Duplication and sequencing of unreliable jobs

A. Agnetis[1], P. Detti[1], B. Hermans[2] and M. Pranzo[1]

[1] Università degli studi di Siena,
Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Siena, Italy
`{agnetis,detti,pranzo}@diism.unisi.it`
[2] Katholieke Universiteit Leuven,
Research Centre for Operations Research and Business Statistics (ORSTAT), Leuven, Belgium
`ben.hermans@kuleuven.be`

Keywords: unreliable jobs; job duplication; parallel machines; complexity.

1 Introduction

This paper considers a scenario in which a given set of n *jobs* must be processed on a machine subject to possible *breakdowns*. If a breakdown occurs, the remaining jobs (including the job currently being processed) cannot be performed and are therefore lost. On the other hand, if job j is successfully completed, a revenue r_j is gained. We denote the success probability of job j as p_j. No rescheduling or reactions are possible, hence a *preventive disruption management* perspective is adopted (Qi *et al.* 2006), and the problem is to decide the job sequence before the beginning of the actual execution of the sequence, in such a way that the expected revenue is maximized. In what follows, we denote this problem as 1ERM (i.e., expected revenue maximization on a single machine). Given a sequence σ of jobs on the machine, and denoting the i-th scheduled job as $\sigma(i)$, the expected revenue $ER[\sigma]$ is given by

$$ER[\sigma] = p_{\sigma(1)}r_{\sigma(1)} + p_{\sigma(1)}p_{\sigma(2)}r_{\sigma(2)} + \ldots + p_{\sigma(1)} \cdots p_{\sigma(n-1)}p_{\sigma(n)}r_{\sigma(n)}. \tag{1}$$

It is known that 1ERM can be solved at optimality (Mitten 1960) by sequencing jobs from the greatest to the smallest *Z-ratio*:

$$Z_j = \frac{p_j r_j}{(1-p_j)}. \tag{2}$$

Here we address an extension of 1ERM in which there are m (identical) machines in charge of job processing. In order to hedge against unrecoverable interruptions, we adopt the technique of *duplicating* work on different machines, which is a common strategy in many computing centers (Zhou *et al.* 2016, Benoit *et al.* 2013). In particular, this means that there are m *copies* of each job, one to be executed on each machine. The revenue r_j is gained if *at least one copy* of j is successfully carried out, i.e., even if all copies are completed, the revenue is attained only once. When no job duplication is allowed, problems on parallel machines have been addressed in (Agnetis *et al.* 2017, Agnetis *et al.* 2009, Lee and Yu 2008).

The problem addressed here may arise when considering the execution of a set of tasks on a multi-processor environment, composed by different servers geographically distributed. In general we can assume that servers may fail, connections can be broken and outages may occur. A strategy to increase reliability is to duplicate the execution of the tasks on two or more independent servers possibly in different geographical locations so that in case of failure of a server the computation is still carried out on the other servers. In any case, the revenue is gained only once when one of the computations is over.

Specifically, we address the following problems.

Definition 1. *EXPECTED REVENUE MAXIMIZATION WITH TWO MACHINES (2ERM) – Given n jobs $\{1, 2, \ldots, n\}$, each having success probability p_j and revenue r_j, and two identical machines M_1 and M_2, find a sequence of the n jobs on M_1 and M_2 so that the expected revenue of attaining at least one copy of each job is maximized.*

Definition 2. *KIT AVAILABILITY MAXIMIZATION (mKAM) – Given n jobs $\{1, 2, \ldots, n\}$, each having success probability p_j, and m identical machines, find a sequence of the n jobs on each machine so that the probability of attaining at least one copy of each job is maximized.*

We discuss the problem complexity and introduce a rule based on a modified Z-ratio (2) that, given the sequence on machine M_1, allows to derive an optimal sequence for M_2. Furthermore, when the Z-ratios (2) of the jobs are all equal to 1, the rule allows to easily build an optimal solution for 2ERM.

2 The Expected Revenue Maximization Problem with two machines

We let $P = \prod_{j=1}^{n} p_j$ and assume that $p_j < 1$ for all j (if for some job $p_j = 1$, such a job is obviously processed first with no consequence on the other jobs). Let us first state a result concerning the following situation. Suppose that a job sequence $\bar{\sigma}_1$ has been fixed on M_1, and we let $\bar{p}_j$ denote the cumulative probability up to job j on M_1 in sequence $\bar{\sigma}_1$, i.e.,

$$\bar{p}_j = \prod_{k:k \prec j} p_k.$$

Moreover, we denote by Z'_j the *modified Z-ratio* of job j, defined as

$$Z'_j = Z_j(1 - \bar{p}_j). \tag{3}$$

The problem of finding the optimal sequence on M_2 given $\bar{\sigma}_1$ is solved as shown in the following lemma:

Lemma 1. *If a job sequence $\bar{\sigma}_1$ is fixed on M_1, expected revenue is maximized sequencing the jobs on M_2 by nonincreasing values of*

$$Z'_j = Z_j(1 - \bar{p}_j). \tag{4}$$

Proof. The proof uses an interchange argument. Consider a sequence σ_2 on M_2, and let P_i be the cumulative success probability of job i in σ_2, i.e., $P_i = p_i \prod_{k:k \prec i} p_k$. (Note that P_i includes the probability of job i itself.) Given a fixed sequence $\bar{\sigma}_1$ on M_1 and the associated probabilities $\bar{p}_i$ and $\bar{p}_j$, assume that in σ_2 there are two consecutive jobs j and i such that $j \prec i$ and $Z'_i > Z'_j$. Let σ'_2 be the sequence obtained swapping i and j in σ_2. The expected revenue of $(\bar{\sigma}_1, \sigma'_2)$ can be expressed as

$$ER(\bar{\sigma}_1, \sigma'_2) = A + r_i(P_i + \bar{p}_i - P_i\bar{p}_i) + r_j(P_i p_j + \bar{p}_j - P_i p_j \bar{p}_j) + B$$

while

$$ER(\bar{\sigma}_1, \sigma_2) = A + r_j(P_j + \bar{p}_j - P_j\bar{p}_j) + r_i(P_j p_i + \bar{p}_i - P_j p_i \bar{p}_i) + B,$$

where A and B denote the contribution of jobs preceding and respectively following i and j on M_2 in the two schedules. Denoting with Q the cumulative probability of jobs preceding i and j on M_2, in $(\bar{\sigma}_1, \sigma'_2)$ one has $P_i = Qp_i$ and in $(\bar{\sigma}_1, \sigma_2)$, $P_j = Qp_j$, one has that $ER(\bar{\sigma}_1, \sigma'_2) - ER(\bar{\sigma}_1, \sigma_2) > 0$ if and only if

$$r_i p_i - r_i p_i \bar{p}_i + r_j p_i p_j - r_j p_i p_j \bar{p}_j - (r_j p_j - r_j p_j \bar{p}_j + r_i p_j p_i - r_i p_j p_i \bar{p}_i) > 0,$$

i.e.,

$$r_i p_i(1-\bar{p}_i)(1-p_j) > r_j p_j(1-\bar{p}_j)(1-p_i),$$

and hence

$$Z_i(1-\bar{p}_i) > Z_j(1-\bar{p}_j),$$

which holds since $Z'_i > Z'_j$. By repeatedly applying the above argument, the thesis follows. □

A consequence of Lemma 1 is the following.

Lemma 2. *Consider an instance of 2ERM in which $Z_j = 1$ for all jobs $j = 1, \ldots, n$. Then any schedule in which the jobs are reversely sequenced on the two machines is optimal.* □

Regarding the computational complexity of the problem, we recall that when duplications are not allowed, the problem with 2 machines and unreliable jobs is known to be strongly NP-hard (Agnetis *et. al.* 2009). Concerning 2ERM, it is possible to prove the following result.

Theorem 1. *2ERM is strongly NP-hard.* □

The proof consists in showing that the combinatorial problem PRODUCT PARTITION can be polynomially reduced to 2ERM. PRODUCT PARTITION was proved strongly NP-hard by (Ng *et al.* 2010).

3 The Kit Availability Maximization Problem

The following results can be established for KAM.

Theorem 2. *When there are two machines and n job types (2KAM), the problem can be solved in $O(n)$.*

The problem in which there are m machines and only two job types (1 and 2) consists in deciding the number x of machines that follow the sequence 12, so that $m - x$ will follow the sequence 21. The following result can be established.

Theorem 3. *mKAM with two job types can be solved in $O(\log m)$.* □

References

Agnetis, A., Detti, P., Martineau, P., Scheduling nonpreemptive jobs on parallel machines subject to exponential unrecoverable interruptions, *Computers and Operations Research*, 79, 109-118, 2017.

Agnetis, A., Detti, P., Pranzo, M., Sodhi, M.S., Sequencing unreliable jobs on parallel machines, *Journal of Scheduling*, 12(1), 45-54, 2009.

Benoit, A., Robert, Y., Rosenberg, A.L., Vivien, F., Static Strategies for Worksharing with Unrecoverable Interruptions, *Theory of Comput Systems*, 2013,53,386-423.

Lee, C.-Y., Yu, G., Parallel-machine scheduling under potential disruption, Optimization Letters, 2, 27-37, 2008.

Lee, C.-Y., Yu, G., Single machine scheduling under potential disruption, Operations Research Letters, 35, 541-548, 2007.

Mitten, L.G., An Analytic Solution to the Least Cost Testing Sequence Problem, J. Industrial Engineering, 11, 17, 1960.

Ng, C.T., Barketau, M.S., Cheng, T.C.E., Kovalyov, M.Y., ”Product partition” and related problems of scheduling and systems reliability: Computational complexity and approximation, *European Journal of Operational Research*, 207, 601-604, 2010.

Qi, X., Bard, J.F., Yu, G., Disruption management for machine scheduling: The case of SPT schedules, *International Journal of Production Economics*, 103, 166-184, 2006.

Zhou, A., Wang, S., Cheng, B., Zheng, Z., Yang, F., Chang, R.N., Lyu, M.R., Buyya, R., Cloud service reliability enhancement via virtual machine placement optimization, *IEEE Transactions on Services Computing*, 10 (6), 902-913, 2016.