Scheduling with Setup Considerations:
An MIP Approach

Mohamed. K. Omar, Siew C. Teo and Yasothei Suppiah
Centre of Computer Aided Design and Knowledge Manufacturing (CCADKM)
Faculty of Engineering and Technology, Multimedia University
Malaysia

1. Introduction

Competitions and ever-changing customer requirements are the driven forces behind manufacturers to reevaluate their planning and scheduling methods and manufacturing systems. Customers’ satisfaction in most cases can be measured by the ability of the manufacturing firms to provide goods with reasonably good prices, acceptable quality standard and deliver at the right time. Scheduling plays an important role in all of the important issues that are considered to measure customers’ satisfaction. In recent years, there has been an increased interest in production planning problems in the multi product chemical and pharmaceutical industry. Multi product chemical plants use either a continuous production system or a batch production system. Batch process plants involve small amounts of a large variety of finished products, therefore are suitable for the production of small-volume, high-value added products. In such industry, products are often grouped into incompatible product families, where an intensive setup is incurred, whenever production changes from one product family to another.

A classical example of the multi product chemical plants is the manufacturing of resins. Typically, in the resin production environment, the planning and scheduling task starts by considering a set of orders where each order specifies the product and the amount to be manufactured as well as the promised due date. The most important task of the planner is the so-called batching of orders. Batching of orders is the process of transforming customers’ product orders into sets of batches to be planned and subsequently assigned due date. This process is commonly practiced in the industry such as this, since a batch is frequently shared by several orders with the earliest one determining the batch due date. Moreover, while the planner is carrying out this task, his/her objective is to minimize as much as possible the setups between products that are generated from incompatible families. Therefore, in such manufacturing environment, setup activities cannot be disregarded and the production range is usually composed of a number of incompatible product families, in a way that no setup is required between production of two products belonging to the same family; long and expensive setup operations are required otherwise.

Scheduling is known as a decision-making process of allocating limited resources over time in order to perform a collection of tasks for the purpose of optimizing certain objectives functions (Baker 1974). Tasks can have difference in their priority levels, ready time, and
process times. The objective function could be, for example, minimizing completion time, minimizing the number of tardy jobs, or adopting the (JIT) concepts and calls for minimization of earliness and tardiness. There are two issues associated with scheduling problems: how to allocate jobs on machines and how to sequence jobs on each machine. Therefore, the scheduler is mainly concerned with allocation decisions and sequencing decisions. On another issue, one must state at this stage that there is a difference between sequencing and scheduling. Sequencing corresponds to a permutation of the job set in which jobs are processed on a given machine. While scheduling is defined as an allocation of jobs within a more complicated setting of machines, which could allow for preemptions of jobs by other jobs that are released at a later point of time.

In the scheduling literature, setups have for long been considered negligible and hence ignored, or considered as part of the process time. But there are situations that call for treating the setups separately from the process time. In such cases, two problem types exist. In the first type, setups depend only on the job to be produced; hence, it is called sequence-independent. In the second type, setups depend on both the job to be processed and the immediate preceding job; hence it is called sequence-dependent.

This paper aims to explore the scheduling and sequencing problem confronted by planners in the multi product chemical plants that involve sequencing of jobs originated from incompatible families making it a situation that requires sequencing of jobs with sequence-dependent setup time. Our intension is to focus on these types of scheduling problems and suggest two mixed integer programming (MIP) formulations. The first formulation considers a single machine situation and aims to minimize total tardiness, while the second formulation attempts to minimize the sum of total earliness/tardiness for parallel machine situation.

This paper is organized as follows: Section 2 presents the literature review. Section 3 introduces a typical multi product chemical production environment. Section 4 presents problem description and formulation. We present our computational example in Section 5. Finally, we present our conclusions and remarks in Section 6.

2. Literature review

Enormous solutions have been proposed for machine scheduling problems, and we do not attempt to cover it all here. However, interested readers are referred to the reported reviews by Allahverdi et al. (1999), Yang and Liao (1999), Cheng et al. (2000), Potts and Kovalyov (2000) and Allahverdi et al. (in press). However, we will provide a brief review related to our work for total tardiness for single machine and the case of earliness/tardiness for parallel machines situation.

2.1 Single machine total tardiness problem

Tardiness is the positive lateness a job incurs if it is completed after its due date and the objective is to sequence the jobs to minimize total tardiness. In the weighted case, each job’s tardiness is multiplied by a positive weight. The weighted tardiness problem in a single machine is NP-hard in the strong sense (Lenstra et al. (1977)). Adding the characteristics of jobs originated from incompatible families increases the difficulty of the problem of minimizing the total weighted tardiness on a single machine. Many practical industrial situations require the explicit consideration of setups and the development of appropriate
scheduling tools. Among the reported cases, Pinedo (2002) describes a manufacturing plant making paper bags where setups are required when the type of bag changes. A similar situation was observed in the plastic industry by Das et al. (1995). The aluminium industry has a casting operation where setups, mainly affecting the holding furnaces are required between the castings of different alloys (see Gravel et al. (2000)).

Previous research done in the case of incompatible job families has been focused mostly on single batch machine problems. Fanti et al. (1996) focused on makespan as the performance measurement. Kemf et al. (1998) investigated a single machine having a second resource requirement, with a goal of minimizing makespan and total completion time. Dobson and Nambimodom (2001) considered the problem of minimizing the mean weighted flow time and provided an integer programming (IP) formulation. Mehta and Uzsoy (1998) presented a dynamic programming (DP) algorithm as well as heuristics that can provide near optimal solutions where the performance under analysis is total tardiness. Azizoglu and Webster (2001) describe a branch and bound procedure to minimize total weighted completion time with arbitrary job sizes. Their procedure returns optimal solutions to problems of up to 25 jobs. Most recently, Perez et al. (2005) developed and tested several heuristics to minimize the total weighted tardiness on single machines with incompatible job families. Their tests consistently show that the heuristics that uses Apparent Tardiness Cost (ATC) rule to form batches, combined with Decomposition heuristics (DH) to sequence jobs, perform better than others tested, except ATC combined with Dynamic Programming algorithms (DP). Their tests show that ATC-DH and ATC-DP results are close.

The literature is also not extensive either for single machine scheduling problems with sequence-dependent setups, where the objective is to meet delivery dates or to reduce tardiness. However, Lee et al. (1997) have proposed the Apparent Tardiness Cost with Setups (ATCS), a dispatching rule for minimizing weighted tardiness. Among other authors who have treated the problem, we find Rubin and Raagatz (1995) developed a genetic algorithm and local improvement method while Tan and Narasimhan (1997) used simulated annealing as a solution procedure. Tan et al. (2000) presented a comparison of four approaches and concludes, following a statistical analysis, that a local improvement method offers a better performance than simulating annealing, which in turn is better than branch-and-bound. In this comparison, the genetic algorithm had the worst performance.

2.2 Parallel machines with earliness/tardiness problem

Another scheduling approach that considers job earliness and tardiness penalties is motivated by the just-in-time concept (JIT). This approach requires only the necessary units to be provided with the necessary quantities, at the necessary times. Production of one extra unit is as bad as being one unit short. In today’s manufacturing environments, many firms are required to develop schedules that complete each customer’s order at, or near, its due date, and at the same time to ensure the cost-efficient running of the plant.

There are a large number of published research papers that consider scheduling problems, with both earliness and tardiness penalties. These include models with common due dates or distinct due dates, with common/symmetrical penalty functions as well as distinct job dependent penalty functions. Except for a few basic models, most of these scheduling problems have been shown to be NP-Hard. Readers are referred to the work of Webster [1997] and Chen [1997] for discussion, about the complexity boundaries of these problems. Readers interested in earliness-tardiness scheduling are referred to the survey conducted by
Baker and Scudder [1990] and the recent book by T’kindt and Billout [2000]. Readers especially interested in earliness and tardiness scheduling with setup considerations, are referred to the survey article by Allahverdi et al. [in press]. However, we summarize below some published works related to earliness and tardiness scheduling problems considered in this paper.

Kanet [1981] examined the earliness and tardiness problem, for a single machine, with equal penalties and unrestricted common due dates. A problem is considered unrestricted, when the due date is large enough not to constrain the scheduling problem. He introduced a polynomial time algorithm to solve the problem optimally. Hall [1986] extended Kanet’s work and developed an algorithm that finds a set of optimal solutions for the problem based on some optimality conditions. Hall and Posner [1991] solved the weighted version of the problem with no setup times. Azizoglu and Webster [1997] introduced a branch-and-bound algorithm to solve the problem with setup times. Other researchers who worked on the same problems with a restricted (small) due date, included Bagchi et al. [1986], Szwarc [1989, 1996], Hall et al. [1991], Alidee and Dragan [1997] and Mondal and Sen [2001]. None of the previous papers consider sequence-dependent setup times.

The majority of the literature on earliness and tardiness scheduling deals with problems that consider single machine only. Problems with multiple machines have been investigated in only a handful of papers which includes among others, Emmons [1986], Cheng and Chen [1994], De et al. [1994], Li and Cheng [1994], Kramer and Lee [1994], Federgruen and Mosheiov [1996, 1997], Adamopoulis and Pappis [1998] and Chen and Powell [1999]. To the best of our knowledge, there have been very few publications that propose a mixed integer programming solution for parallel machines that consider setup for the earliness and tardiness scheduling problem. Balakrishnan et al. [1999] considers the problem of scheduling jobs on several uniform parallel machines and presented a mixed integer programming formulation. However, their reported experiments show that their approach cannot solve a problem with more than 10 jobs. More recently, Zhu and Heady [2000] proposed a mixed integer programming formulation for minimizing job earliness and tardiness scheduling problem for a non-uniform parallel machine and setup considerations. However, their reported experiments show that their approach cannot solve a problem with more than 10 jobs. Furthermore, their reported formulation suffers from a serious error making their reported job/machine assignment and sequential job orders questionable. And the work of Omar and Teo (2006) whom they corrected Zhu and Heady (2000) and proposed an improved MIP formulation for minimizing the sum of earliness/tardiness in identical parallel machine. Their tests show that their proposed formulation can provide optimal solution for 18 jobs originated from 4 incompatible families.

3. Production environment

A resin manufacturing company in South East Asia will be used to illustrate the production environment. The plant has two production lines and the major types of production reactions include Alklylation, Aicylction and Aminition, leading to the production of over 100 finished products. Figure 1 show the structure of the most active 20 products which are generated from 5 incompatible families.

The plant operates on three shifts, and each production year has 358 days. Working capacity is around 742 tons and 663 tons per month for line one and line two respectively. The operation in each production line is a reaction process, where the chemical reaction takes
place in a reactor; mixing where chemicals are mixed in a thinning tank; filtering, where purities are controlled to meet customer’s request; and packaging. Reaction is the bottleneck operation, hence the working capacities estimated, are based on the reaction process. Demand of finished products is considered to be high and therefore, all products cannot be satisfied from production runs, since some of the available capacity is consumed for setups. The workforce involved on the production is very limited and each shift requires 7 persons to run the process and the company does not practice workforce variation policies.

When the demand estimates for the next year are ready, the marketing division passes these estimates to the production division to prepare the operational budget for the next year. The order batching process starts when the production planner receives customers’ orders with due dates. The ultimate objective of this process is to meet the customers due dates and minimize setup activities. Interested readers are referred to the work reported by Omar and Teo (2007) for detailed solution for the planning and scheduling problem described in this section.

4. Problem formulation

In the production environment described above, the scheduling and sequencing problem can be formulated in various ways. We will present two different formulations that reflect some management policies that the company might wish to implement. First, the management might wish to implement a product/production line dedication policy, and in that case, the two production lines will be treated as a two separate single production lines,
or in another world, two separate single machine situation. For this case, we will provide an MIP modeling approach that aims to minimize total tardiness. In the second case, the company may consider examining the idea which assumes that all products can be at any instant of time produced in any of the production lines. In such a case, we will provide an MIP modeling approach that treats this situation as identical parallel machines and aims to minimize the total earliness and tardiness.

It is worth noting that MIP codes have a weakness when confronted with real life industrial scheduling and sequencing problems that involve hundreds of products, since the computational time will increase exponentially as the number of integer variables increase. Consequently, the decision maker may not be able to obtain results in real time to be of any use for implementation purposes. However, MIP codes are beneficial to researchers for testing the performance of their developed heuristics, which are normally developed for industrial application and tested against other heuristics, a dangerous procedure practiced by researchers (see Ovacik and Uzsoy (1994)).

4.1. Single machine problem formulation

Notations

Parameters

\( m \) = number of families.
\( n_i \) = number of jobs in family \( i \).
\( n \) = total number of jobs
\( d_{ij} \) = due date of \( j^{th} \) job in family \( i \).
\( p_{ij} \) = processing time of \( j^{th} \) job in family \( i \).
\( s_i \) = setup time of family \( i \).

Decision Variables:

\( X_{ijk} \) = \( \begin{cases} 1 & \text{if job } j \text{ from family } i \text{ is placed in position } k \\ 0 & \text{otherwise} \end{cases} \)

\( Y_{ik} \) = \( \begin{cases} 1 & \text{if setup } s_i \text{ is needed before a job at position } k \\ 0 & \text{otherwise} \end{cases} \)

\( C_{ik} \) = completion time of the job at position \( k \).
\( T_{ijk} \) = tardiness of the \( j^{th} \) job in family \( i \) at position \( k \)

Formulation

\[
\text{Min } \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k=1}^{n} T_{ijk} \tag{1}
\]

Subject to:

\[
\sum_{i=1}^{m} \sum_{k=1}^{n} X_{ijk} = 1 \quad k = 1, 2, ..., n \tag{2}
\]

\[
\sum_{k=1}^{n} X_{ijk} = 1 \quad i = 1, 2, ..., m; \ j = 1, 2, ..., n_i \tag{3}
\]
In the above formulation, equation (1) represents the objective function, which is to minimize total tardiness. Equations (2) and (3) state that each position can be occupied by only 1 job and each job can be processed only once. Equation (4) checks whether or not the preceding job and the following job are from the same family. If so, there is no setup time between them. Otherwise, a family setup time of the job in position $k$ exists. Equation (5) states the completion time of the job in the first position. Equation (6) calculates the completion time from the second position to the last position of the sequence. Equation (7) determines the tardiness values for all positions. Equations (8) and (9) give the non-negativity constraints.

4.2. Parallel machines problem formulation

**Notations**

**Parameters:**

$m$ = number of families.

$r$ = number of production lines.

$n$ = total number of jobs.

$f_j$ = family of job $j$. $f_j = 1, 2, \ldots, m$

$d_j$ = due date for job $j$.

$p_{jl}$ = processing time of job $j$ at production line $l$.

$e_j$ = earliness penalty for job $j$.

$t_j$ = tardiness penalty per period for job $j$.

$S_{jk}$ = setup time from family of job $j$ to family of job $k$.

$G_{jk} = \begin{cases} s_{jk} & \text{if } f_j \neq f_k \\ 0 & \text{if } f_j = f_k \end{cases}$

$M = A$ large positive number.
Decision Variables:

\( E_j \) = earliness for job \( j \).

\( T_j \) = tardiness for job \( j \).

\( C_j \) = completion time of job \( j \).

\( \alpha_{jl} = \begin{cases} 1 & \text{if job } j \text{ is the first processed in line } l. \\ 0 & \text{otherwise} \end{cases} \)

\( \theta_{jk} = \begin{cases} 1 & \text{if job } k \text{ has been scheduled right after job } j. \\ 0 & \text{otherwise} \end{cases} \)

\( \beta_{jl} \) : Continuous variable restricted to the range [0, 1], denoting that job \( j \) has been scheduled in line \( l \) but not in first place.

Formulation

Min

\[
\sum_{j=1}^{n} (E_j E_j + t_j T_j)
\] (10)

Subjects to:

\[
\sum_{l=1}^{r} (\alpha_{jl} + \beta_{jl}) = 1, \ j = 1,2,\ldots,n
\] (11)

\[
\alpha_{jl} + \beta_{jl} \leq \beta_{kl} + 1 - \theta_{jk}, \ j = 1,2,\ldots,n; k = 1,2,\ldots,n; l = 1,2,\ldots,r
\] (12)

\[
\sum_{j=1}^{n} \alpha_{jl} = 1, \ l = 1,2,\ldots,r
\] (13)

\[
\sum_{j=1}^{n} \alpha_{jl} + \sum_{l=1}^{r} \theta_{jl} = 1, \ j = 1,2,\ldots,n
\] (14)

\[
\sum_{l=1}^{r} \theta_{jl} \leq 1, \ j = 1,2,\ldots,n
\] (15)

\[
C_i \geq C_j + G_{ik} + \sum_{l=1}^{r} p_{jl} \beta_{jl} + M\theta_{ik} - M, \ j = 1,2,\ldots,n; k = 1,2,\ldots,n
\] (16)

\[
C_j \geq \sum_{l=1}^{r} p_{jl} (\alpha_{jl} + \beta_{jl}), \ j = 1,2,\ldots,n
\] (17)

\[
C_j + E_j - T_j = d_j, \ j = 1,2,\ldots,n
\] (18)

\[
d_j \theta_{jk} \leq d_k \theta_{jk}, \ j = 1,2,\ldots,n; k = 1,2,\ldots,n
\] (19)

\[
C_j, E_j, T_j \geq 0, \ j = 1,2,\ldots,n
\] (20)

In the above formulation, equation (10) represents the objective function, which is to minimize the weighted sum of earliness-tardiness. Equation (11) states that each job must be assigned to one production line. Equation (12) enforces both the job and its direct successors in the processing sequence to be manufactured on the same line. Equation (13) states that each job, if first, can only be processed first on one line. Equation (14) enforces that a job is
either the first to be processed, or succeed another in the processing sequence. Equation (15) ensures that every job should at most be directly succeeded by another job in the processing sequence, unless it is last in the sequence. Equation (16) ensures that the processing start time for a job can never be lower than the completion time of its direct predecessor job in the processing sequence. Equation (17) states that completion time of a job must be later or equal to its processing time. Equation (18) measures the degree to which each job is tardy or early. Equation (19) states that the due date of a job must be the same or earlier than its direct successor job. Equation (20) is the non-negativity constraint.

5. Computational example

The models are illustrated using the data shown in Tables 1 and 2. The data presented in Table 1 is for a scheduling and sequencing problem that consists of 10 jobs that can be originated from 2, 3 or 4 incompatible families. As it could be seen in Table 1, the setup time required when the production runs change from one family to another is fixed. On the other hand, Table 2 is used to create variable setup times among the different product families. It is worth noting that in our computations for parallel machines situation, earliness penalty was calculated using the value of \((J+1)^{-1}\) whereas the tardiness penalty was kept to be equal to one.

<table>
<thead>
<tr>
<th>10 jobs originated from 2 incompatible families</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>DD</td>
</tr>
<tr>
<td>Setup</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10 jobs originated from 3 incompatible families</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>DD</td>
</tr>
<tr>
<td>Setup</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10 jobs originated from 4 incompatible families</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>DD</td>
</tr>
<tr>
<td>Setup</td>
</tr>
</tbody>
</table>

J=Job, F=Family. P=process time in hours. DD=Due date in hours. Setup in hours.

Table 1. Ten jobs originated from different incompatible families with constant setup time
In this study, the MIP models were developed using OPL Studio version 3.6 and solved using CPLEX version 8. Models were executed with Pentium IV 2.80Hz. processor, while Microsoft Excel is used to export and import data and solution. The data required by the developed MIP models, and to be used for illustrative purposes is presented in Tables 1 and 2.

<table>
<thead>
<tr>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>F3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Families setup time matrix

<table>
<thead>
<tr>
<th>No. of jobs</th>
<th>No. of families</th>
<th>Sequence</th>
<th>Total tardiness</th>
<th>No. of setups</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>1→4→8→6→10→5→2→3→9→7</td>
<td>141</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>8→1→4→2→5→6→10→9→3→7</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>1→8→6→4→5→10→9→2→3→7</td>
<td>154</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3 Summary of the computational results
6. Concluding remarks

The results of the performance of the developed MIP for single and parallel machines are summarized in Table 3 and the sample of the results is presented in Gantt chart is shown in Figures 2-5. Examining the results presented in Table 3 reveals that for all cases tested, total tardiness and number of setups increases as the number of incompatible families involved
in the scheduling activities increases. As it is expected, when more resources are involved in the scheduling activities, the total tardiness will decrease, when an additional machine is added, tardiness improved almost 5 times.

As a summary, in this research we presented the characteristics of an industrial environment where products (jobs) that originated from incompatible families are required to be scheduled and sequenced to meet some predetermined due dates. We presented two MIP formulations to solve such scheduling and sequencing requirements. The first MIP formulation aims to minimize the total tardiness while the second MIP formulation adopts the just-in-time concept and calls for minimizing the sum of earliness and tardiness for parallel machines. Moreover, we presented computational examples that consider fixed and variable setup times when the production runs changes from one product family to another.

It is worth noting that with MIP models, computational time grows in an almost exponential manner as the problem size is increased. This known fact is considered as a major drawback for using MIP in real industrial application, none the less, MIP models are the only way to check optimality of heuristics solutions employed to solve industrial size scheduling problems.

7. References


A major goal of the book is to continue a good tradition - to bring together reputable researchers from different countries in order to provide a comprehensive coverage of advanced and modern topics in scheduling not yet reflected by other books. The virtual consortium of the authors has been created by using electronic exchanges; it comprises 50 authors from 18 different countries who have submitted 23 contributions to this collective product. In this sense, the volume can be added to a bookshelf with similar collective publications in scheduling, started by Coffman (1976) and successfully continued by Chretienne et al. (1995), Gutin and Punnen (2002), and Leung (2004). This volume contains four major parts that cover the following directions: the state of the art in theory and algorithms for classical and non-standard scheduling problems; new exact optimization algorithms, approximation algorithms with performance guarantees, heuristics and metaheuristics; novel models and approaches to scheduling; and, last but least, several real-life applications and case studies.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: