Future Generation Computer Systems 49 (2015) 1–7

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

An incentive-based heuristic job scheduling algorithm for utility grids

Heyang Xu, Bo Yang *

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

HIGHLIGHTS

- We consider incentives for both grid users and resource providers.
- We optimize successful execution rate of jobs, combined cost, and profit fairness.
- We propose a heuristic scheduling algorithm called CGPA algorithm.
- CGPA algorithm results in better scheduling in terms of the three objectives.

ARTICLE INFO

Article history:
Received 16 April 2014
Received in revised form
23 October 2014
Accepted 1 February 2015
Available online 11 February 2015

Keywords:
Utility grid
Scheduling
Incentive
Cost
Profit fairness

ABSTRACT

Job scheduling in utility grids should take into account the incentives for both grid users and resource providers. However, most of existing studies on job scheduling in utility grids only address the incentive for one party, i.e., either the users or the resource providers. Very few studies on job scheduling in utility grids consider incentives for both parties, in which the cost, one of the most attractive incentives for users, is not addressed. In this paper, we study the job scheduling in utility grid by optimizing the incentives for both parties. We propose a multi-objective optimization approach, i.e., maximizing the successful execution rate of jobs and minimizing the combined cost (incentives for grid users), and minimizing the fairness deviation of profits (incentive for resource providers). The proposed multi-objective optimization approach could offer sufficient incentives for the two parties to stay and play in the utility grid. A heuristic scheduling algorithm called Cost-Greedy Price-Adjusting (CGPA) algorithm is developed to optimize the incentives for both parties. Simulation results show that the CGPA algorithm is effective and could lead to higher successful execution rate, lower combined cost and lower fairness deviation compared with some popular algorithms in most cases.

1. Introduction

Grid computing and related technologies have been proposed for satisfying the increasing demand for the scientific computing communities for more computing capacity [1]. In a grid computing system, applications can utilize multiple computational resources that may be heterogeneous and distributed at widespread geographic locations [2]. Recently, grid computing has been moving towards a pay-as-you-go model, in which resource providers expect an economic compensation for the computational resources or services offered to grid users [3]. Grid systems with economic interactions between users and resource providers are generally termed as utility grids [4].

Job scheduling in utility grid is a complex problem [5], in which different concerns should be taken into account and then be properly addressed. Basically, these concerns can be classified into two categories, i.e., concerns of grid users and those of grid resource providers. From the perspective of grid users, there are two major concerns in job scheduling, i.e., *successful execution rate of jobs* (SERoJ) (e.g., [6,7]), and *cost* incurred (e.g., [3,5]). These two concerns are important ones since grid users generally hope to successfully complete their submitted jobs (i.e., to meet the jobs' respective deadlines), and at the same time, at the lowest possible cost. If, on the contrary, jobs frequently miss their deadlines, or the cost incurred is high, then users tend to lose interest in the grid system and may finally leave the system. Therefore, increasing SERoJ (e.g., [6,7]) is an incentive for grid users, and so is reducing cost (e.g., [3,5]), which shall be considered in job scheduling. On the other hand, from the perspective of grid resource providers, a major concern in job scheduling is *profit fairness* as resource providers generally hope to have equal opportunities to offer their

* Corresponding author.
E-mail addresses: xuheyang124@126.com (H. Xu), yangbo@uestc.edu.cn (B. Yang).

http://dx.doi.org/10.1016/j.future.2015.02.002

resources and gain fair profits according to their capacities [6,7]. Therefore, increasing profit fairness, or equivalently reducing the *fairness deviation of profits* (FDoP, details are in Section 4.1), is desirable from the perspective of resource providers [6], which is an incentive for providers that shall also be considered in job scheduling.

From above discussions, it can be seen that in a utility grid there are different incentives for the two parties, the users and the resource providers. These incentives for both parties shall be considered and properly addressed so that a more comprehensive job scheduling algorithm for utility grid can be developed, which could help to provide both parties with sufficient incentives to stay and play in the grid, leading to a sustainable system [6].

In the literature, many studies have investigated job scheduling in utility grids [1,3,5,8–16] (details are given in Section 2). Most of the studies presented take into consideration only the incentive for grid users without considering the incentive for resource providers. For example, recently Gkoutioudi and Karatza [15] use an accelerated genetic algorithm to decrease the failure rate of jobs submitted by grid users, which is from the perspective of incentive for users. As cost is another major concern of grid users, cost-based approaches [5,8,9] have been proposed for job scheduling in utility grid. For example, Garg et al. [3] present a cost metric to manage the trade-off between job response time and the cost spent to execute job; and in their later work [5] the minimization of the *combined cost* which is the sum of cost of all users is proposed. However, it can be noted that the studies based on cost-based approaches [5,8,9] are also only from the perspective of incentive for users. In the literature, there are very few studies on job scheduling in utility grids which consider the incentives for both parties. Xiao et al. [6] propose to optimize both the SERoJ (incentive for users) and the FDoP (incentive for resource providers); nevertheless, the cost, which is one of the most attractive incentives for users [5], is not considered in deriving the scheduling algorithms.

In this paper, we make an endeavor to study the job scheduling in utility grid by addressing the major incentives for both parties in the grid, i.e., maximizing the SERoJ and minimizing the combined cost (incentives for grid users), and at the same time minimizing the FDoP (incentive for resource providers). The job scheduling is thus formulated as a multi-objective optimization problem. Then, we develop a heuristic scheduling algorithm called Cost-Greedy Price-Adjusting (CGPA) algorithm, which can effectively map jobs to resources and obtain approximate optimal solution. Simulation results show that the proposed CGPA algorithm is effective and could lead to higher SERoJ, lower combined cost and lower FDoP compared with popular algorithms (i.e., FCFS [17], MMC [5], LPGA [5], MinCTT [3]) in most cases.

The rest of this paper is organized as follows. Section 2 gives a review of related work on job scheduling in grid environment. Section 3 presents the system model used in this paper. We put forward the optimization objectives for users and resource providers and propose a heuristic scheduling algorithm called CGPA algorithm in Section 4. Simulations, results and analyses are given in Section 5. We conclude this paper in Section 6.

2. Related work

Many literatures [5,8–10,18–20,15,16,6,7,21,22] have studied the job scheduling problem in grid environment. In general, existing grid scheduling algorithms can be classified into two types based on *scheduling time*, i.e. immediate and batch mode scheduling [1]. In immediate mode, a job is scheduled as soon as it arrives at the scheduler. While in batch mode, jobs are collected into a group and mapped to resources at the end of a fixed scheduling interval. The immediate mode is suitable for the situation of low arrival rate, while batch scheduling can take better advantage of job and resource information [10]. Batch mode scheduling is commonly adopted in literatures which investigate job scheduling problem in grid environment [3,5,7].

Chandak et al. [21] compare the efficiency of many existing heuristic-based task allocation methods and classify them into several groups, such as economic heuristic, population-based heuristic, and so on. Several simple algorithms are introduced in [11–13,17]: First Come First Served (FCFS), Round Robin (RR), Sufferage, Greedy, Min–min and Max–min. These algorithms mainly focus on minimizing *makespan*, which is the finishing time of the latest job [12], without considering the budget requirement of users. Buyya et al. investigate market-oriented economic models for grid resource management [14] and propose several grid brokers, such as Nimrod/G [23] and Gridbus [24]. Nimrod/G is a computational economy driven resource broker and supports commodity market economic model. The Gridbus toolkit is driven by the requirements of grid economy and can provide grid technologies for service-oriented utility computing. In recent years, cost-based and market-oriented models have been widely applied to grid resource management and many market-based scheduling approaches [5,18–20,15,16,6,7] have been proposed. These approaches are basically based on two different economic models, i.e. auction model and commodity market model. Wolski et al. [25] propose an economic-based model, G-Commerce, and compare the efficiency of the two different economic models by simulations. They conclude that a commodity market is a better choice for controlling grid resources than auctions. Thus the scheduling approach presented in this paper is based on commodity market model.

In [18], the authors study service provisioning in grid system based on G-commerce and extend it to allow for trading and pricing of substitutable goods, which is more close to real grid environment. In [19], the authors formulate a nonlinear programming model to maximize the aggregate utilities of all grid users and propose an optimization-based resource pricing algorithm. The grid user utility is defined as a function of resource units allocated to the grid user. In [20], the authors propose a service selection algorithm, Service Providers Search Engine (SPSE), to find appropriate services for users with multiple QoS requirements. In [15], the authors use an accelerated genetic algorithm to schedule jobs in grid with multiple criteria, i.e. job response time and job successful rate. In [3], the authors propose two meta-scheduling heuristics that minimize the execution cost and response time of applications. They also present a cost metric to manage the trade-off between the execution cost and response time. In [5], the authors study how to decrease the aggregate cost of all jobs under some QoS requirements. They propose a constrained linear programming model and a linear programming genetic algorithm (LPGA) to minimize the combined spending of all users. Although the approaches introduced by above researches strive to improve performances of user experiences (such as cost and execution time), they ignore the incentives for resource providers, for example, the FDoP which is considered in this paper.

Huang et al. [16] provide a macroeconomics-based approach, which concerns the total profit of the whole grid market and is well suited for service-oriented grid. Xiao et al. [6] propose a series of incentive-based algorithms to maximize the success execution rate of jobs and to minimize fairness deviation among resource providers. Also, their research is based on immediate mode scheduling which is different from our work. Izakian et al. [7] adopt a continuous double auction method to allocate grid resources. They extend the study, [6], by further taking resource utilization rate and load balancing level into consideration. However, all the above studies ignore considering job execution cost. If the cost of executing jobs is too high, grid users will lose interest in the