Computers & Operations Research 39 (2012) 320–327

Contents lists available at ScienceDirect

Computers & Operations Research

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

A cutting plane approach for integrated planning and scheduling

T. Kis *, A. Kovács

Computer and Automation Institute, Kende str. 13-17, H-1111 Budapest, Hungary

ARTICLE INFO

Available online 14 April 2011

Keywords:
Production planning and scheduling
Mathematical programming
Cutting planes
Parallel machine scheduling

ABSTRACT

In this paper we propose a branch-and-cut algorithm for solving an integrated production planning and scheduling problem in a parallel machine environment. The planning problem consists of assigning each job to a week over the planning horizon, whereas in the scheduling problem those jobs assigned to a given week have to be scheduled in a parallel machine environment such that all jobs are finished within the week. We solve this problem in two ways: (1) as a monolithic mathematical program and (2) using a hierarchical decomposition approach in which only the planning decisions are modeled explicitly, and the existence of a feasible schedule for each week is verified by using cutting planes. The two approaches are compared with extensive computational testing.

1. Introduction

Hierarchical production planning and scheduling deals with tactical and operational decisions. The two types of decisions differ in their scope and time horizon [1]. We focus on *planning* on a weekly basis the objective being to determine the most cost effective way of distributing the workload between the weeks, while *scheduling* is concerned with allocating resources to jobs to be performed during the same week. The main advantage of hierarchical planning and scheduling is that at each decision level, only the most relevant information is used. E.g., when taking planning decisions, resource capacities are aggregated and the fine details of dealing with single resources are neglected. In contrast, when solving scheduling problems, only the weekly or daily assignments have to be scheduled [2]. It is often mentioned that these decisions are worth to be separated to ease the work of decision makers at either level. However, the two types of decisions are strongly related, since both the overloading and the underloading of the weekly production capacities have undesired effects. Namely, if the weekly assignment cannot be met, then the plan has to be reworked. On the other hand, a loose plan may cause unnecessary delays and thus incurs penalties which could be avoided by more careful planning. To remedy this situation, *integrated planning and scheduling* has been suggested by various authors [3,4].

We will study a scheduling problem in a parallel machine environment, where each job has a release time and a due-date, the release time being the first week of the time horizon where the job may be started and the due-date is the week where the job should be completed. Each job has to be assigned to a week and those jobs assigned to a given week must be scheduled on the parallel machines so that the load of every machine is no more than one week. The objective is to minimize the earliness/tardiness penalty costs incurred by completing some of the jobs before or after their due-dates. Albeit this setting is a simplification of real-world planning and scheduling problems, where there may be additional constraints on feasible solutions, the decomposition approach proposed in this paper may be generalized to richer problem formulations, and our main purpose here is to asses its merits in a "laboratory" environment.

While most of the known hierarchical approaches for solving hard scheduling problems reduce the problem size by decomposing the problem along the resources, our approach decomposes the problem along the types of decisions: the upper level assigns the jobs to weeks, and the lower level schedules the jobs assigned to a given week. Though this is a very natural decomposition approach, the computational advantages are not apparent at once. We use a compact problem formulation in which the decision variables represent only the assignment of jobs to weeks; but there will be no explicit variables for representing the schedule of those jobs assigned to the same week. Instead, we verify whether those jobs assigned to the same week can be completed during one week by using cutting planes, or as a last resort, by solving a parallel machine scheduling problem. In contrast to most previous approaches, we generate not only infeasibility or "no-good" cuts, but other problem specific cuts as well, and we try to generate violated cuts not only when an integer solution is found, but in all search-tree nodes.

After a brief literature review (Section 2), we provide a formal problem statement in Section 3. In Sections 4 and 5 we propose two alternative formulations: a monolithic mathematical program, and a compact one suitable for decomposition, respectively. To strengthen

* Corresponding author. Tel.: +36 1 2796156; fax: +36 1 4667503.
E-mail addresses: tamas.kis@sztaki.hu (T. Kis), akovacs@sztaki.hu (A. Kovács).

0305-0548/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2011.04.006

the second formulation, we derive cutting planes from lower bounds for the bin-packing problem (Section 5.1), along with separation algorithms (Section 5.2). The cutting planes are used in a branch-and-cut algorithm (Section 6), whose effectiveness is compared to solving the integrated planning and scheduling problem as a monolithic mathematical program in Section 7.

2. Literature review

2.1. Parallel machine scheduling and bin-packing

By the parallel machine scheduling problem we mean the minimization of the makespan of n jobs on m identical parallel machines. For parallel machine scheduling, the worst case performance of $4/3-1/(3\,\mathrm{m})$ of the LPT rule (longest processing time first) is derived by Graham [5]. We will heavily exploit the strong connection between the parallel machine scheduling and the bin-packing problems, see Coffman et al. [6]. In that paper a new algorithm, called MULTIFIT, is presented, which uses ideas from bin-packing algorithms, and it is shown that it produces a schedule of makespan at most 1.22 times the optimum. However, its running time is larger than that of LPT, since in each iteration the FIRST-FIT-DECREASING bin-packing algorithm is run, which takes as much time as a single run of the LPT heuristic for parallel machine scheduling, and the desired number of iterations is about 7 for a large number of machines (over 8). This connection is pushed further by Hochbaum and Shmoys by developing the first polynomial time approximation scheme for the parallel machine scheduling problem [7]. In contrast to previous approaches, no weight function over the jobs is applied when deriving the approximation ratio of the algorithm. A thorough survey of approximation algorithms for bin-packing can be found in [8]. The lower bounds L_1 and L_2 (for bin-packing) are proposed in [9] to be used in exact algorithms. These bounds are enhanced in [10]. These lower bounds will be used in Section 5.1, where we give their precise definitions.

A cutting plane based approach for solving the parallel machine scheduling problem is proposed by Mokotoff [11]. The novelty of the method is a cutting plane which is valid for a specific face of the single-node fixed charge network model, and in fact can be derived from the well-know flow-cover inequality [12].

2.2. Hierarchical decomposition

Hierarchical decomposition approaches are applied widely in the field of production planning and scheduling. Although the decisions made on the different levels are strongly related, solving these problems in an integrated way is often considered to be computationally intractable. It is therefore typical to apply single- or multi-pass heuristics. In the single-pass case, one fixed upper level plan is unfolded on the lower level, see e.g., [2,13]. Obviously, a shortcoming of this approach is that bad planning decisions may result in situations where no detailed schedules can meet all production goals. Multi-pass heuristics aim at relieving such situations by iterating between the two levels, and modifying the upper level plan according to the problems identified in the previous iteration [14,3]. Sawik [15] compares monolithic and hierarchical MIP formulations of an assembly line scheduling problem. In the hierarchical model, the upper level assigns jobs to resources and the lower level sequences them. The two levels are joined in a single-pass heuristic, and computational experiments have shown that the quicker hierarchical decomposition approach finds optimal solution for most of the instances.

Subsequently, we focus on exact solution methods that use hierarchical decomposition. One of the problems frequently addressed is the *multi-machine assignment and scheduling problem* (MMASP): a set of jobs, characterized by individual time windows, are to be scheduled on unrelated parallel machines to minimize the total assignment cost. In all of the following papers, the master problem assigns jobs to machines, while a separate subproblem belongs to each machine, sequencing the jobs on that machine. Jain and Grossmann [16] apply a MILP/CP approach, and add an infeasibility or “no-good” cut for the complete set of jobs scheduled on the machine where infeasibility is detected. Hooker and Ottoson [17] introduce logic-based Benders decomposition, and illustrate the approach on MMASP. The same type of infeasibility cuts is used, though an indication is made that these cuts can be strengthened based on the CP proof of infeasibility. Sadykov and Wolsey [18] compare several monolithic and MIP/CP hybrid decomposition approaches. The new results include a tight MILP formulation. Their decomposed approaches detect infeasibility or “no-good” cuts in internal nodes of the branch-and-bound tree, after a suitable rounding of the LP solutions. Sadykov [18] investigates the solution of the one-machine subproblem of the above multi-machine assignment problem, which corresponds to $1|r_j|\sum w_jU_j$. Two new classes of cuts are introduced for this problem. The first class is infeasibility cuts of low cardinality, which are found by a modified version of Carlier’s branch-and-bound algorithm [19]. The second class consists of a completely different type of cuts based on the edge-finding constraint propagation rule. Bockmayr and Pisaruk [20] investigate the problem of generating infeasibility cuts by CP for MILP in a general setting. The application of these ideas to MMASP leads to infeasibility cuts. MMASP has been generalized to cumulative resources in [21], and solved by a hybrid MIP/CP approach following the above decomposition scheme. MMASP is extended to multi-stage processes in [22]. The same assign/schedule decomposition approach is taken. The main difference due to the multi-stage processes is that the single-machine subproblems are no longer independent, hence, a single subproblem involving all machines and jobs is solved, but the resulting cuts may not be valid and cut off the optimal solution. A different, multi-product continuous plant scheduling problem with a single processing unit, subject to sequence-dependent setup times, is discussed in [23]. A decomposition approach is proposed, where the upper level sets production levels and inventories for macro time periods, and the lower level sequences the production activities. If the lower-level problem proves infeasible, then integer and logic cuts are fed back to the upper level. Both levels are described by and solved as a MILP.

Artigues et al. [24] investigate a hybrid decomposition based approach for an integrated employee timetabling and job-shop scheduling problem which is an extension of the classical job-shop scheduling problem. A decomposition-based CP formulation is proposed, which assigns jobs (possibly partially) to time periods (shifts). Guyon et al. [25] study a similar problem. In the proposed solution approach, there is a master problem for creating a timetable for the employees, while the subproblem checks if a feasible job schedule exists for the given timetable. It is exploited that the subproblem corresponds to a maximum flow problem, and hence, a minimum cut is fed back to the master problem upon infeasibility. An initial set of cuts is generated in a pre-processing step.

A review of solution approaches has been presented by Grossmann et al. [26]. The possible ways of integrating production planning and scheduling are surveyed in [4].

3. The integrated production planning and scheduling problem

In this section we give a formal definition of the scheduling problem studied in the paper. Suppose that the time horizon is divided into τ equal length periods. The common length of the