Stochastic scheduling with abandonment : Necessary and sufficient conditions for the optimality of a strict priority policy.

Jean-Philippe Gayon[1], Gang Chen[2], Pierre Lemaire[3]

[1] Université Clermont Auvergne, CNRS, LIMOS, INP Clermont Auvergne, France
j-philippe.gayon@uca.fr
[2] Business School, Sun Yat-sen University, Guangzhou 510275, China
chengang5@mail.sysu.edu.cn
[3] Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France,
pierre.lemaire@grenoble-inp.fr

Mots-clés : *Stochastic scheduling, Abandonment, Impatience, Index policies, Dynamic programming*

1 Abstract

We consider a stochastic scheduling problem where a job is characterized by a general service time distribution, an exponentially distributed lifetime and a reward. A job abandons the system if its waiting time in the queue is larger than its lifetime. The objective is to maximize the total expected revenue, without preemption. For this problem, our main result is to establish a set of necessary and sufficient conditions for the optimality of an index policy, in the case of two types of jobs and homogeneous service times. In the process we provide a new set of sufficient conditions under which an index policy is optimal even with general heterogeneous service times. For the case with heterogeneous and exponential service times, we conjecture that this set of conditions remains necessary and sufficient when one parameter is identical (reward or lifetime rate). We prove this conjecture for the necessity part when allowing preemption, and for the sufficient part in the special case of three jobs.

2 Contributions

When all jobs have the same service time distribution, we obtain with a different approach the sufficient conditions established by [3]. When considering two types of jobs, we extend Proposition 4 of [2], who assume exponential service times, to general service times. Combining the two sets of conditions leads to a new set of necessary and sufficient conditions for the optimality of an index policy. To the best of our knowledge, this paper is the first to establish necessary and sufficient conditions for the optimality of index policies for a stochastic scheduling problem with abandonment. On this topic, we can quote [2] : "Corollary 3 and Propositions 3 and 4 provide us sets of sufficient conditions that lead to the optimality of index policies. They are not necessary conditions however, and index policies might be optimal even when none of these conditions holds."

For the case with heterogeneous and exponential service times, we conjecture that our proposed set of conditions remains necessary and sufficient when one parameter is identical (reward or lifetime rate). This conjecture has been checked by extensive numerical experiments. We prove this conjecture for the necessity part when allowing preemption. For the sufficient part, we prove this conjecture in the special case of three jobs, consistently with the results of [1].

3 Main result

We consider a stochastic scheduling problem where jobs can be categorized into two types, with n_i jobs of type $i = 1, 2$ waiting for service. The lifetime of type i jobs is a random variable D_i exponentially distributed (with mean γ_i). The service time of type i jobs is a random variable X_i (with mean $1/\mu_i$). The service times and lifetimes are independent continuous random variables. A job abandons the system if its waiting time in the queue is larger than its lifetime and can not abandon once in service. A positive reward w_j is earned when job j is taken into service. The service is performed in a nonpreemptive manner. The decision epochs for our dynamic control problem are time zero and service completion instants. The objective is to maximize the expected revenue in the set of non-idling and nonpreemptive dynamic scheduling policies. Note that an idling policy, i.e., a policy under which the server may idle in the presence of jobs, is suboptimal in the sense of this optimization problem [2].

A strict priority (SP) policy is a policy which always gives priority to the same type of job, at any system state and decision epoch. The policy which gives a strict priority to type i jobs will be denoted SP_i.

Theorem 1 *Assume that $X_1 =_{st} X_2 =_{st} X$. Policy SP_1 is optimal for any initial state (n_1, n_2) if and only if*

$$(S) : \begin{cases} w_1 \geq w_2 \text{ and} \\ w_1 \mathbb{P}(D_1 < X) \geq w_2 \mathbb{P}(D_2 < X). \end{cases}$$

We prove that (S) is a set of sufficient conditions by combining results from [3] and [2]. The necessary condition $w_1 \geq w_2$ appears when there are many jobs in the system. To prove its necessity, we formulate the problem as a stochastic dynamic program and study the asymptotic behavior of the value function when n_1 and n_2 go to infinity. The necessity of condition $w_1 \mathbb{P}(D_1 < X) \geq w_2 \mathbb{P}(D_2 < X)$ is shown by considering two jobs in the system.

Références

[1] Nilay Tanik Argon, Serhan Ziya, and Rhonda Righter. Scheduling impatient jobs in a clearing system with insights on patient triage in mass casualty incidents. *Probability in the Engineering and Informational Sciences*, 22(3) :301–332, 2008.

[2] Evin Uzun Jacobson, Nilay Tanık Argon, and Serhan Ziya. Priority assignment in emergency response. *Operations Research*, 60(4) :813–832, 2012.

[3] Sheldon M Ross. A sequential scheduling problem with impatient jobs. *Naval Research Logistics*, 62(8) :659–663, 2015.