Renumber Coevolutionary Multiswarm Particle Swarm Optimization for Multi-objective Workflow Scheduling on Cloud Computing Environment

Hai-Hao Li, Zong-Gan Chen, Zhi-Hui Zhan* (Corresponding Author), Ke-Jing Du, Jun Zhang

Department of Computer Science, Sun Yat-sen University, Guangzhou, P. R. China, 510275
Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, P.R. China
Collaborative Innovation Center of High Performance Computing, National University of Defense Technology, Changsha 410073, China
Engineering Research Center of Supercomputing Engineering Software, Ministry of Education, P.R. China
*zhanzhh@mail.sysu.edu.cn

ABSTRACT

Resources scheduling is a significant research topic in cloud computing, which is often modeled as a cost-minimization and deadline-constrained workflow scheduling model. This is a constrained single objective problem that to minimize the overall workflow execution cost while meeting deadline constraints. In this paper, we offer a new horizon to convert this single-objective problem to a multi-objective problem and present coevolutionary multiswarm particle swarm optimization (CMPSO) to find the non-dominated solutions with different execute cost and time. Meanwhile, the renumber strategy is adopted in CMPSO to make the learning efficient. CMPSO is compared with a renumber PSO (RNPSO) by setting the execute time in the CMPSO's non-dominated solutions as the deadline constraint of RNPSO. Results show that CMPSO not only offers many non-dominated solutions with different prices and execute time, but also obtains better solution than RNPSO under a same deadline.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search – *Heuristic methods*.

General Terms

Algorithms, Performance, Design, Experimentation

Keywords

Cloud computing; scheduling; renumber; particle swarm optimization; guiding point

1. INTRODUCTION

Cloud computing is a rapid developing computing type that groups a mass of resources and uses virtualization technique to provide service over the Internet [1][2]. Its parallel and distributed computing ability also make it efficiency for executing workflows which require a high-level computing environment because of the complex and large amount of data.

Fig. 1 illustrates a workflow with a set of tasks $T = \{t_1, t_2, \dots, t_n\}$, which have parent-children relationship. The values on the edges denote the data transfer time from the parental task to the children task if they are not executed on the same virtual machines (VMs). Therefore, a child task can only be executed after all its parent tasks have been finished and all the data have been transferred. For example, for task t_5, if it is executed on the same VM with task t_2, but not the same with task t_3. Then the task t_5 can be executed only after both t_2 and t_3 have finished and t_3 has transferred the data to t_5, with data transfer time being 2.

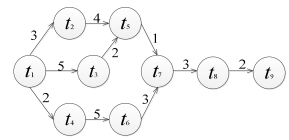


Fig. 1. A sample of workflow.

The cloud workflow scheduling model has been defined in [3], [6], and [7], which is also briefly described as follows. A schedule is defined as $S = (T, R, M, TEC, TET)$ where T represents a set of tasks, $R = \{r_1, r_2, \dots, r_n\}$ is a set of resources, M represents the tasks to resources mappings, TEC is short for 'total execution cost', and TET is short for 'total execution time'. Every resource r_j has a VM type VM_j and an estimated lease start time LST_j and lease end time LET_j. M represents a mapping and is consisted of the forms $m(i, j)$ for every task. The elements in $m(i, j)$ express that task t_i is scheduled to run on resource r_j and is anticipated to begin at start time ST_i and finish at end time ET_i. The calculations of TEC and TET are shown as Eqs (1) and (2).

$$TEC = \sum_{i=1}^{|R|} C_i \times \lceil LET_i - LST_i \rceil \tag{1}$$

$$TET = \max\{ET_t : t \in T\} \tag{2}$$

(A) For a single-objective problem, the optimization objective is to minimize the value of TEC while to make sure the value of TET meets the deadline constraint. That is, the objective is as Eq (3) and the constraint is as Eq (4).

$$Minimize \quad f = TEC \tag{3}$$

$$TET \le deadline \tag{4}$$

(B) For a multi-objective problem, there are two objectives TEC and TET to be minimized. Thus the solution is a schedule S with non-dominated TEC and TET as Eqs (5).

$$Non\text{-}dominated \quad \begin{cases} f_1 = TEC \\ f_2 = TET \end{cases} \tag{5}$$

Rodriguez and Buyya [3] adopted the single-objective optimization model in their work and proposed particle swarm optimization (PSO) [4][5] to find a resource scheduling sequence

GECCO'15 Companion, July 11-15, 2015, Madrid, Spain
ACM 978-1-4503-3488-4/15/07.
http://dx.doi.org/10.1145/2739482.2764632

on that model.. Later, Chen et al. [6] and Li et al. [7] also followed such a model and proposed to use dynamic objective genetic algorithm and renumber PSO (RNPSO) to solve the problem respectively. In this paper, we extend the scheduling model from the single-objective problem to a multi-objective problem and present coevolutionary multiswarm particle swarm optimization (CMPSO) [8] to find the non-dominated solutions with different execute cost and time.

As RNPSO uses a resource renumber strategy to sort the cloud resources according to the price per unit time, it insures the learning process in PSO make sense. The RNPSO has been evaluated by Li et al. [7] and has shown better performance than traditional PSO. Therefore, in this paper, we adopted the resource renumber strategy into CMPSO.

2. Performance Evaluation

The parameters of RNPSO and CMPSO are presented in Table 1.

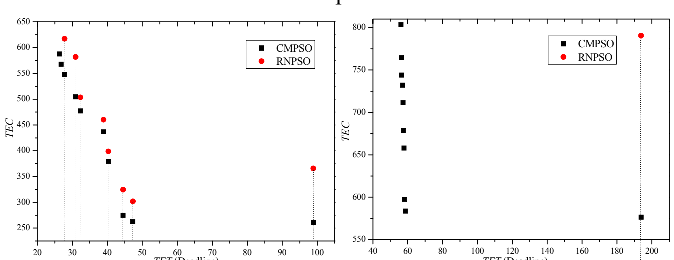
Table 1 Parameter in RNPSO and CMPSO

Parameter	RNPSO	CMPSO	Others
popsize	100	50×2=100	Archive size of CMPSO 10
ω	0.5	0.9~0.4	
c_1,c_2	2.0	c_1,c_2,c_3 4.0/3	
r_1,r_2	[0,1]	r_1,r_2,r_3 [0,1]	

We examine RNPSO and CMPSO on two different scales of cloud computing environments, which are the same with the ones used in [7]. Case 1: 50 tasks and 10 resources; Case 2: 100 tasks and 10 resources.

In each scale environment, we run CMPSO 10 independent times, with each time 4000 generations. The non-dominated solutions of the 10 times are presented in Table 2 and Table 3 respectively.

In the first two rows of the tables, 10 non-dominated solutions obtained by CMPSO with different *TET* and *TEC* are presented. For each solution, we set the *TET* obtained by CMPSO as the deadline constraint for RNPSO and present the *TEC* result obtained by RNPSO in the third row to compare with CMPSO. Moreover, the probability of finding feasible solution under such *TET* deadline constraint is also presented in the fourth row.



(a) Case 1 (b) Case 2

Fig. 2. Result of RNPSO and CMPSO. (a) Case 1; (b) Case 2.

We also plot the solutions found by RNPSO and CMPSO in Fig. 2. The results in the tables and figure show that CMPSO is better than RNPSO. CMPSO can obtain different non-dominated solutions with different *TET* and *TEC* values, while RNPSO can only obtain solutions whose *TET* (deadline) value is large. The reason may be that CMPSO doesn't need to consider the deadline as a constraint so it can evolve smoothly and provide a set of non-dominated solutions.

3. Conclusions

In this paper, we firstly transfer the single-objective problem in [7] to a multi-objective problem by considering the execute time as a second objective instead of a deadline constraint. Meanwhile, we proposed CMPSO to solve the multi-objective optimization problem. The results show that GPCMPSO can not only offer many non-dominant choices with serial different prices and execute time, but also it can work out a more beneficial solution than RNPSO under a same deadline. Additionally, CMPSO is able to provide outcomes with an excessively pressed deadline while RNPSO struggling on finding a feasible answer.

4. Acknowledgments

This work was supported in part by the National Natural Science Foundations of China (NSFC) with No. 61402545, the Natural Science Foundations of Guangdong Province for Distinguished Young Scholars with No. 2014A030306038, the Project for Pearl River New Star in Science and Technology, Guangzhou, China, the Fundamental Research Funds for the Central Universities, the NSFC Key Program with No. 61332002, the NSFC for Distinguished Young Scholars with No. 61125205, and the National High-Technology Research and Development Program (863 Program) of China No.2013AA01A212.

5. References

[1] X. F. Liu, Z. H. Zhan, K. J. Du, and W. N. Chen, "Energy aware virtual machine placement scheduling in cloud computing based on ant colony optimization approach," in *Proc. Genetic Evol. Comput. Conf.*, 2014, pp. 41-47.

[2] Z. H. Zhan, X. F. Liu, Y. J. Gong, J. Zhang, S. H. Chung, and Y. Li, "Cloud computing resource scheduling and a survey of its evolutionary approaches," *ACM Computing Surveys*, Accepted, 2015.

[3] M. A. Rodriguez and R. Buyya, "Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds," *IEEE Transaction on Cloud Computing*, vol. 2, no. 2, April-June 2014.

[4] M. Shen, Z. H. Zhan, W. N. Chen, Y. J. Gong, J. Zhang, and Y. Li, "Bi-velocity discrete particle swarm optimization and its application to multicast routing problem in communication networks," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 7141-7151, Dec. 2014..

[5] Z. H. Zhan, J. Zhang, Y. Li, and Y. H. Shi, "Orthogonal learning particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 15, no. 6, pp. 832-847, Dec. 2011.

[6] Z. G. Chen, K. J. Du, Z. H. Zhan, and J. Zhang, "Deadline constrained cloud computing resources scheduling for cost optimization based on dynamic objective genetic algorithm," in *Proc. IEEE Congr. Evol. Comput.*, 2015, Accepted.

[7] H. H. Li, Y. W. Fu, Z. H. Zhan, and J. J. Li, "Renumber strategy enhanced particle swarm optimization for cloud computing resource scheduling," in *Proc. IEEE Congr. Evol. Comput.*, 2015, Accepted.

[8] Z. H. Zhan, J. J. Li, J. Cao, J. Zhang, H. Chung, Y. Li, and H. Shi, "Multiple populations for multiple objectives: A coevolutionary technique for solving multiobjective optimization problems," *IEEE Transaction on Cybernetics*, vol.43, no.2, pp. 445-463, April, 2013.

Table 2 Results in Case 1 Environment

TET	98.896	47.322	44.526	40.334	38.930	32.359	30.957	27.808	26.829	26.337
TEC	260.206	262.607	274.662	378.940	436.678	477.202	504.196	547.276	567.730	587.628
RNPSO	365.614	301.748	324.589	398.487	460.274	503.342	581.928	617.268	-	-
prob	100%	100%	100%	100%	100%	90%	70%	70%	0%	0%

Table 3 Results in Case 2 Environment

TET	193.867	58.757	58.114	57.886	57.587	57.333	57.034	56.634	56.335	56.157
TEC	576.481	583.615	597.251	657.841	678.295	711.643	732.097	744.023	764.477	803.488
RNPSO	790.589	-	-	-	-	-	-	-	-	-
prob	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%