Computers & Industrial Engineering 97 (2016) 191–201

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Resource consumption minimization with a constraint of maximum tardiness on parallel machines

Zhaohui Liu [a], Wen-Chiung Lee [b], Jen-Ya Wang [c,*]

[a] Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China
[b] Department of Statistics, Feng Chia University, Taichung 40724, Taiwan
[c] Department of Computer Science and Information Management, Hungkuang University, Shalu 43302, Taiwan

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form 4 May 2016
Accepted 4 May 2016
Available online 7 May 2016

Keywords:
Scheduling
Parallel machines
Makespan
Resource consumption

ABSTRACT

In response to the effects of global warming and environmental concerns, energy consumption has become a crucial issue. In this study, we consider a parallel-machine scheduling problem where the objective is to minimize the sum of resource consumption and outsourcing cost given that the maximum tardiness of all jobs does not exceed a given bound. We show that the problem is polynomially solvable for the pre-emption case and strongly NP-hard for the non-preemption case. A branch-and-bound algorithm and a hybrid metaheuristic algorithm are proposed to obtain exact and approximate solutions. Some experimental results are given to evaluate the algorithms.

1. Introduction

Parallel-machine scheduling is an important research topic in the field of job scheduling. According to the definition in Pinedo (2012), machines in parallel with different speeds (*Qm*) comprise a machine environment in which *m* machines are functioning in parallel with different speeds. If all machines have the same speed, *Qm* will be reduced to a primitive one, i.e., identical machines in parallel (*Pm*). Furthermore, parallel-machine scheduling is of equal importance in both theoretical and practical aspects. From a theoretical viewpoint, it is a generalization of the single machine and a special case of the flexible flowshop. From a practical viewpoint, it is important because the occurrence of resources in parallel is common in the real world. Parallel-machine scheduling problems have been studied intensively in the past decades. Thus, we only survey scheduling problems with tardiness objectives.

Parallel-machine scheduling is also a typical methodology for improving time efficiency. Most related studies have focused on tardiness, earliness, makespan, or waiting time. For example, Schaller (2014) presented several genetic algorithms to minimize the total tardiness on identical parallel machines. Mandel and Mosheiov (2001) considered a scheduling problem of minimizing the maximum earliness on parallel identical machines. They proposed a heuristic and two upper bounds for obtaining the optimal schedules. Ozturk, Begen, and Zaric (2014) employed a branch-and-bound method to minimize the makespan of jobs processed by multiple identical machines. In those traditional studies, time efficiency was the top priority.

For parallel-machine scheduling problems, tardiness is an important research topic, since it is closely related to customer satisfaction. For example, Ruiz-Torres, Lopez, and Ho (2007) stated that tracking the percentage of tardy jobs is important due to the relation to customer satisfaction. Lin, Pfund, and Fowler (2011) and Lin, Fowler, and Pfund (2013) also pointed out that the total tardiness is a measure of customer satisfaction. Therefore, tardiness-related research is commonly seen in practice.

Such parallel-machine scheduling for minimizing tardiness can be roughly divided into two categories. One aims to obtain the optimal schedules for minimizing tardiness. Shim and Kim (2007) minimized the total tardiness on identical parallel machines. They derived dominance properties and lower bounds, and developed a branch-and-bound algorithm. Tanaka and Araki (2008) further utilized the Lagrangian relaxation technique to obtain a tight lower bound for the same problem. Schaller (2009) developed a new lower bound that is based on the concepts developed by Shim and Kim (2007). Jouglet and Savourey (2011) addressed a parallel-machine total weighted tardiness scheduling problem with release dates. They developed some dominance rules based on traditional single-machine properties and showed how it is possible to deduce whether or not certain jobs can be processed by a particular machine in a particular context.

* Corresponding author.
E-mail address: jywang@sunrise.hk.edu.tw (J.-Y. Wang).

http://dx.doi.org/10.1016/j.cie.2016.05.011
0360-8352/

Other studies aimed to find approximate schedules when the problem size is large. Biskup, Herrmann, and Gupta (2008) proposed a general heuristic approach to solve several important types of parallel-machine scheduling problems and provided computational experiments for the parallel-machine total tardiness problem. Chen and Chen (2009) examined a flexible flow line problem with unrelated parallel machines at each stage and with a bottleneck stage on the line. The objective of the problem is to minimize the total tardiness. They proposed two bottleneck-based heuristics with three machine selection rules. Chiang, Cheng, and Fu (2010) addressed a parallel-machine scheduling problem with diffusion operations in a wafer fabrication facility. In the target problem, jobs arrive at the batch machines at different time instants, and only jobs belonging to the same family can be processed together. They proposed a memetic algorithm to search for the optimal or near-optimal batch formation and batch sequence for the total weighted tardiness problem. Saricicek and Celik (2011) proposed the tabu search and simulated annealing algorithms for the same problem with a job splitting property. Della Croce, Garaix, and Grosso (2012) presented a heuristic algorithm for the parallel machines total weighted tardiness problem. The algorithm combines generalized pairwise interchange neighborhoods, dynasearch optimization, and a new machine-based neighborhood whose size is non-polynomial in the number of machines. Their computational results showed that the proposed algorithm was a significant improvement over the current state of the art for this problem. Drobouchevitch and Sidney (2012) considered a problem of scheduling identical non-preemptive jobs with a common due date on parallel machines. The objective is to determine an optimal value of the due date and an optimal allocation of jobs onto machines so as to minimize a total cost function, which is the function of earliness, tardiness, and due date values. Janiak, Janiak, Kovalyov, Kozan, and Pesch (2013) studied an identical parallel-machine scheduling problem in which a common due window has to be assigned to all jobs. If a job is completed within the due window, then it incurs no scheduling cost. Otherwise, an earliness or tardiness cost is incurred. The job completion times as well as the due window location and size are decision variables. The objective is to find a job schedule as well as the location and size of the due window such that the sum of the costs associated with job earliness, job tardiness, and due window location and size is minimized.

However, power consumption or carbon emission did not receive equal attention in the above studies. Classical scheduling problems typically strive for increased efficiency in terms of job completion times. For instance, the makespan is minimized by finishing all the jobs in the shortest possible time. However, in response to the effects of global warming and environmental concerns, energy consumption has become a crucial issue. Koomey (2007) showed that the electricity used for computer servers worldwide, including their associated cooling and auxiliary equipment, cost 7.2 billion US dollars in 2005, double the cost of 2000. Therefore, the electricity issue is not negligible, for energy consumption plays a vital role in the overall cost structure. Therefore, energy consumption has drawn the attention of researchers. Liu, Zhang, Yang, Chen, and Huang (2013) indicated that the world's primary energy consumption grew 2.5%, and it was projected to increase by approximately 30% by 2030. Fang and Lin (2013) pointed out that more and more industries not only pursue better work efficiency but also focus on energy-saving issues. Slower paces are attracting more consideration in different industries. For example, in the maritime industry, ships with lower voyage speeds are much preferred due to their lower consumption of energy. Demands for lower electricity costs are also arising in the IT industry. Cloud computing service providers host a larger number of processors that are deployed to fulfill the orders of computing tasks placed by clients. In addition to customer satisfaction and efficient utilization of facilities, lower electricity consumption is also crucial to the service providers.

Outsourcing is a common strategy to avoid tardiness penalties in the field of job scheduling. For example, Lee and Sung (2008) considered two single-machine scheduling problems. One objective is to minimize maximum tardiness and outsourcing costs, and the other is to minimize the total tardiness and outsourcing costs. Neto, Godinho, and da Silva (2015) tried to minimize the sum of the outsourcing and tardiness penalty at the same time. They developed an algorithm and outsourcing was allowed. Zhang and Xu (2016) proposed a resource-constrained scheduling problem. In this problem, a construction contractor (i.e., upper-level decision maker) aims to minimize resource consumption. On the other hand, an outsourcing partner (i.e., lower-level decision maker) tries to minimize the tardiness penalty. To deal with the problem, they proposed a fuzzy method, and outsourcing was also allowed. Though outsourcing can avoid or even eliminate the tardiness penalty of a company, the total consumption of the earth's resources is still serious. Consequently, a more effective scheduling algorithm for environmental protection is called for.

Motivated by the above observations, we consider a parallel-machine scheduling problem to minimize the sum of resource consumption and outsourcing cost given that the maximum tardiness of all jobs cannot exceed a given bound. In many realistic situations, a manufacturer might own several machines with different processing speeds. Moreover, as in Ji, Wang, and Lee (2013) and Yeh et al. (2015), this problem also considers individual due dates and outsourcing in order not to cause serious total tardiness. This consideration makes this problem more practical.

The remainder of this paper is organized as follows. In Section 2, we formulate the problem. In Section 3, we consider the problem for the preemption case. In Section 4, we show that the problem is strongly NP-hard for the non-preemption case. Some dominance properties and a lower bound are proposed, and a branch-and-bound algorithm is developed. In Section 5, a hybrid genetic algorithm is provided. The computational experiments are given in Section 6 and the conclusion is presented in the last section.

2. Problem formulation

There are a set of n jobs to be scheduled on a set of m parallel machines. All the jobs are available at time 0. Each job can be processed on any machine and only once. Job j has a processing time p_j and a due date d_j. Machine i has a speed v_i and a cost β_i (i.e., resource consumption) to produce one unit of production. For instance, the resource consumption of job j is $\beta_i p_j / v_i$ if it is processed on machine i. Without loss of generality, we assume that $\beta_1/v_1 \leqslant \beta_2/v_2 \leqslant \cdots \leqslant \beta_m/v_m$. Moreover, the decision maker can outsource the jobs in order to meet the constraint that the maximum tardiness of all jobs cannot exceed a given bound B. In that case, there is an outsourcing unit cost of α, where $\alpha \geqslant \beta_m/v_m$. For instance, the cost to outsource job j is αp_j. For a job j in a schedule π, the completion time and tardiness are denoted by $C_j(\pi)$ and $T_j(\pi) = \max\{0, d_j - C_j(\pi)\}$ respectively. Under the above assumptions, the problem is to minimize the sum of resource consumption and outsourcing cost, i.e.,

$$\begin{aligned} &\min f(\pi) = \sum_{i=1}^{m} \beta_i \sum_{j \in M_i} p_j \Big/ v_i + \alpha \sum_{j \in O} p_j, \\ &\text{s.t. } \max\{T_j(\pi) | j = 1, 2, \ldots, n\} \leqslant B, \end{aligned} \tag{2.1}$$

where $j \in M_i$ means that job j is allocated to machine i, $j \in O$ means that j is an outsourced job, B is a bound, and $f(\pi)$ is the objective function.