheduling workflows with budget constraints. In: Gorlatch, S., Danelutto, M. (eds.) *Integrated Research in GRID Computing*. CoreGRID Series. Springer, Heidelberg (2007)
8. Kang, D.K., Kim, S.H., Youn, C.H., Chen, M.: Cost adaptive workflow scheduling in cloud computing. In: *ICUIMC*. ACM (2014)
9. Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*. Springer, New York (2002)
10. Fang, W., Lu, Z., Wu, J., Cao, Z.: RPPS: a novel resource prediction and provisioning scheme in cloud data center. In: *IEEE Ninth International Conference on Services Computing* (2012)
11. Kim, H., Ha, Y., Kim, Y., Joo, K.-N., Youn, C.-H.: A VM reservation-based cloud service broker and its performance evaluation. In: Leung, V.C.M., Lai, R., Chen, M., Wan, J. (eds.) *CloudComp 2014*. LNICST, vol. 142, pp. 43–52. Springer, Heidelberg (2015)
12. OpenStack. <https://www.openstack.org/>
13. Montage, An Astronomical Image Mosaic Engine. <http://montage.ipac.caltech.edu/>
14. GoGrid. <https://www.datapipe.com/gogrid/>