SCHEDULING DETERIORATING JOBS ON A SINGLE PROCESSOR

SID BROWNE
Columbia University, New York, New York

URI YECHIALI
Tel-Aviv University, Tel-Aviv, Israel

(Received October 1987; revisions received August 1988, February 1989; accepted March 1989)

N jobs are to be processed sequentially on a single machine. While waiting for processing, jobs deteriorate, causing the random processing requirement of each job to grow at a job-specific rate. Under such conditions, the *actual* processing times of the jobs are no longer exchangeable random variables and the expected makespan is no longer invariant under any scheduling strategy that disallows idleness. In this paper, we analyze the effects of different deterioration schemes and derive optimal scheduling policies that minimize the *expected* makespan, and, for some models, policies that minimize the *variance* of the makespan. We also allow for random setup and detaching times. Applications to optimal inventory issuing policies are discussed and extensions are considered.

Typical single processor stochastic scheduling models deal with N jobs waiting to be processed sequentially, with job i having positive random processing requirement X_i. It is usually assumed that X_i is independent of X_j $j \neq i$, $i, j = 1, \ldots, N$, so that the expected makespan (completion time of the N jobs) is *invariant* under any scheduling policy that disallows idleness. Thus, research has centered on minimizing (weighted) flow times (e.g., Conway, Maxwell and Miller 1967) or maximizing rewards (e.g., Ross 1983).

In this paper, we introduce cases where jobs can deteriorate as they await service, causing their processing times to grow (at job-specific rates) during their wait. For these types of models, the makespan is no longer invariant and is a function of the scheduling policy, as are the actual processing times. We consider the class of nonpreemptive processing strategies and find policies to minimize expected makespans for different deterioration schemes and, in some cases, to minimize the variance of the makespan. These models are discussed in greater detail in Browne (1988, Chap. 5) where they were developed to deal with the control of some queueing and communication systems (see also, Browne and Yechiali 1989).

1. LINEAR DETERIORATION

As stated above, we are interested only in nonpreemptive strategies; we also do not allow the processor to idle if jobs are available so that we need consider only the class Π, where policy $\pi \in \Pi$ is a permutation of the index set $I = \{1, 2, \ldots, N\}$ such that $\pi(i) = j$ means that job j (of I) is the ith one to be processed.

As all models of deterioration to be discussed in the sequel yield objective functions of similar form, we state here for reference a well known result (see Rau 1971 who stated it with summation reversed, where it is intimately related to a class of optimal search problems; see also Kelly 1982).

Lemma 1. *The sum*

$$\sum_{i=1}^{N} \mu_{\pi(i)} \prod_{r=i+1}^{N} \gamma_{\pi(r)} \tag{1}$$

is minimized (maximized) when calculated over the permutation ordered by increasing (decreasing) values of $\mu_i/[\gamma_i - 1]$.

(The proof is direct upon an interchange argument.)

Consider N jobs, all available for processing at time 0, with *initial* processing requirements X_i (that is, the (random) time to complete job i if it is processed first). If job i's processing is delayed until t, we assume the initial requirement deteriorates in such a manner that its processing requirement grows linearly with the delay to

$$Y_i(t) = X_i + \alpha_i t$$

where α_i is job i's (specific) processing growth rate. We assume further that a job stops decaying as soon

Subject classifications: Inventory/production, policies, maintenance/replacement: scheduling deteriorating jobs. Probability, renewal processes: modeling probabilistic deterioration. Production/scheduling: scheduling deteriorating jobs.

0030-364X/90/3803-0495 $01.25

as it is put on the processor. For notational convenience consider policy $\pi_0 = (1, 2, \ldots, N)$. Let Y_i denote the actual processing time of job i in π_0, and let $S_k \equiv \sum_1^k Y_i$ denote the completion time of the kth job in π_0 with $S_0 = 0$. Then $Y_j \equiv Y_j(S_{j-1}) = X_j + \alpha_j S_{j-1}$, $j = 1, \ldots, N$, or equivalently, $S_j - (1 + \alpha_j)S_{j-1} = X_j$, $j = 1, \ldots, N$ which exhibits the solution

$$S_j = \sum_{i=1}^{j} X_i \prod_{r=i+1}^{j} (1 + \alpha_r) \tag{2}$$

(where an empty product is defined to be 1).

As S_N is the makespan of π_0, identification of the proper terms in Lemma 1 shows that the expected makespan is minimized when the jobs are scheduled by increasing values of $E(X_i)/\alpha_i$, the ratio of expected initial processing requirement to growth rate.

The decomposition in (2) exhibits clearly how the makespan is simply the sum of delays to all *future* jobs caused by the initial requirements.

Furthermore, as in some applications it is of interest to minimize the *variance* of the makespan rather than its expectation, utilization of (2) suggests

$$\mathrm{Var}(S_N(\pi)) = \sum_{i=1}^{N} \mathrm{Var}(X_{\pi(i)}) \prod_{j=i+1}^{N} (1 + \alpha_{\pi(j)})^2 \tag{3}$$

which upon identification with Lemma 1 implies that the *variance* of the makespan is minimized when the jobs are scheduled by increasing $\mathrm{Var}(X_i)/[(1 + \alpha_i)^2 - 1]$.

Consider the case where job i deteriorates rather as external shocks arrive via a job-specific homogenous Poisson stream of intensity λ_i. Every shock arriving while job i is waiting for processing inflicts random damage, causing job i's processing time to grow by a random jump having mean d_i. The (increased) processing times remain constant between shocks. Let Y_j be the actual processing time of job j (in π_0), $S_j = \sum_{i=1}^{j} Y_i$, $Z_j = E(S_j)$ and let $N_j(t)$ denote the Poisson count of shocks to job j in $(0, t]$ with parameter λ_j. Then if D_{jk} is the *jump* in job j's processing time caused by the kth shock with mean $E(D_{jk}) = d_j$, clearly

$$Y_j = X_j + \sum_{k=1}^{N_j(S_{j-1})} D_{jk}, \quad j = 1, \ldots, N$$

which, upon expectating, yields (for the expected completion times)

$$Z_j = \sum_{i=1}^{j} E(X_i) \prod_{r=i+1}^{j} (1 + \lambda_r d_r). \tag{5}$$

Once again, a glance at Lemma 1 shows that the expected makespan is minimized when the jobs are scheduled in increasing order of $E(X_i)/\lambda_i d_i$.

This policy is also optimal to a first order approximation when in fact the job's processing requirement grows by a *shot noise* type process (see Browne).

In general, it is clearly the linearity of the (expected) rate of growth (during delay) that enables the reduction of the expected makespan to form, such as Equations 2 and 5, hence allowing the optimality of a simple index policy upon application of Lemma 1. Therefore, as long as the deterioration is of a type that causes the jobs' processing times to grow in a (job-specific) Lévy process (that is, a process with stationary independent increments that is continuous in probability, see e.g., Prabhu 1980), an index policy of the form {[EXPECTED INITIAL REQUIREMENT]/[EXPECTED GROWTH RATE]} will minimize the expected (total) completion time or makespan. For example, consider the case where the actual processing requirement of job i in π_0 is

$$Y_i = X_i + \alpha_i S_{i-1} + B_i(S_{i-1}) + \sum_{k=1}^{N_i(S_{i-1})} D_{ik} \tag{6}$$

where $B_i(t)$ denotes a Brownian motion with positive drift μ_i. Then (neglecting the possibility of negative processing times), expectating (6) yields the expected growth rate $\alpha_i + \mu_i + \lambda_i d_i$, and the index is obvious.

In fact, consideration of the following simple inequality suffices to show that these indices correspond to the Gittin's index (Whittle 1981) even though the problem is not directly a multiarmed bandit.

Proposition 1. *If*

$$\frac{E(X_1)}{\alpha_1} < \frac{E(X_2)}{\alpha_2} < \cdots < \frac{E(X_N)}{\alpha_N}$$

then

$$\frac{E(X_K)}{\alpha_K} > \frac{\sum_{i=1}^{K-1} E(X_i) \prod_{r=i+1}^{K-1} (1 + \alpha_r)}{\prod_{r=1}^{K-1} (1 + \alpha_r) - 1} \qquad K = 1, \ldots, N.$$

Proof

$$\sum_{i=1}^{K-1} E(X_i) \prod_{r=i+1}^{K-1} (1 + \alpha_r) < \frac{EX_K}{\alpha_K} \sum_{i=1}^{K-1} \alpha_i \prod_{r=i+1}^{K-1} (1 + \alpha_r) = \frac{EX_K}{\alpha_K} \left[\prod_{r=1}^{K-1} (1 + \alpha_r) - 1 \right].$$

Furthermore, consideration of the interpretation of the initial processing requirement enables us to immediately deduce results for setup times. Specifically, assume that a random time τ_i (independent of X_i)

must be devoted to setting up or adjusting the machine to process job i—during which job i continues to deteriorate. We may assume as well that a random (mutually independent) time θ_i must be expended after processing job i to switch-out (e.g., detach or *cleanse* the machine) of job i. Letting α_i denote the general expected rate at which job i's processing grows during its delay, it is immediate that the expected makespan is minimized by scheduling the jobs in increasing order of the index

$$[\mathrm{E}(X_i) + \mathrm{E}(\tau_i)(1 + \alpha_i) + \mathrm{E}(\theta_i)]/\alpha_i \qquad (7)$$

as now job i's initial service requirement is the setup time for i, the growth caused by this setup, the original X_i, and the time to detach or switch-out of job i.

These results can be treated in the framework of optimal stock depletion with stochastic field lives or optimal inventory issuing policies (Derman and Klein 1958, Brown and Solomon 1973, Albright 1976). The model is usually developed in the context of N identical spares of an item (e.g., batteries) of different ages waiting in a stockpile or on a shelf. Items are issued into the field sequentially upon the *death* of the present working item. Research has previously centered on characterizations of the field life function $\mathrm{L}(x)$ for FIFO or LIFO to be optimal, where if x is an item's shelf life, $\mathrm{L}(x)$ is its field life.

Here we have considered N different items—type i with initial random field life X_i, which if put into the field at t will yield (expected) field life $\mathrm{E}(Y_i(t)) = \mathrm{E}(X_i) - \alpha_i t$. For practical purposes α_i is assumed to be of an order of magnitude such that each spare has a positive expected field life in every permutation $\pi \in \Pi$, $i = 1, \ldots, N$.

As such, it is immediate that expected (total) field life is maximized by issuing the items sequentially into the field by increasing values of $\mathrm{E}(X_i)/\alpha_i$. The fact that the X_i's are of different types motivates the use of setup and detaching times as in Equation 7, because if we allow differences among the spares it is only natural to expect that we would need different times to detach the previously used dead spare and to hook up and adapt a new one.

2. CONCLUSIONS AND OPEN PROBLEMS

We have established policies to minimize expected makespans for some cases of deteriorating jobs. It would be of great interest to determine policies to minimize (as is usual) the sum of weighted completion times. However, if c_j is the waiting cost rate of job j, utilization of Equation 2 yields for the total cost under policy π_0

$$C(\pi_0) = \sum_{j=1}^{N} c_j \sum_{i=1}^{j} X_i \prod_{r=i+1}^{j} (1 + \alpha_r) \qquad (8)$$

which does not yield to an easy analysis even for special cases and is conjectured to be NP-hard, a proof of which still awaits.

However, consideration of an adjacent pairwise interchange yields (where $\pi_1 = (1, 2, \ldots, j-1, j+1, j, j+2, \ldots, N)$, that is, π_1 simply interchanges the jth and $j+1$st terms in π_0)

$$\begin{aligned} C(\pi_0) - C(\pi_1) &= S_{j-1}[c_{j+1}\alpha_j(1 + \alpha_{j+1}) - c_j\alpha_{j+1}(1 + \alpha_j)] \\ &\quad \cdot [X_j c_{j+1}(1 + \alpha_{j+1}) - X_{j+1} c_j(1 + \alpha_j)] \\ &\quad + \sum_{r=j+2}^{N} c_r \prod_{k=j+2}^{r} (1 + \alpha_k)[X_j\alpha_{j+1} - X_{j+1}\alpha_j] \end{aligned}$$

from which the following proposition is apparent.

Proposition 2. *If*

$$\frac{\mathrm{E}(X_1)}{\alpha_1} < \cdots < \frac{\mathrm{E}(X_N)}{\alpha_N}$$

and

$$\frac{\alpha_1}{c_1(1 + \alpha_1)} < \cdots < \frac{\alpha_N}{c_N(1 + \alpha_N)}$$

then π_0 minimizes the weighted expected completion times.

Another open problem is that of nonlinear type deterioration. Consider the simple case of exponential growth

$$Y_i(t) = X_i e^{\beta_i t} \qquad (9)$$

yielding the completion times (in π_0)

$$S_j = S_{j-1} + X_j e^{\beta_j S_{j-1}}$$

from which no closed form expression for the expected makespan appears to exist, although note that for β_j *small* enough for all j (i.e., neglecting $O(\beta_j^2)$ terms), it is obvious that the expected (approximated) makespan is minimized when the jobs are scheduled by decreasing β_i.

It would be interesting to see if any other deterioration functions yield an index policy.

REFERENCES

ALBRIGHT, S. C. 1976. Optimal Stock Depletion With Stochastic Lives. *Mgmt. Sci.* **22,** 852–857.

BROWN, M., AND H. SOLOMON. 1973. Optimal Issuing Policies Under Stochastic Field Lives. *J. Appl. Prob.* **10,** 761–768.

BROWNE, S. 1988. Optimal Dynamic Operating Policies for Cyclic-Type Queues. Unpublished Ph.D. Dissertation, New York University, New York.

BROWNE, S., AND U. YECHIALI. 1989. Dynamic Priority Rules for Cyclic Type Queues. *Adv. Appl. Prob.* **10,** 432–450.

CONWAY, R. W., W. L. MAXWELL AND L. W. MILLER. 1967. *Theory of Scheduling.* Addison-Wesley, Reading, Mass.

DERMAN, C., AND M. KLEIN. 1958. Inventory Depletion Management. *Mgmt. Sci.* **4,** 450–456.

KELLY, F. P. 1982. A Remark on Search and Sequencing Problems. *Math. Opns. Res.* **7,** 154–157.

PRABHU, N. U. 1980. *Stochastic Storage Processes.* Springer-Verlag, New York.

RAU, G. 1971. Minimizing a Function of Permutations of *n* Integers. *Opns. Res.* **19,** 237–240.

ROSS, S. M. 1983. *Introduction to Stochastic Dynamic Programming.* Academic Press, San Francisco.

WHITTLE, P. 1981. *Optimization Over Time: Vol. I.* John Wiley, New York.