Journal of Optimization in Industrial Engineering 14 (2014) 27-36

Multi-objective and Scalable Heuristic Algorithm for Workflow Task Scheduling in Utility Grids

Vahid Khajehvand[a,*], Hossein Pedram[b], Mostafa Zandieh[c]

[a] *Assistant Professor, Department of Computer Engineering and Information Technology, Qazvin Branch, Islamic Azad University, Qazvin, Iran*

[b] *Associate Professor, Department of Computer Engineering and Information Technology, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran*

[c] *Associate Professor, Department of Industrial Management, Shahid Beheshti University, G.C., Tehran, Iran*

Received 12 September, 2012; Revised 29 January, 2013; Accepted 20 February, 2013

Abstract

To use services transparently in a distributed environment, the Utility Grids develop a cyber-infrastructure. The parameters of the Quality of Service such as the allocation-cost and makespan have to be dealt with in order to schedule workflow application tasks in the Utility Grids. Optimization of both target parameters above is a challenge in a distributed environment and may conflict one another. We, therefore, present a novel heuristic algorithm for scheduling a workflow application on Utility Grids. Our proposed algorithm optimizes the allocation-cost and makespan in a scalable and very low runtime. The results of the wide-spread simulation indicate that the proposed algorithm is scalable against an increase in the application size and task parallelism of the application. The proposed algorithm effectively outperforms the current algorithms in terms of the allocation-cost, makespan and runtime scalability.

Keywords: Utility Grids, Resource Provisioning, Workflow Scheduling, Multi-objective Optimization, Scalability.

1. Introduction

Grid computing is capable of controlling a wide variety of heterogeneous distributed resources to execute computation and data intensive applications. Grid computing has recently been oriented towards pay-as-you-go models. In these models, the resource providers receive fees from the users for presenting computing and data services. That is why; the industry pioneers such as IBM, HP, Intel and SUN with a large share in this business are more inclined toward the grid computing. IBM, for instance exploits "e-business on demand" model, HP exploits "Adaptive enterprise" model and Sun Microsystems apply to "pay-as-you-go" model Garg et al. (2010). To conduct large-scale computations, the shared distributed infrastructures create the grid environment software and hardware resources. These infrastructures proved to be efficient for executing applications in sciences such as astronomy Katz et al. (2005), high energy physics Deelman et al. (2002) and others.

To describe an application in a high level form regardless of the distributed computing environment, the workflow is the most common approach. A workflow is represented in a "Direct Acyclic Graph" (DAG) with nodes and edges representing the tasks and data dependencies between the tasks, respectively. A DAG is defined as $G = (V, E)$, where V is a set of nodes, each node representing a task, and E is a set of links, each link representing the execution precedence order between two tasks. For example, a link $(i, j) \in E$ represents the precedence constraint that task v_i needs to be completed before task v_j starts. As a workflow may consist of sub-workflows with multiple entries and exits so the first thing to be done is to add two pseudo-tasks, a top task and a bottom task, with zero execution time indicated by 0 and $n + 1$, respectively. The top task spawns all actual entry tasks of the workflow to be linked to a single node, while the bottom task joins all actual exit tasks to a single node.

Once an application is transformed into the workflow structure, a workflow management system such as Pegasus Deelman et al. (2005) will be ready to control and manage the execution of workflow on the distributed infrastructure. In these environments, indeed, access to the shared computational resources is carried out through the queue-based Local Resource Management (LRM) system.

A competition develops among users caused by the resources-pricing policies so that users begin being involved in a competition with one another only to gain a resource with an affordable cost and an efficient processing capability. Similarly, resource providers are driven into a competition with one another to sell their idle resource slots to the users in order to gain more profits as well as enhance the resource utilization. The

* Corresponding author E-mail: Khajehvand@qiau.ac.ir

proposed genetic algorithm finds an optimized mapping of the tasks to the resources minimizing both financial cost and makespan. This approach developed in [19, 25] presents the cost-based model in which the resource providers advertise the available resource slots to the users. To minimize the application makespan under the minimum resource allocation-cost, the presented multi-objective genetic algorithm is capable of provisioning a subset of the resource slots. The main difference between these cost minimization algorithms and our proposed algorithm is that these minimization algorithms rely on a cluster with homogeneous processors. Thus, in (Singh et al, 2009; Singh et al, 2007), the entire resources possess identical CPU ratings and cost processing whereas in the proposed model, all resources are constituted of the heterogeneous clusters with different processing cost and CPU ratings in the real-world Utility Grids environments. Hence, removing this resource homogeneity complicates the identification of an appropriate resource selection.

Since the above-mentioned cost optimization algorithms (Singh et al, 2009; Singh et al, 2007) are genetic-based ones, the runtime takes a longer time. In case, the slots' characteristics undergo a change during scheduling, the slots' characteristics are to be updated and a rescheduled resulting in a far longer runtime. Hence, these approaches do not serve the purpose in the dynamic environments such as the Grids.

This paper deals with developing a Workflow Planning Cost-based (WPC) model in order to effectively schedule an application in the Utility Grids so that the application time and allocation-cost can be minimized. In fact, the WPC model allows the users to make a trade-off between an application time and allocation-cost. Next, a First-fit Cost-Time Trade-off (FCTT) heuristic algorithm is employed to solve the workflow scheduling problem. The FCTT is a heuristic algorithm that schedules an application in a form that both the time and the allocation-cost can be optimized according to the trade-off factor. A trade-off factor shows the preference of the allocation-cost optimization to the turnaround-time. In Khajevand et al. (2012), we presented a preliminary version of the proposed algorithm so that it selects a task with a minimum first fit cost-makespan objective function. However, this paper was not considered the issue of scalability with different workflow sizes, task parallelism and heterogeneous resources. Finally, to study and evaluate the efficiency of the proposed algorithm on the proposed model, a handful of experiments have been conducted and simulated. The main contributions of the paper are as follows:

1) Developing a WPC model based on provisioning the resources for scheduling a workflow, so that the application makespan and allocation-cost can be minimized.
2) Developing a multi-objective FCTT heuristic algorithm based on the WPC model with the following characteristics: (a) the scalability and a better performance due to an increase in the workflow size. (b) the scalability and a better performance according to an increase in the degree of the task parallelism.

The rest of this paper is organized as follows: Section 2 introduces a workflow planning problem and execution environment. A proposed detailed model and heuristic algorithm is described in Section 3. Section 4 involves a simulation setup and its relevant experiments in order to evaluate the efficiency of the proposed algorithm. In Section 5, the results have been analysed statistically. Finally, section 6 ends with a conclusion.

2. Workflow Planning Problem

To execute large-scale applications by developments in computer science, the collaborative use of distributed resources managed by different autonomous domains is made possible. In this environment, the resources available to these applications are shared between multiple users, so the optimization of the throughput or utilization of the resources is of importance. Consequently, the user has no explicit control on the allocation of resources; hence the performance of the application will be unpredictable in advance. Similarly, the resources do not take the users' preferences into consideration. Thus more advanced methods are needed that would allow the resources to differentiate between the users and deliver multiple qualities of service.

Performance optimization for applications in such a distributed environment is a difficult problem. Sometimes, offline methods such as manual negotiation between resources and users are used for allowing users to achieve the desired performance. However, these offline methods are not scalable against an increase in the application size and task parallelism of the application.

The user submits the application characteristics to the application-level scheduler only to be executed on the grid environment. The user expects to have his application executed with the minimal time and allocation-cost. Certainly, the users exploit trade-off factor in order to show a preference for cost to time. In cases where this factor is not specified by the users, the default trade-off factor is considered as equal.

In fact, the application-level scheduler acts as a mediator between the resource providers and users. Due to the reports of the available slots obtained from the resource providers, the application-level scheduler plans the application. The entire slots exploited in planning the application, will be submitted to LRM in order to provision the resources. Each computational resource is equipped with a number of the processors, the memory and the network interfaces showing an independent processing unit. The entire resources are fully-connected while being capable of executing all application-tasks. All of the computational resources can act as a service-provider (site) for time-slots.

topological sort. Eventually, the FCTT algorithm selects a task whose execution of all parents' tasks is completed from the obtained list scheduling in order to schedule the task.

There are many choices for each task, only the choice capable of minimizing the multi-objective cost metric of Eq. (1) will be selected as the best solution. According to the best solution, the Earliest Start Time (EST) needs to be computed to execute immediate successor tasks and this procedure will be carried on so long as the execution of the whole application tasks will be finished.

The FCTT pseudo-code is presented in algorithm 1 operating according to the WPC model. The FCTT algorithm obtains the list-scheduling, application characteristic and the available list of the slots of all resources as an input parameter (lines 1, 2). Moreover, the EST is initialized with simulation current time (line 3). The application-level scheduler carries out the planning of each application task due to available slots list characteristics with an eye on the multi-objective cost metric presented in Eq. (1), (lines 4 to 14). Initially, a list of unplanned tasks eligible to be executed is selected (line 5). Next, the eligible tasks are defined as the ones whose parents' tasks execution is completed. These very same tasks have not been executed yet. The available slots list of each resource is obtained by line 7. In line 8, the EST of the task T on all the resources is computed. Eventually, the Earliest Finish Time (EFT) of the task T is computed, (line 9).

The EST is computed on the basis of the completion-time of the latest parents tasks T. Next, the best slot capable of executing the task is selected for each task T on each resource. In cases, the selected resource does not match with the resource executing the parents' tasks, the data-transfer time needs to be added to the EST.

Once the best slot to execute task T is obtained on each resource, the resource minimizing the multi-objective cost metric in Eq. (1) will be selected as the best resource (line 10). Now, it comes to allocating the task T to a selected resource (line 11) as well as updating the slots list of the selected resource (line 12). This procedure needs to be continued as long as there still exists an eligible task (lines 4 to 14). At the end of the completion of the whole application tasks, the slots assigned to the application tasks will be released.

4. Simulation Setup

To conduct an experimental evaluation of the efficiency of algorithm 1, the GridSim [28, 29] is used to simulate the application-level scheduler in the Utility Grids environment. The Grids environment modelled in this simulation consists of ten sites. These sites belong to a subset of the European Data Grid (EDG) spread across five interconnected countries via a high-speed network [1, 30]. The characteristics of the resources are shown in table 1. The mean bandwidth value of the resources is 10 Gb/s with a mean latency time of 150 s.

The workload simulated on these sites follows the workload model generated by Lublin Lublin et al. (2003). The main purpose of the use of this model is to create a realistic simulation environment where the tasks compete with one another. Table 2 shows the workload parameters values applied to in the Lublin model.

To conduct experiments, a parameterized graph generator is used to create a synthetic workflow application Topcouglu et al. (2002). Table 3 shows the workflow application characteristics as in Singh et al. (2009).

At this stage, the scheduling algorithm using the best-effort QoS for scheduling, is simulated and tagged as the BE. As the number of the resources is m and the resources are heterogeneous in terms of CPU rating and allocating-cost, a heuristic algorithm needs to be taken into account to select a suitable resource in the best-effort QoS. In BE, the exploited heuristic method selects a resource with the minimum number of tasks in the waiting and running queues. The majority of the resource management systems make it possible for users to obtain the number of the tasks in the waiting and running queues Singh et al. (2009).

For our experiments, we use the GridSim to simulate benchmark algorithms, testbed platform environment and workload on an Intel Core 2 Duo CPU T9600, 2.80 GHz computer.

Algorithm 1: The pseudo-code for the FCTT algorithm	
Input:	A list-scheduling obtained from the HEFT algorithm An application characteristics The resource characteristics and the available slots to each resource
Output:	The workflow scheduling
1	Get the list of the available time slots for all resources
2	*UnScheduledTask* = get the list of the tasks which is sorted by the rank obtained from the HEFT algorithm due to Eq. (5).
3	Assign the simulation current time to the *EST*.
4	**While** *UnScheduledTask* is not empty **do**
5	*EligibleTasks* = select all tasks which executions of their parents have been completed.
6	**for** each *T* in the *EligibleTasks* **do**
7	Acquire the available slots of each resource.
8	Compute the *EST* of the task T on each resource.
9	Compute the Earliest Finish Time (*EFT)* of the task T according to best EST.
10	Find a Time Slot (*TS*) which is feasible for the task *T* while minimizing the multi-objective cost-based metrics defined by Eq. (1).
11	Allocate the task *T* to the *TS* on the resource *r*.
12	Update the list of available slots to the resource *r*.
13	**end for**
14	**end while**
15	Compute the makespan and allocation-cost of the application.

In last row of Table 5, the outputs of the algorithms on the test problems are summarized via an average metric. According to this average values, it is clear that for all metrics, the proposed algorithms is superior. Moreover, to have a better sense of performance of the algorithms, Fig. 1 to Fig.3 are plotted for makespan, allocation-cost and runtime, respectively. According to these figures, on the makespan and allocation-cost FCTT and MOGA have close manners and outperforms BE. However, in runtime the proposed algorithm dominates the two other algorithms considerably.

Besides non-statistical test, we need to have a statistical test for comparing the algorithms. Table 6 shows a statistical comparison of the algorithms. This table presents the *P*-values of the tests. In the statistical difference to the MOGA algorithm hypothesis tests of this table, like any other hypothesis tests, *P*-values are compared with the level of 0.05 significance. In case the null hypothesis is rejected (*P*-value of the test becomes less than our considered significant level), the boxplot will be assessed and the superior algorithm will be determined. The MOGA algorithm is proved to be better than BE algorithm [19, 25]. In order to prove that the proposed approach is more effective than the MOGA and BE algorithms, we performed a Kruskal-Wallis test between metrics of the MOGA and proposed algorithms. Table 7 shows the results of this test. Due to the table, the proposed algorithm operates in two metrics- makespan and cost- the same as the MOGA does. However, in terms of runtime, the proposed algorithm shows a significant

Table 5
The outputs of the algorithms on the metrics

Test problems	BE			MOGA			Proposed Approach (FCTT)		
	Makespan	Cost	Runtime	Makespan	Cost	runtime	Makespan	Cost	Runtime
p1	38	185488	67033	16	8273	182380	16	5266	70
p2	199	890175	72421	50	51552	125317	47	29959	43
p3	455	1817690	207761	67	111106	179351	56	87689	71
p4	967	3116929	507615	83	243622	188335	72	170474	83
p5	2573	5944557	3222543	136	324209	169125	135	340184	86
p6	3889	7186587	7245006	352	1250115	178488	319	533121	55
p7	3944	9014009	5834740	352	4163217	163963	406	1188330	72
p8	2081	5200360	2423552	99	457395	266074	85	422030	202
p9	2092	7746812	2460495	108	632793	384885	98	474088	315
p10	4027	10100000	9851479	122	1246753	4233005	98	857269	735
Average	2026.5	5120261	3189265	138.5	848904	607092	133.2	410841	173.2

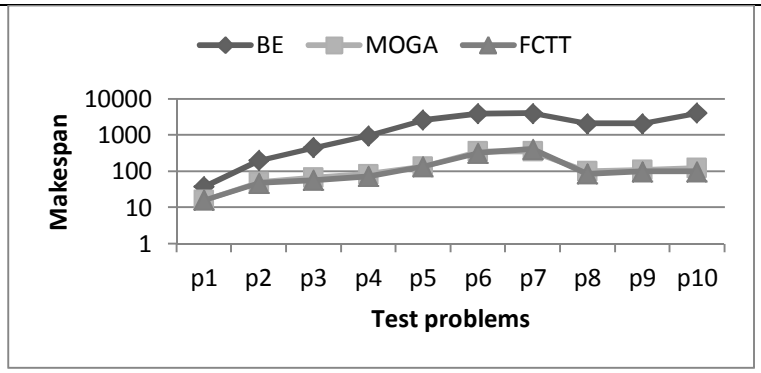

Fig. 1. comparision of the three algorithms on the makespan through test problems

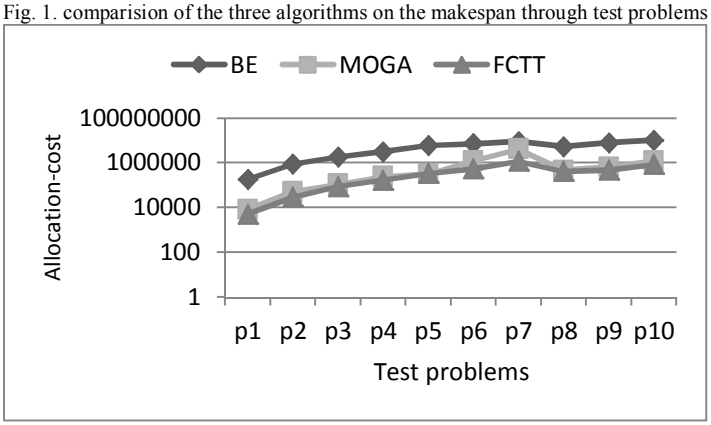

Fig. 2. comparision of the three algorithms on the allocation-cost through test problems

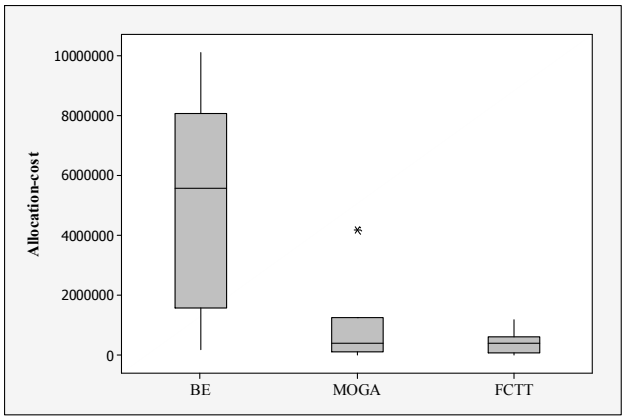

Fig. 5. Allocation-cost boxplot of the three algorithms on the test problems

6. Conclusion

The present paper involves designing, implementing and evaluating the FCTT heuristic algorithm for scheduling a workflow application. The paper seeks to optimize the multi-objective cost-time based on the proposed WPC model. To develop a real distributed environment, the resources workload is simulated based on the Lublin model. Due to many experiments conducted on a generated syntactic workflow, it was shown that the FCTT heuristic algorithm is far more effective than the

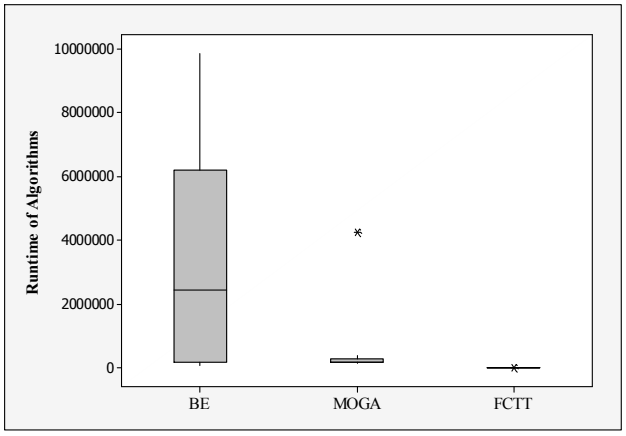

Fig. 6. Runtime boxplot of the three algorithms on the test problems

existing algorithms in terms of the cost-time optimization and scalability for scheduling the workflow application.

To determine the impact of the workflow size and task size on the allocation-cost, makespan and runtime in terms of the number of the application tasks, in this paper, a few experiments were conducted. Next, it is followed by an analysis of a comparison between the FCTT, MOGA and BE algorithms. This comparison includes both environments, the statistical and the non-statistical test. Moreover, in the statistical part, a non-parametric test is conducted.

As a result, it is shown the FCTT algorithm is scalable due to an increase in the workflow tasks as well as capable of scheduling huge workflow applications with the lowest runtime in the heterogeneous environment. However, in terms of runtime, the proposed algorithm shows a significant difference to the existing algorithms.

References

[1] Abrishami, S. Naghibzadeh, M. and Epema, D. H. (2012), "Cost-driven scheduling of grid workflows using partial critical paths," Parallel and Distributed Systems, IEEE Transactions on, vol. 23, pp. 1400-1414.

[2] Buyya R. and Murshed, M. (2002), "Gridsim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing," Concurrency and Computation: Practice and Experience, vol. 14, pp. 1175-1220.

[3] Czajkowski, K. Foster, I. and Kesselman, C. (2005), "Agreement-based resource management," Proceedings of the IEEE, vol. 93, pp. 631-643.

[4] Deelman, E. Kesselman, C. Mehta, G. Meshkat, L. Pearlman, L. Blackburn, K. et al., (2002), "GriPhyN and LIGO, building a virtual data grid for gravitational wave scientists," in 11th IEEE International Symposium on High Performance Distributed Computing (HPDC-11), Edinburgh, Scotland, UK.

[5] Deelman, E. Singh, G. Su, M. H. Blythe, J. Gil, Y. Kesselman, C., et al., (2005), "Pegasus: A framework for mapping complex scientific workflows onto distributed systems," Scientific Programming, vol. 13, pp. 219-237.

[6] Deelman, E., Singh, G., Livny, M., Berriman, B. and Good, J. (2008), "The cost of doing science on the cloud: the montage example," in Proceedings of the 2008 ACM/IEEE conference on Supercomputing, NJ, USA , pp. 1-12.

[7] Dongarra, J. J. Jeannot, E. Saule, E. and Shi, Z. (2007), "Bi-objective scheduling algorithms for optimizing makespan and reliability on heterogeneous systems," in Proceedings of