A Job Shop Scheduling Problem with Due Dates Under Conditions of Uncertainty

Wojciech Bożejko[1(✉)], Paweł Rajba[2], Mariusz Uchroński[1], and Mieczysław Wodecki[3]

[1] Department of Automatics, Mechatronics and Control Systems, Wrocław University of Technology, Janiszewskiego 11-17, 50-372 Wrocław, Poland
{wojciech.bozejko,mariusz.uchronski}@pwr.edu.pl
[2] Institute of Computer Science, University of Wrocław, Joliot-Curie 15, 50-383 Wrocław, Poland
pawel@cs.uni.wroc.pl
[3] Department of Telecommunications and Teleinformatics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
mieczyslaw.wodecki@pwr.edu.pl

Abstract. In the work we consider a job shop problem with due dates under conditions of uncertainty. Uncertainty is considered for operation execution times and job completion dates. It is modeled by normal and Erlang random variables. We present algorithms whose constructions are based on the tabu search method. Due to the application of the probabilistic model, it was possible to obtain solutions more resistant to data disturbances than in the classical approach.

Keywords: Scheduling · Job shop · Weighted tardiness · Uncertainty

1 Introduction

The paper deals with the job-shop problem, widely regarded as one of the most difficult combinatorial problems. Due to many of its practical applications, it has been intensively researched for many years. There are few works that dealt with the deterministic job shop problem with additional parameters and constraints. One of the few is the work of Balas et al. [1], which deals with the problem of the job shop with time windows, penalties for untimely completion of jobs and machine setups. A review of different variants of multi-machine problems with machine setups is contained in Sharma and Janin [8]. There are definitely fewer works devoted to the job shop problem with uncertain parameters. Kim et al. [5] presented a modification of classical construction algorithms solving deterministic scheduling problems. Shoval and Efatmaneshnik [9] consider the problem in which jobs execution times are random variables with a normal distribution. The method of determining the set of feasible schedules is presented, as well as the criteria for selecting an appropriate sub-optimal schedule from this set.

M. Paszynski et al. (Eds.): ICCS 2021, LNCS 12742, pp. 198–205, 2021.
https://doi.org/10.1007/978-3-030-77961-0_17

There are several practical reasons for introducing due dates. First, there might be dependencies to other departments or contracted delivery dates as a part of the over supply chain processes. The other reason is that there might be time window when machines are not available e.g., because of some maintenance activities. In this paper we consider a job shop problem with due dates for jobs completion and penalties for tardy jobs, applying total weighted tardiness criterion. Uncertain parameters, i.e., processing times and due dates are random variables with the normal or Erlang distributions. Normal distribution is the natural choice when we expect a standard behavior of some phenomenon, obviously we can also leverage very useful mathematical properties in the modeling transformations. Erlang distribution was investigated as, for set of parameters, it introduces a possibility to model events which diverge from the standard behavior. This distribution also offers some useful mathematical properties which have been applied in the modeling activities.

2 Problem Formulation

The literature presents various ways of defining the job shop problem, nevertheless in this paper we will use the definitions and symbols from [4]. Let $\mathcal{J} = \{1, 2, \ldots, n\}$ be a set of jobs, $\mathcal{M} = \{1, 2, \ldots, m\}$ - a set of machines. By $\mathcal{O} = \{1, 2, \ldots, o\}$ we denote a set of operations, where an operation is equal to an action of job execution on a machine in a *machine order*. Therefore, we can partition the set of $\mathcal{O}$ operations into sequences corresponding to individual jobs, i.e., define the j job as sequence of o_j operation, which will be indexed sequentially with the numbers $(l_{j-1} + 1, \ldots, l_j)$. The operations should be performed in a given order (*technological order*), where $l_j = \sum_{i=0}^{j} o_i$ $(j \in \mathcal{J},\ l_0 = 0,$ $\sum_{i=1}^{n} o_i = o)$. The operation $k \in \mathcal{O}$ should be performed on the $\mu_k \in \mathcal{M}$ machine in time (duration) $p_k > 0$. For each job j there is a desired *completion date* d_j and designated a *penalty factor* w_j for exceeding it. The considered variant of the problem consists in determining the moments of starting the execution of jobs on machines so that the technological order is maintained, and minimize the sum of penalties for untimely execution of jobs (*tardiness*).

Over the years of research on the job shop problem, there have emerged many ways of modeling it. In the following part, we will present one of them, the so-called disjunctive graph. A disjunctive graph with weights in vertices can be defined as a pair of sets $G = (V, K \cup D \cup R)$ [4], defined as follows:

1. The set of the vertices of the graph $V = \{1, 2, \ldots, o\} \cup \{s, t\}$, where the vertex i $(i = 1, 2, \ldots, n)$ corresponds to the i-th operation. The weight of a vertex is equal to the execution time of the operation it represents. The remaining (extra) vertices s and t have a weight zero.
2. The set of arcs is the union of the sets:
   (i) *conjunctive arcs* K between successive operations within the job (representing the technological sequence): $K = \bigcup_{j=1}^{n} \bigcup_{i=l_{j-1}+1}^{l_j - 1} \{(i, i+1)\}$,
   (ii) *disjunctive arcs* D between operations performed on the same machine: $D = \bigcup_{i,j \in \mathcal{O}, i \neq j, \nu_i = \nu_j} \{(i, j), (j, i)\}$,

(iii) arcs R, in the form of $(s, l_{j-1}+1)$ and (l_j, t), where $l_{j-1}+1$ is the first and l_j the last operation of the job $j \in \mathcal{J}$.

The subset $D' \subset D$ containing exactly one arc from each pair of arcs is called a *representation* of disjunctive arcs. It can be easily shown that the subgraph $G' = (V, K \cup D' \cup R)$ of the disjunctive graph G represents an feasible solution if and only if G' does not contain cycles.

For a determined feasible representation of D', the time of completion of i-th job C_i should be calculated as the length of the longest path in the graph $G' = (V, K \cup D' \cup R)$ from the vertex s to l_i (l_i – the last operation of job i). Then $T_i = \max\{0, C_i - d_i\}$ is the job's delay (*tardiness*), and $\mathcal{F} = \sum_{i=1}^{n} w_i T_i$ the sum of the costs of tardiness (*penalties*).

3 Uncertainty

We will consider a job shop problem with random parameters, where the job completion times C_i and the due dates d_i are expressed as random variables $\tilde{C}_i$ and $\tilde{d}_i$ with normal or Erlang distributions. More specifically, we will consider the following four cases: (**A**) $\tilde{C}_i \sim N(C_i, \alpha \cdot C_i)$, (**B**) $\tilde{d}_i \sim N(d_i, \beta \cdot d_i)$, (**C**) $\tilde{C}_i \sim \mathcal{E}(C_i, 1)$, (**D**) $\tilde{d}_i \sim \mathcal{E}(d_i, 1)$, where the α, β parameters will be specified at the stage of computational experiments. Additionally, we define a random variable $\tilde{T}_i$ representing the size of delay (tardiness) of i-th job. After randomization, to compare solutions (just like at paper [2]), we use the following comparative criteria: $\mathcal{W}_E = \sum_{i=1}^{n} w_i E(\tilde{T}_i)$, $\mathcal{W}_{ED} = \sum_{i=1}^{n} w_i (E(\tilde{T}_i) + \theta \cdot D^2(\tilde{T}_i))$, where $E(\tilde{T}_i)$ is an expected value of $\tilde{T}_i$, $D^2(\tilde{T}_i)$ is an variation and a parameter $0 < \theta < 1$. Criteria similar to $\mathcal{W}_E$ and $\mathcal{W}_{ED}$ are considered in papers [2,3,7].

Normal Distribution. Here we will consider the following two cases: (**A**) $\tilde{C}_i \sim N(C_i, \alpha \cdot C_i)$ and (**B**) $\tilde{d}_i \sim N(d_i, \beta \cdot d_i)$. If X is a random variable, then by F_X and f_X we will denote its cumulative distribution function and density.

Case A ($\tilde{C}_i \sim N(C_i, \alpha \cdot C_i)$). Then, tardiness is a random variable defined as $\tilde{T}_i = \max\{0, \tilde{C}_i - d_i\}$. For shortening of the notation, let $\mu_i = (d_i - C_i)/\sigma_i$.

Theorem 1. *If the job completion times are independent random variables with a normal distribution $\tilde{C}_i \sim N(C_i, \sigma_i)$ ($\sigma_i = \alpha \cdot C_i$, then the expected value*

$$E(\tilde{T}_i) = (1 - F_{\tilde{C}_i}(d_i)) \left(\frac{\sigma_i}{\sqrt{2\pi}} e^{\frac{-(\mu_i)^2}{2}} + (C_i - d_i) \left(1 - F_{N(0,1)}(\mu_i)\right) \right).$$

For calculating the value of the criterion function $\mathcal{W}_{ED}$ it is necessary to know the standard deviation of the random variable $\tilde{T}_i$. Because $D^2(\tilde{T}_i) = E(\tilde{T}_i^2) = (E(\tilde{T}_i))^2$, we will present a method for determining the random variable $\tilde{T}_i^2$.

Theorem 2. *If the times of execution of jobs are independent random variables with a $\tilde{C}_i \sim N(C_i, \sigma_i)$ ($\sigma_i = \alpha \cdot C_i$, $i \in \mathcal{J}$), then the expected value of the random* *variable* $\tilde{T_i}^2$ *constituting the square of delay:*

$$E(\tilde{T_i}^2) = \left(1 - F_{N(0,1)}(\nu_i)\right)\left(\frac{(C_i - d_i)\sigma_i}{\sqrt{2\pi}} e^{\frac{-(\mu_i)^2}{2}} \right.$$
$$\left. +(C_i^2 + \sigma_i^2 + d_i^2 - 2d_i\sigma_i)\left(1 - F_{N(0,1)}(\mu_i)\right)\right).$$

Both proven theorems allow quick calculation of the first two central moments for random dates of jobs' completion.

Case B $(\tilde{d_i} \sim N(d_i, \beta \cdot d_i))$**.** Tardiness of job $i \in \mathcal{J}$ is $\tilde{T_i} = \max\{0, C_i - \tilde{d_i}\}$. We will now present two theorems that allow us to calculate the value of $\mathcal{W}_E$ and $\mathcal{W}_{ED}$. We omit the proofs of theorems because they are similar to the proofs of Theorems 1 and 2.

Theorem 3. *If the requested due dates for completing jobs are independent random variables with a normal distribution* $\tilde{d_i} \sim N(d_i, \sigma_i)$ $(\sigma_i = \beta \cdot d_i,\ i \in \mathcal{J})$*, then the expected tardiness value*

$$E\left(\tilde{T_i}\right) = F_{N(0,1)}(-\mu_i)\left(C_i F_{N(0,1)}(-\mu_i) + \frac{\sigma_i}{\sqrt{2\pi}} e^{-(-\mu_i)^2}2 - d_i F_{N(0,1)}(-\mu_i)\right).$$

Theorem 4. *If the latest due dates for completing jobs are normally distributed independent random variables* $\tilde{d_i} \sim N(d_i, \sigma_i)$ $(\sigma_i = \beta \cdot d_i,\ i \in \mathcal{J})$*, then the expected value of the random variable* $\tilde{T_i}^2$

$$E(\tilde{T_i}^2) = F_{N(0,1)}(-\mu_i)\left((d_i^2 + \sigma_i^2 + C_i^2 - 2C_i d_i)F_{N(0,1)}(-\mu_i) \right.$$
$$\left. + \frac{\sigma_i(C_i - d_i)}{\sqrt{2\pi}} e^{\frac{-(\mu)^2}{2}}\right).$$

Erlang Distribution. It results directly from the definition of the Erlang distribution that the density function of the random variable $\tilde{C_i} \sim \mathcal{E}(\nu_i, \lambda_i)$

$$f_{\tilde{C_i}}(x) = \begin{cases} \frac{1}{(\nu_i-1)!}\lambda_i^{\nu_i} x^{\nu_i - 1} e^{-\lambda_i x}, & \text{if } x > 0, \\ 0, & \text{if } x \le 0. \end{cases} \tag{1}$$

We will consider two cases: (**C**) $\tilde{C_i} \sim \mathcal{E}(C_i, 1)$ and (**D**) $\tilde{d_i} \sim \mathcal{E}(d_i, 1)$.

Case C $(\tilde{C_i} \sim \mathcal{E}(C_i, 1))$**.** In this case, the tardiness in the execution times $\widetilde{T_i} = \max\{0, \widetilde{C_i} - d_i\}$. Similarly, as in the case of normal distribution, we present two theorems claim that is used when calculating the value of the criterion comparing the solution.

Theorem 5. *If the job completion times* $\tilde{C_i}$ *are independent random variables with the Erlang distribution* $\mathcal{E}(C_i, 1)$*, then the expected value of the job tardiness*

$$E(\tilde{T_i}) = (1 - F_{\mathcal{E}(C_i,1)}(d_i))\left(C_i\left(1 - F_{\mathcal{E}(C_i+1,1)}(d_i)\right) - d_i\left(1 - F_{\mathcal{E}(C_i,1)}(d_i)\right)\right).$$

Theorem 6. *If the job completion times $\tilde{C}_i$ are independent random variables with the Erlang distribution $\mathcal{E}(C_i, 1)$, $i \in \mathcal{J}$, then the expected value*

$$E(\tilde{T_i}^2) = \left(1 - F_{\mathcal{E}(C_i,1)}(d_i)\right)\left(C_i(C_i+1)\left(1 - F_{\mathcal{E}(C_i+2,1)}(d_i)\right) - 2d_iC_i\left(1 - F_{\mathcal{E}(C_i+1,1)}(d_i)\right) + d_i^2\left(1 - F_{\mathcal{E}(C_i,1)}(d_i)\right)\right).$$

In summary, when the job completion times arerandom variables with the Erlang distribution, then the values of the function $\mathcal{W}_E$ and $\mathcal{W}_{ED}$ are calculated using the Theorem 5 and 6.

Case D ($\tilde{d_i} \sim \mathcal{E}(d_i, 1)$)**.** Similarly, as for case **C**, we determine the cumulative distribution function of the density of job tardiness $i \in \mathcal{J}$ given as $\widetilde{T_i} = \max\{0, C_i - \widetilde{d_i}\}$. Then, we can proceed to calculate the expected values of random variables $\tilde{T_i}$ and $\tilde{T_i}^2$.

Theorem 7. *If the deadlines are independent random variables with the Erlang distribution $\tilde{d_i} \sim \mathcal{E}(d_i, 1)$, $i = 1, 2, \ldots n$, then the expected value of tardiness $E(\tilde{T_i}) = F_{\mathcal{E}(\delta,1)}(C_i)(C_iF_{\mathcal{E}(\delta,d_i)}(C_i) - \frac{\delta}{d_i}F_{\mathcal{E}(\delta+1,1)}(C_i))$, where $\delta = d_1 + \ldots + d_i$.*

Theorem 8. *If the due dates for completing jobs are independent random variables with the Erlang distribution $\tilde{d_i} \sim \mathcal{E}(d_i, 1)$, then the expected value of the square of tardiness for $\delta = d_1 + \ldots + d_i$ is*

$$E(\tilde{T_i}^2) = F_{\mathcal{E}(\delta,1)}(C_i)\left((\delta+1)F_{\mathcal{E}(\delta+2,1)}(C_i) - 2C_i\delta F_{\mathcal{E}(\delta+1,1)}(C_i) + C_i^2F_{\mathcal{E}(\delta,1)}(C_i)\right).$$

Equalities (Theorem 7 and 8) will be used in $\mathcal{W}_E$ and $\mathcal{W}_{ED}$ calculation.

4 Computational Experiments

In order to carry out computational experiments, a simplified version of the tabu search algorithm for solving the job shop problem presented in the work [4] was adopted. Later in this section the deterministic algorithm will be denoted by $\mathcal{AD}$, whereas probabilistic one by $\mathcal{AP}$, where $\mathcal{AP}_E$ denoted probabilistic algorithm with criterion $\mathcal{W}_E$ and $\mathcal{AP}_{ED}$ – with $\mathcal{W}_{ED}$ criterion.

Test instances of 82 examples of deterministic data were taken from OR-Library [6]. Additionally, in accordance with the uniform distribution, the coefficients of the penalty function for the delay of the jobs w_i were drawn from the set $\{1, 2, \ldots, 10\}$. Similarly, the requested completion dates for the d_i jobs were drawn from the set $[P_i, (3/2)P_i]$, where $P_i = \sum_{j=1}^{m} p_{i,j}$, $i = 1, 2, \ldots, n$. The set of these (deterministic) data was marked with Ω. Let F be the solution value determined by the tested algorithm, and F^* the value of the reference solution. Relative error of the solution F given by $\delta = \frac{F - F^*}{F^*} \cdot 100\%$ indicates by how many percent the solution of the algorithm is worse/better than reference one.

Let $\tilde{\eta}$ be an instance of probabilistic data, $\mathcal{Z}(\tilde{\eta})$ data set generated from $\tilde{\eta}$ by disturbance of task execution times according to the assumed schedule. We have further: $\mathtt{A}_{ref}$ – algorithm designating reference solutions, $\mathtt{A}$ – algorithm whose resistance are tested (in our case $\mathtt{A}_{\mathcal{P}}$ or $\mathtt{A}_{\mathcal{P}+\mathcal{B}}$), $\pi_{(\mathtt{A},x)}$ – solution designated by the algorithm $\mathtt{A}$ for data x, $F(\pi_{(\mathtt{A},x)}, y)$ – value of criterion function of solution $\pi_{(\mathtt{A},x)}$ for the instance y. Then $\Delta(\mathtt{A}, \tilde{\eta}, \mathcal{Z}(\tilde{\eta})) = \frac{\sum_{\varphi \in \mathcal{Z}(\tilde{\eta})} F(\pi_{\mathtt{A},\tilde{\eta}}, \varphi) - \sum_{\varphi \in \mathcal{Z}(\tilde{\eta})} F(\pi_{(\mathtt{A}_{ref}, \varphi)}, \varphi)}{\sum_{\varphi \in \mathcal{Z}(\tilde{\eta})} F(\pi_{(\mathtt{A}_{ref}, \varphi)}, \varphi)}$, we call *resistance of solution* $\pi_{(\mathtt{A},\tilde{\eta})}$ (designated by the algorithm $\mathtt{A}$ for the instance $\tilde{\eta}$) on the set of distributed data $\mathcal{Z}(\tilde{\eta})$.

If $\tilde{\mathcal{P}}$ is a set of instances of probabilistic data of the examined problem, then the expression $\mathcal{S}(\mathtt{A}, \tilde{\mathcal{P}}) = \frac{1}{|\tilde{\mathcal{P}}|} \sum_{\tilde{\eta} \in \tilde{\mathcal{P}}} \Delta(\mathtt{A}, \tilde{\eta}, \mathcal{Z}(\tilde{\eta}))$ we call a *resistance coefficient* of the algorithm $\mathtt{A}$ on the set $\tilde{\mathcal{P}}$. For one deterministic instance, we generate three instances of probabilistic data from the set $\mathcal{D}$. In total, *probabilistic data set* $\tilde{\mathcal{P}}$ has 1125 instances.

The most important results of the computational experiments are presented in Tables 1 and 2. The $\mathcal{AD}$ column contains the results of the deterministic algorithm. Columns labeled $\mathcal{AP}_E$ and $\mathcal{AP}_{ED}$ contain the results of the probabilistic algorithm based on the expected value or the expected value and variance, respectively. In addition to the mean errors, additional columns were added with the percentage of cases for which the probabilistic algorithm gave results not worse than the deterministic algorithm (column %NG) and the corresponding percentage for which the algorithm gave better results (in tables this column is marked as %L).

Table 1. The robustness of the $\mathcal{AD}$ and $\mathcal{AP}$ algorithms expressed as a percentage for random times with **normal distribution.**

Random	α	$\mathcal{AD}$	$\mathcal{AP}_E$	%NG	%L	$\mathcal{AP}_{ED}$	%NG	%L
Durations	0.1	60.93	68.34	70	51	73.17	56	49
	0.2	69.91	78.58	66	62	88.81	49	44
	0.3	71.13	81.12	62	60	88.65	46	45
	0.4	80.41	85.68	63	62	92.16	52	52
	Mean	**70.60**	**78.43**	**65**	**59**	**85.70**	**51**	**48**
Due dates	0.1	24.49	23.82	72	65	24.40	65	61
	0.2	7.28	5.80	66	63	5.61	67	66
	0.3	3.06	2.16	77	76	2.17	72	71
	0.4	1.74	1.14	80	80	1.14	78	78
	Mean	**9.15**	**8.23**	**74**	**71**	**8.33**	**70**	**69**

Part 'due dates' of Table 1 presents the results of computational experiments for the case in which the requested due dates for completing jobs are random variables with a normal distribution. This time we note that the resistance coefficients are better for the probabilistic algorithm and this is the case for all β parameter values. Moreover, also the percentage of cases for which the probabilistic algorithm was not worse or better is clearly in favor of the probabilistic algorithm.

Table 2. The robustness of the $\mathcal{AD}$ and $\mathcal{AP}$ algorithms expressed as a percentage for random times with **Erlang distribution.**

Random	$\mathcal{AD}$	$\mathcal{AP}_E$	%NG	%L	$\mathcal{AP}_{ED}$	%NG	%L
Durations	57.20	57.14	90	54	268.32	45	38
Due dates	51.86	52.20	90	56	53.50	78	55

The jobs execution times in 'durations' part of Table 2 are random. We note that the resistance coefficients are slightly better for the probabilistic algorithm. Also, the percentage of cases for which the probabilistic algorithm was not worse or better than the deterministic algorithm is significantly in favor of the probabilistic algorithm. In turn, 'due dates' part of Table 2 presents test results for the case in which the required due dates for completing jobs have the Erlang distribution. We note that the resistance coefficients are slightly better for the deterministic algorithm. On the other hand, however, if we look at the percentage of cases for which the probabilistic algorithm was not worse or better than the deterministic algorithm, there is a noticeable advantage of the probabilistic algorithm.

5 Summary

The performed computational experiments have shown that the solutions determined by the probabilistic version of algorithm are more robust to disturbances (that is: stable) than the ones determined by the deterministic algorithm, for majority of considered variants. Thus, the concept of uncertainty modeling and comparing the solutions (being random variables) for solving the job shop problem with uncertain parameters were confirmed.

Acknowledgments. The paper was partially supported by the National Science Centre of Poland, grant OPUS no. 2017/25/B/ST7/02181.

References

1. Balas, E., Simonetti, N., Vazacopoulos, A.: Job shop scheduling with setup times, deadlines and precedence constraints. J. Sched. **11**(4), 253–262 (2008). https://doi.org/10.1007/s10951-008-0067-7
2. Bożejko, W., Rajba, P., Wodecki, M.: Stable scheduling of single machine with probabilistic parameters. Bull. Pol. Acad. Sci. Tech. Sci. **65**(2), 219–231 (2017)

3. Bożejko, W., Rajba, P., Wodecki, M.: Robustness of the uncertain single machine total weighted tardiness problem with elimination criteria applied. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) DepCoS-RELCOMEX 2018. AISC, vol. 761, pp. 94–103. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-91446-6_10
4. Grabowski J., Wodecki M.: A very fast tabu search algorithm for job shop problem. In: Rego, C., Alidaee, B. (eds.) Metaheuristic Optimization Via Memory and Evolution, pp. 117–144. Kluwer Academic Publishers (2005)
5. Kim, J.G., Jun, H.B., Bang, J.Y., Shin, J.H., Choi, S.H.: Minimizing tardiness penalty costs in job shop scheduling under maximum allowable tardiness. Processes **8**(11), 1398 (2020)
6. OR-Library. http://www.brunel.ac.uk/~mastjjb/jeb/info.html
7. Rajba, P., Wodecki, M.: Sampling method for the flow shop with uncertain parameters. In: Saeed, K., Homenda, W., Chaki, R. (eds.) CISIM 2017. LNCS, vol. 10244, pp. 580–591. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59105-6_50
8. Sharma, P., Jain, A.: A review on job shop scheduling with setup times. J. Eng. Manuf. **230**(3), 517–533 (2016)
9. Shoval, S., Efatmaneshnik, M.: A probabilistic approach to the Stochastic Job-Shop Scheduling problem. Procedia Manuf. **21**, 533–540 (2018)