

Research Article

Stochastic Machine Scheduling to Minimize Waiting Time Related Objectives with Emergency Jobs

Lianmin Zhang,¹ Lei Guan,² and Ke Zhou³

¹ School of Management and Engineering, Nanjing University, Nanjing 210093, China

² School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

³ Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong

Correspondence should be addressed to Lei Guan; guanlei@bit.edu.cn

Received 14 March 2014; Accepted 24 April 2014; Published 11 May 2014

Academic Editor: Xiaolin Xu

Copyright © 2014 Lianmin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider a new scheduling model where emergency jobs appear during the processing of current jobs and must be processed immediately after the present job is completed. All jobs have random processing times and should be completed on a single machine. The most common case of the model is the surgery scheduling problem, where some elective surgeries are to be arranged in an operation room when emergency cases are coming during the operating procedure of the elective surgeries. Two objective functions are proposed to display this practice in machine scheduling problem. One is the weighted sum of the waiting times and the other is the weighted discounted cost function of the waiting times. We address some optimal policies to minimize these objectives.

1. Introduction

Machine scheduling problems have attracted researchers for decades since they play an important role in various applications from the areas of operations research, management science, and computer science. There are lots of works of the literature published on these problems. However, most of them study the deterministic case where all the information about jobs and machines is completed without any uncertainty; for example, job processing times are assumed to be exactly known in advance. But it is really hard to know the exact values of these parameters in practical situations and thus such an assumption is hardly justifiable. Instead, one sometimes can only roughly estimate the values of these parameters or their probability distributions. As Albert Einstein said, “As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality.”

In addition, the majority of machine scheduling problems studied in the literature assume that all jobs are prepared for processing before scheduling. In reality, however, it is a common phenomenon that new jobs may come randomly from time to time. The necessity to account for random

coming of jobs is a key drive of developing stochastic approaches to scheduling problems.

In this paper, we study the single machine stochastic scheduling problem with emergency jobs. In our study, we focus on optimal policies to schedule current jobs with some objectives on a single machine when emergency jobs appear during the processing of current jobs and must be processed immediately after the present job is completed. All processing times of jobs are not known. The most common case of the model is the surgery scheduling problem, where some elective surgeries are to be arranged in an operation room when emergency cases are coming during the operating procedure of the elective surgeries. Those emergency cases must be operated as soon as possible. However, since the present surgery occupies the operation room, they should wait to operate until the current surgery is completed.

As indicated by Rothkopf and Smith [1], there are two basic classes of delay costs considered in scheduling problems. The first class includes linear costs, which consider no discount of the value of money over time, whereas the other class involves exponential functions to represent the discounts as a function of time. In our research, two objective functions are expected to minimize. One is the weighted sum

of jobs' waiting time and the other is the weighted discounted function of jobs' waiting times. Our first objective belongs to the first class, while our second objective function belongs to the second class. We will give optimal policies to solve those problems.

The remainder of this paper is organized as follows. In Section 2, the related works of the literature will be reviewed. Notations and some definitions of the model are provided in Section 3. And specially, Section 3.1 investigates the optimal policy to schedule jobs with the objective to minimize the total weighted waiting time of jobs. The optimal policy for scheduling jobs to minimize the discount cost of the waiting time of jobs is included in Section 3.2. Concluding remarks and further research are presented in Section 4.

2. Literature Review

There has been an enormous amount of work on single machine scheduling. We do not intend to do a complete review of results in the area and restrict our attention to works of the literature directly related to the matter of this paper.

The first is scheduling jobs on machines minimizing the total weighted completion time of jobs. It has attracted a great deal of attention, partly because of its importance as a fundamental problem in scheduling and also because of new applications, for instance, in manufacturing, compiler optimization, and parallel computing. Researchers have studied it extensively in different environments, such as allowing different release times of jobs, multimachine scheduling, and online scheduling. We here just review the scheduling problem on a single machine and all jobs have the same release times.

Perhaps the most famous scheduling policy with this objective in the deterministic case is Smith's rule, also known as the WSPT (weighted shortest processing time) rule. It schedules jobs in the order of nonincreasing ratio $\{w_j/p_j\}$, where p_j denotes the deterministic processing time of job j and w_j is the weight assigned to job j , and it is shown to be optimal for scheduling jobs on a single machine by Smith [2]. Extending this rule to stochastic environment, firstly Rothkopf [4, 5] develops the WSEPT (weighted shortest expected processing time) rule, which schedules jobs in the order of nonincreasing ratio $\{w_j/\mathbb{E}[P_j]\}$. Rothkopf [4, 5] proves that the WSEPT rule is optimal for single machine scheduling with identical release dates. The technique adopted is to reduce the problem to the deterministic case by taking the expectations of the processing times. This policy, however, fails in the multimachine case with general processing times, even if the weights are identical; that is, $w_i \equiv w$. For more details about this research, we refer readers to see [3].

Another stream related to our research is to minimize total weight discounted cost function, such as Rothkopf [4, 5]. Recently, Cai et al. [6] consider the problem of scheduling a set of jobs on a single machine subject to random breakdowns. They study the preemptive-repeat model, which addresses the situation where, if a machine breaks down during the processing of a job, the work done on the job prior to the breakdown is lost and the job will have to be started from the beginning again when the machine resumes its

work. They obtain the optimal policy for a class of problems to minimize the expected discounted cost from completing all jobs. For more results in the area, see the survey [3].

What differentiates our paper from the above works of literature is that we consider a new scheduling problem where the emergency jobs are constantly coming and should be placed as soon as possible. To the best of our knowledge, there is no previous research studying this problem. Therefore, we open up a new research direction for the scheduling problem. In addition, different from previous works of the literature considering the completion time of jobs, we focus on the waiting time of jobs. It is more reasonable in our scheduling environment.

3. Models and Optimal Policies

We here consider the stochastic scheduling problem with emergency jobs. Specially, what we want to study is how to schedule current jobs on a single machine when emergency jobs appear during the processing of jobs and must be processed immediately after the present job is completed. Without loss of generality, in the following we assume that the emergent stream of jobs has a Poisson process with rate λ . The processing time Z_j for the emergency job j is arbitrarily distributed with the mean μ .

Let $J = \{1, 2, \dots, n\}$ be a set of jobs to be processed nonpreemptively on a single machine. In other words, if a job starts to process, other jobs must wait to be processed until the present job is completed. Assume that the processing time P_j of current job j is a random variable. We denote W_j as the waiting time of job j and let γ_j be the unit waiting cost of job j . In the following, we will propose two objectives and study the optimal policies to minimize those functions.

3.1. Objective: $\mathbb{E}[\sum \gamma_i W_i]$. In this subsection, we will first focus on the objective function to minimize total weighted waiting cost, written as $\mathbb{E}[\sum \gamma_i W_i]$ for short. We here will address that the well-known policy in stochastic single machine scheduling, WSEPT, is also an optimal policy for our model.

Theorem 1. *WSEPT, that is, sequencing jobs in nonincreasing order of $\gamma_j/\mathbb{E}[P_j]$, is an optimal nonpreemptive policy to minimize total weighted waiting cost.*

Proof. Consider the schedule $1, 2, \dots, n$, denoted as Π . For the first job 1, its waiting time is 0. Based on the fact that new jobs' coming follows the Poisson process, it is easy to know that the expected number of emergency jobs, before job 2 starts to process, is $(\lambda/(1 - \lambda))\mathbb{E}[P_1]$. Therefore, the expected waiting time of job 2 is $(1 + \lambda\mu/(1 - \lambda))\mathbb{E}[P_1]$, since each new job has mean processing time μ . Let C_j be the completion time of job j . Similarly, we can obtain that

$$\begin{aligned}\mathbb{E}[W_i] &= \mathbb{E}[C_{i-1}] + \frac{\lambda\mu\mathbb{E}[P_{i-1}]}{1 - \lambda}, \\ \mathbb{E}[C_{i-1}] &= \mathbb{E}[W_{i-1}] + \mathbb{E}[P_{i-1}], \\ \mathbb{E}[P_0] &= 0.\end{aligned}\tag{1}$$

Thus, we have

$$\begin{aligned}
 \mathbb{E} \left[\sum_{i=1}^n \gamma_i W_i \right] &= \sum_{i=1}^n \gamma_i \mathbb{E} [W_i] \\
 &= \sum_{i=1}^n \gamma_i \left(\mathbb{E} [C_{i-1}] + \frac{\lambda \mu \mathbb{E} [P_{i-1}]}{1 - \lambda} \right) \\
 &= \sum_{i=1}^n \gamma_i \mathbb{E} [W_{i-1}] + \sum_{i=1}^n \gamma_i \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) \mathbb{E} [P_{i-1}] \\
 &= \sum_{i=1}^n \gamma_i \mathbb{E} [W_{i-2}] + \sum_{i=1}^n \gamma_i \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) \mathbb{E} [P_{i-2}] \\
 &\quad + \sum_{i=1}^n \gamma_i \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) \mathbb{E} [P_{i-1}] \\
 &= \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) \sum_{i=1}^n \gamma_i \sum_{k=1}^{i-1} \mathbb{E} [P_k].
 \end{aligned} \tag{2}$$

Without loss of generality, we assume that jobs are ordered in nonincreasing order of $\gamma_j/\mathbb{E}[P_j]$; that is, $\gamma_1/\mathbb{E}[P_1] \geq \gamma_2/\mathbb{E}[P_2] \geq \dots \geq \gamma_n/\mathbb{E}[P_n]$. If this sequence, denoted by Π for brevity, is not an optimal policy, there must exist job j in some optimal policy such that job $j + 1$ is processed before job j . Without loss of generality, let $1, 2, \dots, j - 1, j + 1, j, j + 2, \dots, n$, denoted by Π' , be such optimal policy and we hence have $\Phi(\Pi) > \Phi(\Pi')$. Comparing these two policies and observing that the waiting cost of jobs $1, 2, \dots, j - 1, j + 2, \dots, n$ is not changed, we obtain

$$\begin{aligned}
 \Phi(\Pi) - \Phi(\Pi') &= \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) \\
 &\quad \times \left(\sum_{i=j}^{j+1} \gamma_i \sum_{k=1}^{i-1} \mathbb{E} [P_k] - \gamma_{j+1} \sum_{k=1}^{j-1} \mathbb{E} [P_k] \right. \\
 &\quad \left. - \gamma_j \sum_{k=1}^{j-1} \mathbb{E} [P_k] - \gamma_j \mathbb{E} [P_{j+1}] \right) \\
 &= \left(\frac{\lambda \mu - \lambda + 1}{1 - \lambda} \right) (\gamma_{j+1} \mathbb{E} [P_j] - \gamma_j \mathbb{E} [P_{j+1}]) < 0.
 \end{aligned} \tag{3}$$

Therefore, $\Phi(\Pi) < \Phi(\Pi')$, which leads to a contradiction. With the adjacent pairwise interchange, we can know that WSEPT is an optimal policy. This completes the proof. \square

Remarks. (1) In fact, the optimal policy for the objective with minimizing total expected waiting cost is equal to that

with minimizing total expected completion time. This can be shown according to the following equation:

$$\begin{aligned}
 \mathbb{E} \left[\sum \gamma_i C_i \right] &= \mathbb{E} \left[\sum \gamma_i (W_i + P_i) \right] \\
 &= \mathbb{E} \left[\sum \gamma_i (W_i) \right] + \sum \gamma_i \mathbb{E} [P_i].
 \end{aligned} \tag{4}$$

It is well known that WSEPT optimally minimizes the total expected completion time for single machine problem, which coincides with the result we obtained. But here we consider a new scheduling model. Hence, this theorem extends the well-known optimal policy to a new scheduling environment.

(2) The theorem shows that it is simple to construct the policy to minimize the total weighted waiting time of jobs. The only information for the optimal policy is the weight and the expected processing time of jobs, while the exact distribution of processing times of jobs is needed so as to obtain optimal policy with simple structure in most works of the literature in stochastic scheduling.

3.2. Objective: $\sum \gamma_i \mathbb{E}[1 - \exp(-\alpha W_i)]$. In this section, we study the objective of weighted discounted cost of the waiting times; that is, $\sum \gamma_i \mathbb{E}[1 - \exp(-\alpha W_i)]$. Here α is the discount factor. Before giving the optimal policy for our model, we will release two lemmas.

Lemma 2. Consider

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^\lambda. \tag{5}$$

Proof. Let x be a random variable and be followed with Poisson distribution. We then have

$$\sum_{x=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^x}{x!} = 1. \tag{6}$$

By multiplying e^λ to both sides of the equality, we have $\sum_{x=0}^{\infty} (\lambda^x/x!) = e^\lambda$. \square

The other lemma which will be used is stated as follows.

Lemma 3. For any a , if x is followed with Poisson distribution, then $\mathbb{E}[e^{ax}] = e^{\lambda(e^a - 1)}$.

Proof. Since x is followed with Poisson distribution, we obtain

$$\begin{aligned}
 \mathbb{E} [e^{ax}] &= \sum_{x=0}^{\infty} e^{ax} \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x e^{ax}}{x!} \\
 &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^a)^x}{x!} = e^{-\lambda} e^{e^a \lambda} = e^{\lambda(e^a - 1)}.
 \end{aligned} \tag{7}$$

\square

We now can prove our second main theorem as follows.

Theorem 4. Sequencing jobs with the nonincreasing order of

$$\frac{\gamma_j}{1 - f_j} \quad (8)$$

are an optimal policy to minimize the objective $\sum \gamma_i \mathbb{E}[1 - \exp(-\alpha W_i)]$, where

$$f_j = \mathbb{E} \left[e^{(-\alpha - \lambda + \lambda E[e^{-\alpha Z_1}]) P_j} \right]. \quad (9)$$

Proof. Assume that $1, 2, \dots, n$ is a feasible schedule. Let Y_j be the random variable that denotes the total time of processing job j and processing emergence jobs arriving during the processing of job j . And let N_j denote the random variable as the number of emergence jobs during the processing of job j . Observing that the emergence jobs arrive according to the Poisson distribution with rate λ , we know that random variable $(N_j \mid P_j)$ has a Poisson distribution with mean $(\lambda/(1 - \lambda))P_j$. Thus,

$$\begin{aligned} \mathbb{E} \left[e^{-\alpha Y_j} \right] &= \mathbb{E} \left[e^{-\alpha(P_j + N_j Z_j)} \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[e^{-\alpha(P_j + N_j Z_j)} \mid P_j \right] \right] \\ &= \mathbb{E} \left[e^{-\alpha P_j} \mathbb{E} \left[e^{(-\alpha Z_j) \cdot N_j} \mid P_j \right] \right] \\ &= \mathbb{E} \left[e^{-\alpha P_j} e^{(\lambda/(1-\lambda))P_j(\mathbb{E}[e^{-\alpha Z_j}] - 1)} \right] \\ &= \mathbb{E} \left[e^{-\alpha P_j - (\lambda/(1-\lambda))P_j + \mathbb{E}[e^{-\alpha Z_j}]P_j} \right] = f_j. \end{aligned} \quad (10)$$

Based on the fact that all $\{Y_k\}_{k=1}^{k=n}$ are independent random variables, we thus can know that

$$\mathbb{E} \left[e^{-\alpha W_j} \right] = \mathbb{E} \left[e^{-\sum_{k=1}^{j-1} \alpha Y_k} \right] = \prod_{k=1}^{j-1} f_k. \quad (11)$$

Therefore, the objective function under sequence j_1, j_2, \dots, j_n results is

$$\sum_{i=1}^n r_{j_i} \left(1 - \prod_{k=1}^{i-1} f_k \right). \quad (12)$$

Without loss of generality, we assume that jobs are ordered in nonincreasing order of $\gamma_j/(1 - f_j)$; that is, $\gamma_1/(1 - f_1) \geq \gamma_2/(1 - f_2) \geq \dots \geq \gamma_n/(1 - f_n)$. If this sequence, denoted by Π , is not an optimal policy, there must exist j in some optimal policy such that job $j + 1$ is processed before job j . Without loss of generality, let $1, 2, \dots, j-1, j+1, j, j+2, \dots, n$, denoted by Π' , be such optimal policy and we hence have $\Phi(\Pi) > \Phi(\Pi')$. Comparing these two policies and observing

that the waiting cost of jobs $1, 2, \dots, j-1, j+2, \dots, n$ is not changed, we obtain

$$\begin{aligned} \Phi(\Pi) - \Phi(\Pi') &= \sum_{i=j}^{j+1} r_i \left(1 - \prod_{k=1}^{i-1} f_k \right) - r_{j+1} \left(1 - \prod_{k=1}^{j-1} f_k \right) \\ &\quad - r_j \left(1 - f_{j+1} \prod_{k=1}^{j-1} f_k \right) \\ &= -r_j (1 - f_{j+1}) \prod_{k=1}^{j-1} f_k + r_{j+1} (1 - f_j) \prod_{k=1}^{j-1} f_k \\ &= (r_{j+1} (1 - f_j) - r_j (1 - f_{j+1})) \prod_{k=1}^{j-1} f_k < 0. \end{aligned} \quad (13)$$

Therefore, we have $\Phi(\Pi) < \Phi(\Pi')$, which leads to a contradiction. With the adjacent pairwise interchange, we can know that nonincreasing order of $\gamma_j/(1 - f_j)$ is an optimal policy. This completes the proof. \square

Remark. The structure of the optimal policy is a little complex comparing to the optimal policy in the last section. In addition to weight and processing time of jobs, this optimal policy is also related to the parameters λ, Z_1, α . It is reasonable since the objective function is really complex. It is, however, also easy to calculate or estimate those values with the help of computer. Besides, if the processing time of current and emergent jobs is followed with some special distribution, the optimal policy will have very simple structure. One example is stated as follows.

Example. Assume that $P_i \sim \exp(\nu_j)$ and $Z_i \sim \exp(\mu)$. We can easily calculate that

$$\begin{aligned} f_j &= \mathbb{E} \left[e^{(-\alpha - \lambda + \lambda E[e^{-\alpha Z_1}]) P_j} \right] = \mathbb{E} \left[e^{(-\alpha - \lambda + \lambda(\mu/(\mu + \alpha))) P_j} \right] \\ &= \mathbb{E} \left[e^{(-\alpha - (\lambda\alpha/(\mu + \alpha))) P_j} \right] = \frac{\nu_j}{\nu_j + \alpha + (\lambda\alpha/(\mu + \alpha))}. \end{aligned} \quad (14)$$

It is easy to verify that the optimal policy sequences jobs in nonincreasing order of $\gamma_j(\nu_j + \alpha + (\lambda\alpha/(\mu + \alpha)))$.

4. Conclusions and Future Research

In this paper, we consider a new stochastic scheduling model based on the emergency jobs. We give optimal policies for two classes of objectives: one is the total waiting time of jobs and the other is the weighted sum of an exponential function of the waiting times.

One can consider other objectives in scheduling area, such as the weighted number of late jobs. Specially, we can study the environment that each job j has a due date D_j , which follows the exponential distribution. We believe that it is reasonable to study this case. Taking the surgery scheduling

as an example again, we can consider the due date for each elective surgery. One reason is that each elective surgery should have a due date to wait for operating. Otherwise, the patient will die or will leave to search for another hospital for help. The other reason is that in some sense this date can also be as the commitment to the patients. And sometimes this commitment is consistent with patients' condition. Therefore, it can be interpreted as a random variable. In addition, Denton et al. [7] consider the overtime to the model and study how case sequencing affects patient waiting time, operating room idling time (surgeon waiting time), and operation room overtime. Therefore, maybe we can study the objective which includes both the weighted number of late jobs and overtime:

$$\mathbb{E} \left[\sum_{k=1}^K o_k \max \left\{ 0, \sum_{i \in S_k} P_i - \Delta_k \right\} + \sum_j \gamma_j (W_j - D_j)^+ \right], \quad (15)$$

where o_k is the unit overtime cost, Δ_k is the width of the k th regular time, and S_k is the subset of cases scheduled in the k th day.

Another direction we can study is to take the waiting cost of emergency jobs into consideration. The reason is that emergency surgery which cannot be treated in time will result in substantial losses. It reflects that the unit waiting time cost of emergency jobs is very large in the model. It, thus, is necessary to operate these cases as soon as possible. Clearly, sometimes it is nonsense if the new coming jobs follow the Poisson process, since its influence on the time line is the same. Thus, in this case, maybe we can take another distribution for the emergency jobs' coming. Sometimes it is reasonable since in practice emergency jobs' appearance is related to the timing. For example, Summala and Mikkola [8] show that the fatal accidents appear more frequently during 3am–5am and 2pm–3pm, which is very close to emergency surgeries.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

We thank the Guest Editor and the three anonymous referees for their helpful comments and suggestions. The second author is supported by the fundamental research funds of Beijing Institute of Technology (no. 20122142014).

References

- [1] M. H. Rothkopf and S. A. Smith, "There are no undiscovered priority index sequencing rules for minimizing total delay costs," *Operations Research*, vol. 32, no. 2, pp. 451–456, 1984.
- [2] W. E. Smith, "Various optimizers for single-stage production," *Naval Research Logistics Quarterly*, vol. 3, pp. 59–66, 1956.
- [3] X. Q. Cai, X. Y. Wu, L. Zhang, and X. Zhou, *Scheduling with Stochastic Approaches, Sequencing and Scheduling with Inaccurate Data*, Nova, New York, 2014, edited by U. Sotnikov and F. Werner.
- [4] M. H. Rothkopf, "Scheduling independent tasks on parallel processors," *Management Science*, vol. 12, pp. 437–447, 1966.
- [5] M. H. Rothkopf, "Scheduling with random service times," *Management Science*, vol. 12, pp. 707–713, 1966.
- [6] X. Q. Cai, X. Sun, and X. Zhou, "Stochastic scheduling subject to machine breakdowns: the preemptive-repeat model with discounted reward and other criteria," *Naval Research Logistics*, vol. 51, no. 6, pp. 800–817, 2004.
- [7] B. Denton, J. Viapiano, and A. Vogl, "Optimization of surgery sequencing and scheduling decisions under uncertainty," *Health Care Management Science*, vol. 10, no. 1, pp. 13–24, 2007.
- [8] H. Summala and T. Mikkola, "Fatal accidents among car and truck drivers: effects of fatigue, age, and alcohol consumption," *Human Factors*, vol. 36, no. 2, pp. 315–326, 1994.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

