Hybrid Job Shop and parallel machine scheduling problems: minimization of total tardiness criterion

Frédéric Dugardin, Hicham Chehade, Lionel Amodeo, Farouk Yalaoui and Christian Prins
University of Technology of Troyes, ICD (Institut Charles Delaunay)
France

1. Introduction

Scheduling is a scientific domain concerning the allocation of limited tasks over time. The goal of scheduling is to maximize (or minimize) different criteria of a facility as makespan, occupation rate of a machine, total tardiness … In this area, scientific community usually group the problem with, on one hand the system studied, defining the number of machines (one machine, parallel machine), the shop type (as Job shop, Open shop or Flow shop), the job characteristics (as pre-emption allowed or not, equal processing times or not) and so on. On the other hand scientists create these categories with the definition of objective function (it can be single criterion or multiple criteria). The main goal of this chapter is to present model and solution method for the total tardiness criterion concerning the Hybrid Job Shop (HJS) and Parallel Machine (PM) Scheduling Problem.

The total tardiness criterion seems to be the crux of the piece in a society where service levels become the central interest. Indeed, nowadays a product often undergoes different steps and then traverses different structures along the supply chain, this involve in general a due date at each step. This can be minimized as a single objective or as a part of a multi-objective case.

On the other hand, the structure of a hybrid job shop consists in two types of stages with single and parallel machines. That is why we propose to point out the parallel machine PM problem domain which can be used to solve the hybrid job shop scheduling system. This hybrid characteristic of a job shop is very common in industry because of two major factors: at first some operations are longer than other ones and secondly flexible factory. Indeed, if some operations too long; they can be accelerated by technical engineering but if it is not possible they must be parallelized to avoid bottlenecks. Another potential cause is the flexible factory: if a factory does many different jobs these jobs can perhaps pass through a central operation and so the latter must increase his efficiency.

This work is organized as follow: firstly a state of the art concerning PM is realized. The latter leads us to the HJS problem where we summarize a state of the art on the minimization of the total tardiness and in a second step we present several results concerning efficient heuristic methods to solve the Hybrid Job Shop problem such as Genetic Algorithm or Ant Colony System algorithm. We also deal with multi-objective

optimizations which include the minimization of total tardiness using the NSGA-II see Deb et al., (2000). The Hybrid Job Shop Parallel Machine Scheduling rises in actual industrial facilities, indeed some of the results presented here have direct real application in a printing factory. Here the hybridization between the parallel machine stage and the single stage is provided by the printing and the winding operations which proceed with more jobs than cutting and packaging operations.

To put it in a nutshell, this chapter presents exact and approximate results useful to solve the Hybrid Job Shop problem with minimization of total tardiness.

2. Problem formulation

The hybrid job shop problems or the flexible job shop problem are various considered in this document can be shown using the classical notation $HJS_n \mid \text{prec}, S_{mj}, r_j, d_j \mid \sum T_j$. It can be formulated as follow: $n$ jobs ($j = 1, ..., n$) have to be processed by $m$ machines ($i = 1, ..., m$) of different types gathered in $E$ groups. In this case two types of groups are considered: groups with single machines and groups with identical parallel machines.

Each job has a predetermined route that it has to follow through the job-shop. Only one operation for a job can be processed in a group. The maximal number of operations is equal to the number of groups. All the machines are available at the initial time $0$. No order priority is assigned to the job.

The processing of job $j$ on machine $i$ is referred to as operation $O_{ij}$, with processing time $p_{ij}$. The processing times are known in advance. Job $j$ has a due date $d_j$ and a release date $r_j$, respectively, the last job operation completion time and the first job operation availability. No job can start before its release date and its processing should not exceed its due date. If operation $O_{ik}$ immediately succeeds operation $O_{ij}$ on machine $i$, a setup time $S_{ij,k}$ is incurred. Such setups are sequence dependent see Yalaoui (2003) and $S_{ij,k}$ need not be equal to $S_{ij,k}$. Let $C_j$ denote the completion time of job $j$ and $T_j = \max(C_j - d_j, 0)$ its tardiness. The objective is to find a schedule that minimizes the total tardiness $T = \sum_{j=1}^n T_j$ in such a way that two jobs cannot be processed at the same time on the same machine. The splitting and the pre-emption of the operations are forbidden.

Table 1 shows an instance of the $HJS_n \mid \text{prec}, S_{mj}, r_j, d_j \mid \sum T_j$ problem with four jobs and four machines $4^*4$.

<table>
<thead>
<tr>
<th>Job</th>
<th>$r_j$</th>
<th>$d_j$</th>
<th>Total processing time</th>
<th>Sequence</th>
<th>Processing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>35</td>
<td>27</td>
<td>2-1-4-3</td>
<td>$p_{1,1} = 4, p_{1,2} = 8, p_{1,3} = 10, p_{1,4} = 5$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>22</td>
<td>11</td>
<td>1-2-4</td>
<td>$p_{2,1} = 2, p_{2,2} = 6, p_{2,3} = 3$</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>25</td>
<td>20</td>
<td>1-2-4-3</td>
<td>$p_{3,1} = 7, p_{3,2} = 5, p_{3,3} = 1, p_{3,4} = 7$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>34</td>
<td>14</td>
<td>4-3-1</td>
<td>$p_{4,1} = 7, p_{4,2} = 4, p_{4,3} = 3$</td>
</tr>
</tbody>
</table>

Table 1. Example $4^*4$

The $HJS_n \mid \text{prec}, S_{mj}, r_j, d_j \mid \sum T_j$ problem, extends the classical job shop problem by the presence of identical parallel machines, by allowing for sequence dependent setup times between adjacent operations on any machine and the restriction of jobs arrival dates.
The classical job-shop problem, $J_m || \gamma$, is a well-known NP-hard combinatorial optimization one, see Garey and Johnson, (1979), which makes our problem a NP-hard problem too. The $J_m || \gamma$ problem has been investigated by several researchers. It can be classified in two large families according to the objective function: minimizing makespan and minimizing tardiness.

3. State of the Art

3.1 Parallel Machine

The Hybrid Job Shop is linked in some way to Parallel Machine Job Shop. Indeed as can seen in the Figure 1, an Hybrid Job Shop is composed of different stages which can contain one single machine or parallel machines.

![Figure 1. Example of Hybrid Job Shop](image)

So this type of problem can be described as a sequence of parallel machine problem. Moreover the Parallel Job Shop problem has been widely studied especially for the minimization of the total tardiness. The parallel machine problem consists of scheduling $N$ jobs on $M$ different parallel machines without interruption. The goal here is to minimize the total tardiness. The parallel machine problem is known as NP-Hard, so the minimization of total tardiness in a parallel machine problem is also NP-Hard according to Kouilamas C., (1994) and Yalaouli & Chu, (2002). Different reviews exist in the literature as Kouilamas C., (1994) and Shim & Kim, (2007) and it appears that the one machine problem has been more studied than the multiple machine problems. On the other hand, one can also stress that the objective is mainly to minimize the makespan, the total flow time and more recently the minimization of the total tardiness. We now mention different interesting works for their heuristics or their problem. In 1969 Pritsker et al., (1969) have done the formulation with linear programming. Alidaee & Rosa, (1997) have proposed a method based on the modified due date method of Baker K.R. & Bertrand J.W., (1982). Other priority rules can be found in the work of Chen et al., (1997). Kouilamas has proposed the KPM to extend the PSK method of Panwalker et al., (1993) to parallel machines problem, the former has proposed also a method based on Potts & Van Wassenhove, (1997) on the single machine problem and also an hybrid method with Simulated annealing Kouilamas, (1997). Other authors were interested in this type of
problem, as Ho & Chang, (1991) with their traffic priority index or Dogramaci & Surkis, (1979) with different rules like Early Due Dates, Shortest Processing Time or Minislack. There is the work of Wilkerson & Irwin, (1979) and finally one must mention the Montagne’s Ratio Method (Montagne, 1969).


More recently Armentano and de França Filho, (2007) have proposed a tabu-search with a self adaptive memory, Lallaoui et al., (2007) proposed six efficient approach in order to take the best schedule, one can also mention the work of Mönch and Unbehaun, (2007) who compare their results to the best known lower bound. Anghinolfi and Paolucci, (2007) have proposed an algorithm based on tabu search, simulated annealing and variable neighbourhood search.


3.2 Parallel machine: useful results

One can mention different results which can be useful for a Hybrid Job Shop. Now, we propose a selection of properties and especially dominance ones from different authors.

Assuming the following notations:

- $J$: set of jobs
- $M$: set of machines
- $n$: number of the jobs ($n = |J|$)
- $m$: number of the machines ($m = |M|$)
- $p_i$: processing time of job $i$
- $d_i$: due date of job $i$
- $C_i(\sigma)$: completion time of job $i$ in partial schedule $\sigma$
- $T_i(\sigma)$: tardiness of job $i$ in partial schedule $\sigma$
- $I(\sigma)$: completion time of the last job on machine $k$ in (partial) schedule $\sigma$
- $n_k(\sigma)$: number of jobs assigned to the same machine, in partial schedule $\sigma$
- $S(\bullet)$: set of jobs already included in partial schedule $\bullet$

We will now enumerate the selection of dominance properties:

Proposition 1 (Azizoglu & Kirca, (1998)): There exists an optimal schedule in which the number of jobs assigned to each machine does not exceed $N$ such that:

$$\sum_{i=1}^{n} p_{j[i]} \leq A \quad \text{and} \quad \sum_{i=1}^{N+1} p_{j[i]} \geq A \quad (1)$$

Where $p_{j[i]}$ is the processing time of the job with the $i$th shortest processing time.

Proposition 2 (Azizoglu & Kirca, (1998)): If $d_i \leq p_i$, for all jobs an SPT schedule is optimal.

Proposition 3 (Azizoglu & Kirca, (1998)): For any partial schedule $\sigma$, if $d_i \leq p_i + \min_{m \in M} \Gamma_i(\sigma)$ for all $i$ not in $S(\sigma)$, then it is better to schedule jobs after $\sigma$ in an SPT order.
Proposition 4 (Yalaoui & Chu, (2002)): For a partial schedule $\sigma$ and any job $i$ that is not included in $\sigma$, if there is another job $j$ not included in $\sigma$ that satisfies $p_j \leq p_i$ and $(p_j - p_i) \max k \in M \Gamma_k(\sigma) - 1 \leq \min \{d_i - p_i, \min k \in M \Gamma_k(\sigma)\}$, where $\hat{n}(\sigma)$ denote the number of additional jobs that are scheduled on machine $k$ after partial schedule $\sigma$ in an optimal schedule, then complete schedule $\sigma^*$ are dominated.

Proposition 5 (Yalaoui & Chu, (2002)): For a partial schedule $\sigma$ and any job $i$ that is not included in $\sigma$, if there is another job $j$ not included in $\sigma$ that satisfies $0 \leq p_j - p_i \leq \min r \in S(\sigma) (C_r(\sigma) - p_j)$ and $(p_j - p_i)(\hat{n}(\sigma) - 1) \leq \min \{d_i - p_i, \min k \in M \Gamma_k(\sigma)\} - \min k \in M \Gamma_k(\sigma)$, then complete schedule $\sigma^*$ are dominated.

Proposition 6 (Shim & Kim, (2007)): For any schedule $\sigma$ in which job $i$ and job $j$ are assigned to the same machine and job $j$ precedes job $i$, there is a schedule that dominates $\sigma$, if at least one of the following three conditions holds:

1. $p_j \leq p_i$ and $d_i \leq \max (C_i(\sigma), d_i)$
2. $d_i \leq d_j$ and $C_i(\sigma) - p_i \leq d_i \leq C_i(\sigma)$
3. $C_i(\sigma) \leq d_j$

### 3.3 Hybrid Job shop

Much of the research literature in job shop scheduling deals with pure job shop environments. However, currently most processes involve a hybrid of both the job shop and a flow shop with a combination of flexible and conventional machine tools.

In a classical job shop problem, the elementary product operations follow a completely ordered sequence according to the product to be manufactured. In some structures, each elementary operation may be carried out on several machines, from where, thanks to the versatility of the machines, a greater flexibility is obtained. We can talk about total flexibility if all the machines are able to carry out all the operations, otherwise, it is a partial flexibility. This is what we call the hybrid job shop or the flexible job shop.

This flexibility may also be applied to the flow shop problem leading then to the hybrid flow shop configuration. A hybrid flow shop is constituted of several stages or groups. Each stage is composed by a set of machines. The passing order in the stages for each part to be manufactured is the same one as in Gourgand et al., (2001). In this work, we are particularly interested in the hybrid job shop scheduling problem.

The Hybrid Job Shop Problem (HJSP) is then an important extension of the classical job shop scheduling problem which allows an operation to be processed by any machine from a given set thus creating an additional complexity. The methodology is to assign each operation to a machine and to order the operations on the machines, such that the maximal completion time (makespan) of all operations or the total tardiness is minimized.

Many scheduling optimization problems have been studied in the research works dealing with complex industrial cases with flexibility. The hybrid job shop scheduling problem was one of those studies presented in the literature like Penz, (1996), Dauzere-Peres et al., (1998), Xia and Wu (2005) and many others.

Chen et al., (1999) present a genetic algorithm to solve the flexible job-shop scheduling problem with a makespan criterion to be minimized. The chromosomes representing the problem solutions consist of two parts. The first part defines the routing policy and the second part the sequence of the operations on each machine. Genetic operators are introduced and used in the reproduction process of the algorithm. Numerical experiments show that the algorithm can find out high-quality schedules.
Gomes et al., (2005) present an integer linear programming (ILP) model to schedule flexible job shop. The model considers groups of parallel homogeneous machines, limited intermediate buffers and negligible set-up effects. Orders consist of a number of discrete units to be produced and follow one of a given number of processing routes with a possibility of re-circulation. Good solution times are obtained using commercial mixed-integer linear programming (MILP) software to solve realistic examples of flexible job shops to optimality.

A genetic algorithm-based approach is also developed to solve the considered problem by Chan et al., (2006). The authors try to solve iteratively a resource-constrained operations-machines assignment problem and flexible job-shop scheduling problem. In this connection, the flexibility embedded in the flexible shop floor, which is important to today's manufacturers, is quantified under different levels of resource availability.

Literature review shows that minimizing tardiness in hybrid job shop problems has been an essential criterion. It is the main objective of the work of Scrich et al., (2004). Two heuristics based on Tabu Search are developed in this paper: a hierarchical procedure and a multiple start procedure. The procedures use dispatching rules to obtain an initial solution and then search for improved solutions in neighborhoods generated by the critical paths of the jobs in a disjunctive graph representation. Diversification strategies are also implemented and tested.

Alvarez-Valdez et al., (2005) presented the design and implementation of a scheduling system in a glass factory aiming at minimizing tardiness by means of a heuristic algorithm. The structure basically corresponds to a flexible job-shop scheduling problem with some special characteristics. On the one hand, dealing with hot liquid glass imposes no-wait constraints on some operations. On the other hand, skilled workers performing some manual tasks are modeled as special machines. The system produces approximate solutions in very short computing times.

Minimizing tardiness in a hybrid job shop is one of the objectives in the work of Loukil et al., (2007) that the authors tried to optimize. A simulated annealing is developed and many constraints are taken in consideration such as batch production; existence of two steps: production of several sub-products followed by the assembly of the final product; possible overlaps for the processing periods of two successive operations of a same job. At the end of the production step, different objectives are considered simultaneously: the makespan, the mean completion time, the maximal tardiness and the mean tardiness.

For our case study, two works have discussed the problem of minimizing tardiness in a hybrid job shop. The first was that of Nait Tahar et al., (2004) by using a genetic algorithm. Only one criterion was taken into account which was the total tardiness. The results obtained showed that the genetic algorithm technique is effective for the resolution of this specific problem. Later, an ant colony optimization algorithm was developed by Nait Tahar et al., (2005) in order to minimize the same criterion with sequence dependent setup times and release dates.

4. Case study: industrial

In this section we will describe an industrial case of hybrid job-shop. Firstly we will describe the problem encountered by a company, and then we will develop three ways of solving the problem: one with a genetic algorithm, the second with a meta heuristic based on ant colony system and the third one with a non-dominated sorting genetic algorithm coupled with a simulation software.
The problem is located in the printing factory that could be described in Figure 2. This factory produces printed and unprinted roll from the raw material: paper, plastic film and ink are combined to produce a finish product. The plant employs 90 people to produce high quality packaging. It produces about 1500 different types of finished goods and delivers about 80 orders per week. During the process, each product (job) is elaborated on a given sequence of machines. The tasks performed on these machines are called operation.

As it appears, the factory structure shows an hybrid job shop structure, with some single machine stage (M5, M6) and multiple machines (M7,M8,M9, for instance) stage with identical parallel machines. Setup times are present: when a machine switch from one operation to another a “switching time” is required. The process is divided into four areas: printing, assembly, paraffining, winding and cutting. The process starts in the printing area where a drawing in one or more colours is reproduced on a paper, raw material. Two printing process can be used: photoengraving and flexography. The assembly combines two supports (printed or not) with a binder (adhesive) on their surface forming one. Paraffining put paraffin on the surface. Then the products reach the cutting and winding area. Finally the products are packaged, stored or shipped.

We now describe two methods used to solve this problem using Ant Colony System (ACS) based algorithm and Genetic Algorithm (GA).

Then a third method is presented dealing with a multi-objectif case resolution.

4.1 Genetic Algorithm

The first method of Nait Tahar et al., (2004) uses a genetic algorithm to solve the problem. In a genetic algorithm the solution is represented in a chromosome. The first step is the modeling of the solution in a genetic way, each encoding is specific to one problem. We employ the following encoding with a matrix as it is shown in table 2.
This encoding represents the scheduling in a table of $m$ lines. Each line represents the operations to schedule in the form of $n$ cells ($n$ is the number of jobs). Each cell contains: the job number, the order of the operation in the manufacturing process, the processing time and the operation completion date. The representation of a solution considers the sequences for each job, a machine sequences and not a group sequences. The evaluation (fitness) of an individual is simply the total tardiness.

Since the encoding is chosen, we have to propose mutation and crossover operators. Three crossovers are known for the problems of sequencing: LOX (Linear Order Crossover), OX (Order Crossover) and X1 (One point crossover). We adopt X1 crossover with the studied problem for our encoding. For a parent $P_1$ having a length $t$, a random position $p$ ($p < t$) is generated. To build the child $E_1$, the portion $P_1$ between 1 and $p$ inclusive is copied in $E_1$ using the same positions. Then the portion of $P_2$ between $p$ which is not included and $t$ is swept. Only the non present elements in $E_1$ are copied. The missing elements in $E_1$ are added after, from left to right. The construction of child $E_1$ is identical, by permuting the role of $P_1$ and $P_2$. A chromosome contains all the operations of the problem, and each operation is assigned to only one machine. To prevent a too fast convergence of the algorithm, a mutation is applied to the children with a weak rate. We tested two types of mutation named mut-ch and mut-nb. The first interchange two operations randomly selected from the busiest machine in the chromosome. mut-nb interchanges two operations from the machine having the most total tardiness.

The population stores a fixed number ($T_{pop}$) of chromosomes in a table. These initial solutions are created randomly. For each machine belonging to a single machine group, a sequence is thus randomly generated. For the groups containing several units (identical parallel machines), the operation assignment and sequencing on each machine are also randomly done by balancing the work-load of the machine. For the selection we tested roulette technique and direct tournament. The genetic algorithm is an incremental (steady state) one: the new solutions immediately replace existing solutions in the population. Five procedure have been tested for the replacement: each parent is replaced by its children if there is improvement, the worst parent is replaced by the best descendant, the worst individual of the population is replaced by the better of the solutions, a randomly selected parent is replaced by a randomly selected child, the child replaces an individual chosen uniformly chosen under the median (incremental replacement). Our algorithm is tested with production data, coming from the network of the printing factory. These data are adapted to our algorithm, by creating instances of the same size as the randomly generated (25, 50, 100 jobs). We used a probability of 90% for the crossover and 10% for the mutation probability.

Table 3 gives results form many instances, different columns show name of the instance, its size and its number of operations, moreover we can see the total tardiness of the industrial solution and the one with the solution given by the algorithm. Finally this table shows the improvement between industrial and genetic based solution.
Hybrid Job Shop and parallel machine scheduling problems: 
minimization of total tardiness criterion

<table>
<thead>
<tr>
<th>Instance</th>
<th>Size</th>
<th>Operations</th>
<th>Real total tardiness</th>
<th>GA total tardiness</th>
<th>Improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>25</td>
<td>133</td>
<td>1297.9</td>
<td>919.95</td>
<td>29.12</td>
</tr>
<tr>
<td>A2</td>
<td>25</td>
<td>145</td>
<td>1330.1</td>
<td>915.91</td>
<td>31.14</td>
</tr>
<tr>
<td>A3</td>
<td>25</td>
<td>161</td>
<td>1488.4</td>
<td>1106.33</td>
<td>25.67</td>
</tr>
<tr>
<td>A4</td>
<td>25</td>
<td>155</td>
<td>1451.6</td>
<td>908.12</td>
<td>37.44</td>
</tr>
<tr>
<td>A5</td>
<td>25</td>
<td>114</td>
<td>1237.4</td>
<td>809.75</td>
<td>34.56</td>
</tr>
<tr>
<td>B1</td>
<td>50</td>
<td>339</td>
<td>2311.2</td>
<td>1268.62</td>
<td>45.11</td>
</tr>
<tr>
<td>B2</td>
<td>50</td>
<td>297</td>
<td>2122.4</td>
<td>1229.72</td>
<td>42.06</td>
</tr>
<tr>
<td>B3</td>
<td>50</td>
<td>308</td>
<td>2178.1</td>
<td>1432.75</td>
<td>34.22</td>
</tr>
<tr>
<td>B4</td>
<td>50</td>
<td>325</td>
<td>2219.3</td>
<td>1350.67</td>
<td>39.14</td>
</tr>
<tr>
<td>B5</td>
<td>50</td>
<td>341</td>
<td>2297.4</td>
<td>1335.02</td>
<td>41.89</td>
</tr>
<tr>
<td>C1</td>
<td>100</td>
<td>637</td>
<td>4587.0</td>
<td>2366.89</td>
<td>48.40</td>
</tr>
<tr>
<td>C2</td>
<td>100</td>
<td>524</td>
<td>4233.3</td>
<td>2279.63</td>
<td>46.15</td>
</tr>
<tr>
<td>C3</td>
<td>100</td>
<td>612</td>
<td>4524.9</td>
<td>2308.15</td>
<td>48.99</td>
</tr>
<tr>
<td>C4</td>
<td>100</td>
<td>688</td>
<td>4721.2</td>
<td>2307.72</td>
<td>51.12</td>
</tr>
<tr>
<td>C5</td>
<td>100</td>
<td>653</td>
<td>4642.8</td>
<td>2512.68</td>
<td>45.88</td>
</tr>
</tbody>
</table>

Table 3. Improvement of the industrial solution, Nait Tahar et al., (2004)

The Genetic Algorithm has been coded in C on a 440 Mhz bi processors, it took near from 1000 seconds to get solutions. One can see that the improvement is important from 29 to 51% from the industrial solution use by the factory and based on the “Early Due Date” policy. We have improve significantly the industrial solution. We will see now how this solution can be improved with another meta-heuristic called Ant Colony System optimization.

4.2 Ant Colony System

The ACS Nait Tahar et al.,(2005) attempts to solve the problem imitating the behaviour of ants searching food in the nature. Consider for instance the four-job example of table 1.

![Disjunctive graph for 4x4 instance](image)

Figure 3. A disjunctive graph for 4x4 instance, Nait Tahar et al., (2005)
We have to describe the sequence by a graph (see Figure 3) in order to apply an ant colony based algorithm. The former is a disjunctive graph where node are operations of the job on a certain machine, and the conjunctive arc are weighted by the duration of operations and the arc connecting the node $U$ correspond to the release date $r_j$. Thanks to this description we can apply an ant colony based algorithm sketched by the algorithm 4:

Algorithm 4. A skeleton of the Ant Colony System algorithm (Nait Tahar (2005))

A solution is a path From $U$ to $V$. This path is build by an ant step by step, node by node. The principle is to simulate ants walking trough the graph, at each node they have to choose one arc. The criterion for this choice is the probability of each arc to be taken: this probability grows with the number of ants which have traversed this arc. This mechanism is assumed by the pheromone lay down by each ant.

In this algorithm two things have to be precised: $\tau_{a,b}$ the quantity of pheromone on the arc $(a,b)$, how the next node $b$ is chosen, and finally how the pheromone quantity is updated on each arc.

Here is described how the pheromone quantities are determined:

$$\tau_{a,b} = \sum_{k=1}^{n} \Delta \tau_{a,b}^k$$

(2)

$$\Delta \tau_{a,b}^k = \begin{cases} \frac{Q}{W_k} & \text{if arcs } (a \to b) \text{ belongs to the path of ant } k \\ 0 & \text{otherwise} \end{cases}$$

(3)
With $W_k$ is the total tardiness of the arc selected by the ant $k$, $Q$ is a constant, $n$ is the number of job generated and $\tau_{0,0}$ is the quantity of pheromone at initial time. Now an ant can “walk” through the graph (i.e. it can build a path), partially guided by the pheromone:

$$b = \left\{ \begin{array}{ll}
\arg \max_{b \in \text{Succ}(a)} \frac{r_{a,b}}{r_{a,b} + \eta_{a,b}} & \text{if } \phi \leq \Phi_0 \\
p_k & \text{otherwise}
\end{array} \right. \quad (4)$$

$$p_k^{a,b} = \left\{ \begin{array}{ll}
\frac{\left[ \frac{r_{a,b}(t)}{\eta_{a,b}} \right]^\beta}{\sum_j \frac{r_{j}(t)}{\eta_{j}}^{\beta}} & \text{if } b \in \text{Succ}(a) \\
0 & \text{otherwise}
\end{array} \right. \quad (5)$$

Where $\eta_{a,b}$ is an estimate of desirability of the transition $a,b$ according to the apparent tardiness cost (ATCS) heuristic, Lee et al., (1997), $\text{Succ}(a)$ is the set of adjacent nodes to $a$, $\Phi$ is a random number in $[0,1]$, and $\Phi_0$ is a tuning parameter. Here we simulate the route of an ant $k$ through the graph by a two-level decision making, Dorigo & Gambardella, (1997). At first there is a draw of $\Phi$: if $\Phi$ is lower than $\Phi_0$ then the next node visited by $k$ will have the maximum value of $\left[ \frac{r_{a,b}}{\eta_{a,b}} \right]^\beta$, otherwise the probability will be determined with the equation number (5).

And after making a choice for an arc, we have to update the pheromone according to the algorithm. We do this with the formula:

$$\tau_{a,b} = (1-\theta)\tau_{a,b} + \theta \tau_0 \quad (6)$$

Where $\theta$ $(0<\theta<1)$ is the local pheromone decay parameter and $\tau_0$ is the initial amount of pheromone deposited on each arc. In our case we consider $\tau_0 = (\Sigma_{j=1}^{n} T_j)^{1/\text{EDD}}$ where $(\Sigma_{j=1}^{n} T_j)_{\text{EDD}}$ is the total tardiness given by the Early Due Date. Once all of the ants have completed their path, the intensity of pheromone on each arc is update according to below:

$$\tau_{a,b} = (1-\lambda)\tau_{a,b} + \lambda \sum \Delta \tau_{a,b}^{k} \quad (7)$$

Where $\Delta \tau_{a,b}$, calculated by equation (2) is the pheromone currently laid by ant $k$, and $\lambda$ is the evaporation rate of previous pheromone intensity $(0<\lambda<1)$. Finally the authors compare this algorithm to Genetic Algorithm. In order to compare them to each other, the authors have tested these algorithms on 900 different instances, and they compare the computation time took by ACS and GA. This is possible with the use of $\text{Cycle}^*$ determined by $\text{Cycle}^* = \text{PW1} \times \text{Cycle}_{\text{max}}$ where $\text{Cycle}_{\text{max}}$ is the stopping criterion of the algorithm and PW1 is coefficient showing the weight of $\text{Cycle}_{\text{max}}$ between the two methods, here we choose $\text{Cycle}_{\text{max}} = 3000$ for the ACS and $\text{Cycle}_{\text{max}} = 1000$ for the GA, these values represent the same amount of CPU time.

Finally we obtain better results with ACS than Genetic Algorithm. According to the Figure 5 it can be seen that this the Ant Colony System based algorithm improve its result at each iteration rather than the Genetic Algorithm does.
To conclude, we have introduced an interesting ant colony system for hybrid job shop scheduling problem with sequence dependent setup times and release dates to minimize the total tardiness, encountered in industrial situation. The ant colony method proved to be very efficient for randomly generated and real instances compared to a genetic algorithm.

4.3 Non Dominated Sorting Genetic Algorithm
This part presents an optimization technique built by coupling the ARENA®, Kelton et al., (2003) simulation software with a multi-objective optimizer based on the second version of a nondominated sorting genetic algorithm (NSGA-II) coded in Visual Basic for Application (VBA). This simulation-based-optimization technique is used to optimize the performances of the simulation model representing the considered workshop (the same study case adopted for the ACS and the GA) by testing new scheduling rules different from the only First In First Out (FIFO) rule which was adopted for the machines.

This work was developed in order to assess, by the means of a simulation software, the production system and to have a comprehensive tool in which the whole system’s constraints will be handled as well as those of the logistical and handling system. These additional constraints have required a powerful simulation tool to manage them. In addition to that, the stochastic nature of some system’s parameters (like the downtime of machines, the arrival times of products or others) makes analytical models very complicated or computationally intractable. That is why we have decided to use the simulation based optimization technique as it has been proved to be effective for such kind of applications.

Indeed, simulation is more and more used in today’s industries with the aim of assessing their systems or to study the impact of changing system design parameters, Muhl et al., (2003) and Sahlin et al., (2004). ARENA®, developed by Systems Modelling Corporation, is
one of the softwares that can be used to model industrial systems in different domains like automobile, aeronautics as well as others like hospitals, banks, ...Kelton et al., (2003).
While simulation makes it possible to test potential changes in an existing system without disturbing it or to evaluate the design of a new system without building it, simulation based optimization can be defined by coupling an optimization method with simulation in order to test many parameters that can maximize the performances of the simulated system, Hani et al., (2006).
The coupling of heuristic methods with the ARENA® software, or other simulation software, was the subject of many works like Harmonosky (1995), Drake & Smith (1996) or others.
The second version of the non-dominated sorting genetic algorithm (NSGA-II) is a heuristic algorithm based on the genetic techniques applied by Goldberg (1989) and it was initially implemented by Deb et al., (2000). It is based on the principle of the genetic algorithms by means of creating an initial population, selecting parents in order to get children and finally choose the best solution constructed from genes.
In addition to that, it consists on affecting fronts or groups to the proposed solutions. Front 1 contains the non-dominated solutions of the created population. Those individuals or solutions are then virtually removed from the population. We compare the remaining solutions and the next set of non-dominated solutions is assigned to front 2 and so on until that each individual of the population is affected to a front. Many works and researches have as a main subject the impact of NSGA algorithms on different optimization problems such as in Dolgui et al. (2005) for balancing and optimizing production lines or in Deb and Reddy (2003) and Deb et al. (2004).
Our model allows to simulate the production system with graphical animations starting from the exit of raw materials form the warehouse and until the exit of finished products while modeling their circulations which will be really done by the means of Auto Guided Vehicles (AGV). All the machines were modelled. In addition to the machines and handling systems' characteristics, the model contains as inputs:
- the workshop structure (production areas, warehouse and stock zones ...) on the scale,
- the different job sequences which guide the products forward between the servers,
- the simulation horizon (one year),
- the statistical law representing time between product arrivals.
In order to validate the simulation and to evaluate the production system, performances indicators were introduced in the model and they were compared to the real indicators adopted in the workshop. These indicators to optimize are:
1. The performance rate of each machine ($g_{1}(k)$) (to be maximized)
2. The occupation rate of each machine ($g_{2}(k)$) (to be maximized)
3. The total tardiness time resulting from processing all the jobs on a considered horizon of time $T$ (to be minimized)
The multi objective optimization consists of finding the optimal objective function vector, $g(k) = [g_{1}(k), g_{2}(k), T]$, instead of a unique objective function. It aims at finding a compromise between the set of objectives.
We try to optimize those objectives by choosing the best priority for each queue of the considered machines. The multiplicity of choices could lead to results which are better in the case of using a unique rule as it was shown in the paper of Liu and Wu (2004). Until the beginning of this work, only one rule was tested: FIFO (First In First Out).
In our work, four priority rules were adopted to be tested as a first step: FIFO, LIFO (Last In First Out), SPT (Shortest Processing Time) and LPT (Longest Processing Time). The results obtained by the optimization algorithm will help us to make the final choice. We present in this part the different properties of the developed algorithm. A chromosome specifies the scheduling rule for each machine. The number of genes in the chromosomes of our algorithm is equal to the number of machines. The first step of our algorithm is then to create an initial population. The value of each gene $k_i$ is generated randomly based on a uniform distribution $U[K_{\text{imin}}, K_{\text{imax}}]$. $K_{\text{imin}}$ and $K_{\text{imax}}$ are respectively the minimum and the maximum possible values of $K_i$. As we have four policies then $K_{\text{imin}} = 1$ and $K_{\text{imax}} = 4$. The scheduling rules and the corresponding numbers are shown in table 4.

<table>
<thead>
<tr>
<th>Priority rule</th>
<th>Corresponding number</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIFO</td>
<td>1</td>
</tr>
<tr>
<td>LIFO</td>
<td>2</td>
</tr>
<tr>
<td>SPT</td>
<td>3</td>
</tr>
<tr>
<td>LPT</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4. Priority rules

The binary tournament technique is used to select the parents: two solutions are randomly selected and the best one becomes the first parent. This process is repeated to get the second parent. We choose the two-point crossover operation with a high probability and a very small point mutation probability.

The steps of the algorithm, as shown in Fig. 6, were inspired from the work of Deb et al. (2000). The overall structure of the NSGA-II algorithm is presented in Fig. 7. For more details about the algorithm, reader is referred to Chehade et al. (2007).

![Figure 6. Steps of the NSGA-II algorithm (Deb et al., 2000)](image)

The simulation model and the optimization algorithm interact by means of a VBA procedure as the ARENA® software has a Visual Basic Editor. The coupling process works in the following way:

1. The algorithm starts by executing the first steps of the NSGA-II, to generate an initial population ($M_t$) of size $ns$
2. In order to calculate the fitness functions for the individuals of \((M_t)\), the simulation model is launched on \(ns\) iterations. Each iteration is supposed to calculate one fitness function corresponding to one individual (chromosome). The queues’ priority rules of each machine are read directly from the chromosomes of the algorithm.

3. Rank solutions in \((M_t)\).

4. The algorithm creates then the offspring population \((N_t)\) (of size \(ns\)) by processing genetic functions (selection of parents, crossover, mutation).

5. The simulation model does \(ns\) new iterations in order to evaluate the chromosomes of the \((N_t)\) population. At this stage, we have now the population \((O_t)\) of size \(2ns\).

6. Rank solutions in \((O_t)\).

7. The algorithm executes its remaining steps (decomposition into fronts, crowding distance sorting). We have now the new parent population \((M_{t+1})\).

8. Repeat steps 4 to 7 till the stopping criteria is reached (which is in our case the number of generations).

```
Create the initial population \(M_t\) of size \(ns\)
Evaluate the \(ns\) solutions using simulation
Sort \(M_t\) by non domination
Compute the crowding distance of each solution
REPEAT
Creation of the offspring population \(N_t\): add \(n\) children at the end of \(M_t\) (with genetic operators: selection, crossover and mutation of two parents) and evaluate each solution by simulation
Sort \(N_t\) by non domination
Compute the crowding distance of each solution
Sort the resulting population \(O_t\) of \(2^{ns}\) solutions by non-domination
\(M_{t+1} = 0;\)
\(i = 1\)
WHILE \(|M_{t+1}| + |\text{front}(i)| \leq n\) do
   Add \(\text{front}(i)\) to \(M_{t+1}\)
   \(i = i + 1\)
END WHILE
missing = \(n - |M_{t+1}|\)
IF missing \(
\neq 0\) THEN
   Sort the solutions by descending order of the crowding distance
   FOR \(j = 1\) to missing DO
      Add the \(j^{th}\) solution of \(\text{front}(i)\) to \(M_{t+1}\)
   END FOR
   \(P = M_{t+1}\)
END IF
UNTIL Stopping Criterion
```

Algorithm 7. Overall Structure of the NSGA-II algorithm (Deb et al., 2000)

Table 5 shows a comparison between the real industrial data (RID) and the first results of simulation (SIM) initially get without applying the NSGA-II algorithm. It shows that a very
small gap for the three performances indicators was noticed which is a very good basis to realize later the optimization procedures. Final results obtained after 100 generations showed that individuals of the last population are distributed on three fronts. Table 6 shows the optimization results which are compared to previous simulation results (SIM). The three objectives presented in the table are the performance rate of the machines (PR), the occupation rate of the machines (OR) and the total tardiness time (TT). It shows first the average result for each objective (OABF), the best (BIBF) and worst individual for each objective (WIBF), which gives an idea about the distribution of the individuals in this front. The numbers in brackets for OABF, BIBF and WIBF represent the difference between those parameters and the simulation results (SIM). The last row shows the standard deviation (STD) of those non-dominated individuals. Table 6 shows results where each simulation iteration is set to cover a production horizon of ten years. We adopted a warm-up period of two years. The size of the initial population is 20, the number of generations is 100, the crossover probability is 0.9 and the mutation probability is 0.01. The average of the numerical results of the best front shows that the performance rate and occupation rate are improved by 6.28% and 12.7% respectively. As for the third indicator which is total tardiness, it is reduced by 48.3% on average. As a consequence, the algorithm has showed that it has considerable improvements on the performances of the model.

Table 5. Simulation results compared to real industrial data

<table>
<thead>
<tr>
<th>RID</th>
<th>SIM</th>
<th>GAP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance rate (%)</td>
<td>56</td>
<td>57.2</td>
</tr>
<tr>
<td>Occupation rate (%)</td>
<td>75</td>
<td>74.4</td>
</tr>
<tr>
<td>Total tardiness</td>
<td>53.6</td>
<td>53.1</td>
</tr>
</tbody>
</table>

Table 6. Optimization results

<table>
<thead>
<tr>
<th></th>
<th>PR(%)</th>
<th>OR(%)</th>
<th>TT(hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM</td>
<td>57.2</td>
<td>74.4</td>
<td>53.1</td>
</tr>
<tr>
<td>OABF</td>
<td>63.48 (+6.28%)</td>
<td>87.1 (+12.7%)</td>
<td>27.6 (-48.3%)</td>
</tr>
<tr>
<td>BIBF</td>
<td>68.9 (+11.7%)</td>
<td>94.8 (+20.4%)</td>
<td>19.9 (-62.53%)</td>
</tr>
<tr>
<td>WIBF</td>
<td>57.8 (+0.6%)</td>
<td>76.9 (+2.5%)</td>
<td>36.2 (-31.8%)</td>
</tr>
<tr>
<td>STD</td>
<td>4.63</td>
<td>6.96</td>
<td>6.15</td>
</tr>
</tbody>
</table>

5. Conclusion

In this chapter we have presented different results useful for scheduling tasks through a hybrid job shop system. At first we have dealt with the parallel machine job shop since its structure is near from the multi-processors stages of a Hybrid Job Shop. Then we have presented some theoretical results and their application in the industry. We have developed some examples of modeling industrial lines for a genetic application or an ant colony system application. After this step of modeling, the result show real improvements of the minimization of the total tardiness in an industrial case. These results could be very useful in the semiconductor manufacturing or in the paper industries since the Hybrid Job Shop structure seem to be common in this area.
Hybrid Job Shop and parallel machine scheduling problems:
minimization of total tardiness criterion

6. References


Hybrid Job Shop and parallel machine scheduling problems: minimization of total tardiness criterion


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:
