Computers and Operations Research 98 (2018) 103–112

Contents lists available at ScienceDirect

Computers and Operations Research

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

Review

Survey and extensions of manufacturing models in two-stage flexible flow shops with dedicated machines

F.J. Hwang[a,*], B.M.T. Lin[b]

[a] School of Mathematical and Physical Sciences, University of Technology Sydney, Australia
[b] Institute of Information Management, Department of Information Management and Finance, National Chiao Tung University, Taiwan

ARTICLE INFO

Article history:
Received 18 March 2017
Revised 28 March 2018
Accepted 15 May 2018

Keywords:
Flexible flow shop
Dedicated machines
Computational complexity
Review

ABSTRACT

This study considers the manufacturing environments in which $m+1$ machines are configured as two-stage flexible flow shops with dedicated machines (F2DM). The F2DM scheduling problems arise naturally from practical production and fabrication systems, and they are classified into two categories, whose machine settings are antithetical to each other. In model 1, a single common bottleneck machine is installed at stage 1 and m parallel dedicated machines comprise stage 2. The second model has the m dedicated machines at stage 1 and the bottleneck machine at stage 2. Categorizing the literature according to the performance metrics, we survey the existing research results of the two models and propose several new solution procedures with improved computational complexity. The complexity results are summarized, and suggestions are made for future research.

1. Introduction

Since the seminal study of Johnson (1954), flow shop scheduling has been one of the most extensively studied problems in operations management (Dudek et al., 1992; Emmons and Vairaktarakis, 2013; Reisman et al., 1997). A flow shop has a number of stages organized in series. All of the jobs are processed and routed through the shop stage by stage. Among the many variants of flow shops, flexible flow shops or hybrid flow shops have multiple identical or non-identical parallel machines installed within each stage. Readers are referred to Linn and Zhang (1999); Quadt and Kuhn (2007); Ribas et al. (2010); Ruiz and Vázquez-Rodrígue (2010); Wang (2005) for comprehensive reviews and surveys on the flexible/hybrid flow shop scheduling. The scheduling of flexible/hybrid flow systems is computationally intractable in most problem settings. The complex decision-making process could be broken into three phases: (i) machine allocation, (ii) sequencing, and (iii) release timing (Pinedo, 2009). Machine allocation refers to the assignment of jobs to machines. Next, the jobs on each machine are arranged for processing according to specific sequencing rules. Whilst the processing sequences on all of the machines are known, the starting time of each job on each machine needs to be determined subject to practical constraints, including machine availability, resource availability, and job release dates. The manufacturing setting for flexible flow shops with dedicated machines corresponds to the last two phases of the decisions above. In other words, machine allocation is completed *a priori*. This study surveys two-stage flexible flow shops with dedicated machines (F2DM), in which the two stages are separately equipped with a single machine and a set of parallel dedicated machines. Given the hierarchical relationship between manufacturing environments, F2DM can be regarded as a simplification as well as the fundamentals of the corresponding flexible flow shops. From the practical perspective, F2DM scheduling has a wide range of real-world applications as discussed later. From the theoretical viewpoint, it presents challenging research issues when delineating the computational complexity boundary between easy and hard problems. Furthermore, the relevant results could provide profound insights into flexible flow shop scheduling. This study is motivated by the aforementioned practical relevance and theoretical worth, and the aim of the survey is two-fold. Firstly, to the best of our knowledge there is no relevant review work focusing on this topic, and it is worthwhile having a survey paper reporting the contributions made by the related studies. Secondly, this survey intends to depict and categorize the state of the art of F2DM scheduling and to explore the potential future research issues.

1.1. Problem formulation and notation

Consider a set of m parallel dedicated machines $M_1, \ldots, M_m$ and a single common bottleneck machine M_0. We discuss two

* Corresponding author.
E-mail addresses: feng-jang.hwang@uts.edu.au (F.J. Hwang), bmtlin@mail.nctu.edu.tw (B.M.T. Lin).

https://doi.org/10.1016/j.cor.2018.05.016
0305-0548/

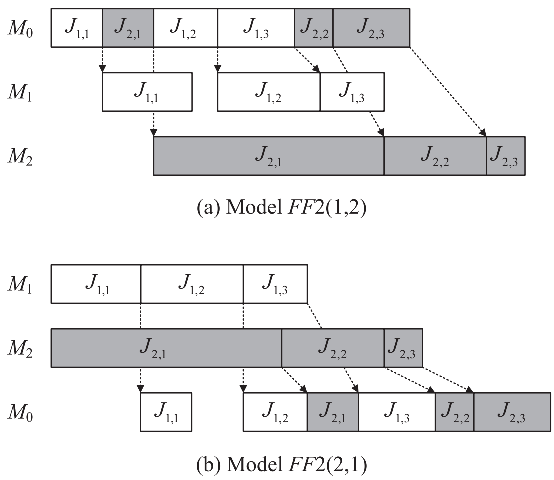

Fig. 1. Two counterpart manufacturing models of F2DM with $m = 2$.

manufacturing environments in which the $m+1$ machines are configured as two-stage flexible flow shops. In model 1, the bottleneck machine M_0 is installed at stage 1 and the m dedicated machines comprise stage 2. This configuration is denoted by $FF2(1,m)$. The second configuration, denoted by $FF2(m,1)$, has the m dedicated machines at stage 1 and the bottleneck machine M_0 at stage 2. There is a job set $\mathcal{J}$ consisting of n jobs for processing in each of the two settings. The schedules start from time zero. No preemption is allowed, and each machine can process at most one job at any time. The jobs $\mathcal{J}$ are categorized into m disjoint types of job sets $\mathcal{J}_1, \ldots, \mathcal{J}_m$, in which n_l jobs $\mathcal{J}_l = \{J_{l,1}, \ldots, J_{l,n_l}\}$ are to be processed on the bottleneck machine M_0 and their dedicated machine M_l, $l \in \{1, \ldots, m\}$ with $\sum_{l=1}^{m} n_l = n$. We denote $p_{l,j}$ and $q_{l,j}$ as the processing times of job $J_{l,j}$, $l \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n_l\}$ on the dedicated machine M_l and on the bottleneck machine M_0, respectively. When flow shop $FF2(1,m)$ is applied, all of the jobs in $\mathcal{J}$ must be processed by the first-stage bottleneck machine M_0, and the jobs in $\mathcal{J}_l$ are routed, after completing the stage-one operations, to their dedicated machine M_l at stage 2. In the counterpart $FF2(m,1)$, the jobs of each type $\mathcal{J}_l$ are processed on their dedicated machine M_l at stage 1, and after their completion they are moved to the bottleneck machine M_0 at stage 2. Fig. 1 shows Gantt charts depicting the two specific models, $FF2(1, 2)$ and $FF2(2, 1)$. It is not difficult to observe that

(1) jobs of the same type are processed on their dedicated machine and the bottleneck machine, as processed in a standard two-machine flow shop;
(2) jobs of different types compete for the processing capacity of the bottleneck machine.

These are two principal characteristics of the F2DM models. The scheduling decisions belong to two categories: (i) the sequencing of jobs $\mathcal{J}_l$ on their dedicated machine M_l for each job type $l \in \{1, \ldots, m\}$ and (ii) the sequencing of all jobs $\mathcal{J}$ on the bottleneck machine M_0. Due to multi-fold factorial growth, the difficulty of determining an optimal solution lies in the number of possible solutions which is exponentially increasing in the problem size.

In addition to the processing times, associated with each job $J_{l,j}$ are (i) weight $w_{l,j}$, reflecting the relative importance, and (ii) due date $d_{l,j}$, by which the job is expected to be finished. We denote the completion time of job $J_{l,j}$ under a specific schedule σ by $C_{l,j}(\sigma)$. The lateness and tardiness of job $J_{l,j}$ in σ are given by $L_{l,j}(\sigma) = C_{l,j}(\sigma) - d_{l,j}$ and $T_{l,j}(\sigma) = \max\{0, C_{l,j}(\sigma) - d_{l,j}\}$, respectively. The maximum lateness is $L_{\max}(\sigma) = \max\{L_{l,j}(\sigma) \mid l = 1, \ldots, m;\ j = 1, \ldots, n_l\}$. The tardiness indicator $U_{l,j}(\sigma)$ is a binary variable defined by

$$U_{l,j}(\sigma) = \begin{cases} 1, & \text{if } C_{l,j}(\sigma) > d_{l,j}; \\ 0, & \text{otherwise.} \end{cases}$$

In the $FF2(1,m)$ model, each dedicated machine M_l, $l \in \{1, \ldots, m\}$, has an associated machine weight w_l, and the machine completion time on M_l in σ is defined by $\mathsf{MC}_l(\sigma) = \max_{1 \le j \le n_l} C_{l,j}(\sigma)$. For notational simplicity, σ may be omitted if no ambiguity would arise. To denote the considered F2DM problem, we use the three-field notation $\alpha|\beta|\gamma$ (Graham et al., 1979), in which $\alpha \in \{FF2(1,m), FF2(m,1)\}$ represents either F2DM model, β describes any imposed job characteristics or processing restrictions, and γ indicates the objective function of interest such as $C_{\max}$ for the makespan, $\Sigma C_{l,j}$ for the total completion time, $\sum w_{l,j} C_{l,j}$ for the total weighted completion time, $L_{\max}$ for the maximum lateness, $\Sigma T_{l,j}$ for the total tardiness, $\Sigma U_{l,j}$ for the number of tardy jobs, $\sum w_{l,j} U_{l,j}$ for the weighted number of tardy jobs, $\sum \mathsf{MC}_l$ for the sum of the completion times of parallel dedicated machines, and $\sum \mathsf{w}_l \mathsf{MC}_l$ for the total weighted dedicated-machine completion time.

The notation *x-batch_y_z* is utilized to specify the setting of batch scheduling in the F2DM models. The *x*-batch mode with *y*-type batch composition and *z*-type batch processing fashion, in which $x \in \{\mathsf{s}, \mathsf{p}\}$, $y \in \{\mathsf{c}, \mathsf{i}\}$, and $z \in \{\mathsf{b}, \mathsf{j}\}$, is applied on the bottleneck machine M_0. If $x = \mathsf{s}$, then the sequential-batch (i.e. sum-batch) mode is considered. Otherwise, $x = \mathsf{p}$ represents the parallel-batch (i.e. max-batch) mode. If $y = \mathsf{c}$, then the compatible batch composition is assumed, viz., jobs of different types can be processed in the same batch. Otherwise, $y = \mathsf{i}$ denotes the assumption of an incompatible batch composition, viz., only jobs of the same type are allowed to be processed in one batch. If $z = \mathsf{b}$, then the batch availability is assumed, viz., all the jobs in a batch have a common completion time, when processing of the last job in that batch is finished. Otherwise, $z = \mathsf{j}$ denotes the assumption of job availability, i.e., the completion of some job in a batch depends on the time at which it is completed. For the comprehensive concept of batch scheduling and the relevant solution techniques, readers are referred to Potts and Kovalyov (2000); Potts and Van Wassenhove (1992).

For the F2DM scheduling problems in which transportation between stage 1 and stage 2 with a single transporter or conveyor is considered, we use *trans* in the β field. Denote by c the capacity of the transporter or conveyor for transportation operations between the two stages, i.e., the transporter can carry up to c jobs in one shipment. The transportation time, i.e. the duration of one trip, from stage $i \in \{1, 2\}$ to the other stage is assumed to be independent of the jobs being conveyed or the dedicated machine from/for which the transporter departs and is denoted by t_i.

Notation:

- m: number of parallel dedicated machines
- M_l: dedicated machine l, $l \in \{1, \ldots, m\}$
- M_0: bottleneck machine
- n_l: number of type-l jobs, $l \in \{1, \ldots, m\}$
- n: total number of jobs, i.e. $n = \sum_{l=1}^{m} n_l$
- $J_{l,j}$: job j of type l, $l \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n_l\}$
- $p_{l,j}$: processing time of job $J_{l,j}$ on the dedicated machine M_l
- p_l: common processing time per job type l on the dedicated machine M_l
- p: common processing time for all jobs on dedicated machines
- $q_{l,j}$: processing time of job $J_{l,j}$ on the bottleneck machine M_0
- q_l: common processing time per job type l on the bottleneck machine M_0