

ASSESSING THE EFFECTS OF MACHINE BREAKDOWNS IN STOCHASTIC SCHEDULING

John R. BIRGE

Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal, Ann Arbor MI 48109, USA

Kevin D. GLAZEBROOK

Department of Statistics, University of Newcastle upon Tyne, UK

Received January 1988

Revised July 1988

In most scheduling problems discussed in the literature it is assumed that the machine (i.e. key resource) is continuously available. Plainly, this is often unrealistic. Here we suggest assessing the effects of machine breakdowns by evaluating the strategy which is optimal when the machine is always available as a strategy for the breakdowns case. The results extend earlier ones of the authors and co-workers.

alternating renewal process * Gittins * renewal function * stochastic scheduling

1. Introduction

Let $J = \{1, 2, \dots, N\}$ be a set of jobs to be processed on a single machine which is subject to breakdown. For $k = 1, 2, \dots$ the k -th breakdown of the machine is associated with two random variables U_k and D_k , both taking values in the positive integers. U_k is the k -th *uptime* for the machine, i.e. the length of the period between the $(k-1)$ -st and the k -th breakdown. D_k is the k -th machine *downtime*, i.e. the length of the k -th breakdown. All of these random variables are independent of each other, and further the uptimes and downtimes are (separately) identically distributed. This process of breakdowns constitutes an alternating renewal process (see Ross, 1970). We further assume that the uptimes have finite first and second moments, μ_1 and μ_2 , that the downtimes have finite first moment μ_D and that the machine is up from time 0 until U_1 .

The processing requirement of job j is a random variable X_j , taking values in the positive integers, $1 \leq j \leq N$. The X_j 's are independent of each other and of the uptimes and downtimes. At

each non-negative point in time t , the state of the system consists of the machine's condition (up or down), the (non-negative) time the machine has been in that condition, the set of completed jobs from J , and the cumulative processing time up to time t of all unfinished jobs in J . Note that this situation implies that the total processing time for a given job is only available after the job has completed. A decision depending only on the system state must be made at t to determine which uncompleted job to process during $[t, t+1]$. A policy π is any rule for making these decisions until all jobs in J are finished. Under π , job j is completed at time $F_j(\pi)$. The objective is to choose a policy which minimizes *expected weighted flow time*

$$C(\pi) = E \left\{ \sum_{j=1}^N w_j F_j(\pi) \right\} \quad (1)$$

where the w_j 's are given positive constants.

For a machine which is continuously available (i.e. downtimes are zero), any optimal policy is known to be determined by a collection of Gittins' indices (see Gittins, 1979, and Glazebrook, 1984)).

If job j has received x units of processing and has still to complete, its Gittins index is defined to be

$$\gamma_j(x) = \sup_{r \geq 1} \left[\frac{w_j \sum_{t=1}^r \rho_j(x+t-1) [\prod_{s=0}^{t-2} \{1 - \rho_j(x+s)\}]}{\sum_{t=0}^{r-1} [\prod_{s=0}^{t-1} \{1 - \rho_j(x+s)\}]} \right],$$

(2)

where

$$\rho(t) = p(X_j = t + 1 | X_j > t), \quad t = 0, 1, 2, \dots,$$

is the *completion rate function* for job j . Hence in the absence of breakdowns, an optimal policy always allocates the machine to any uncompleted job with maximal Gittins index. Glazebrook (1987) pointed out that this remains optimal for a problem with breakdowns when the uptimes are geometrically distributed.

Our objective is to evaluate the simple 'no breakdowns' optimal policy determined by the Gittins' indices in (2) as a policy for our general problem with breakdowns. For convenience denote by π_γ a policy which, whenever the machine is up, processes the uncompleted job with the largest Gittins index. Denote by π^* an optimal policy. We shall seek to evaluate π_γ by putting upper bounds on the quantities

$$C(\pi_\gamma) - C(\pi^*) \tag{3}$$

or

$$\{C(\pi_\gamma) - C(\pi^*)\} \{C(\pi^*)\}^{-1}. \tag{4}$$

Following on from the last remark in the previous paragraph, a natural approach is to bound the quantities in (3) or (4) by a quantity which measures the extent to which the uptimes fail to be geometric. This work is reported in Section 2 and extends results due to Glazebrook (1987). A more general result in Section 3 extends previous work by Birge, Frenk, Mittenthal and Rinnooy Kan (1987) which was restricted to the case of deterministic processing times.

2. Bounds based on stochastic orderings

We shall need some preliminary remarks and definitions relating to the renewal process $N_U(t)$ determined by the up-times, i.e.

$$N_U(t) \triangleq \sup \{k \geq 0 : U_1 + U_2 + \dots + U_k \leq t\}.$$

The *renewal function* $m_U(t)$ is defined as usual to be $E\{N_U(t)\}$. We shall use the notation $N_V(t)$, $N_W(t)$, $m_V(t)$, $m_W(t)$ to represent renewal processes and functions defined with respect to i.i.d. random variables $\{V_1, V_2, \dots\}$ and $\{W_1, W_2, \dots\}$ respectively. Throughout, all random variables are assumed to be positive integer-valued with finite mean. We make use of the following ideas from reliability theory (see Barlow and Proschan, 1975).

Definition 1. Positive integer-valued random variable V is *new better than used in expectation* (N.B.U.E.) if

$$E(V) \geq E(V - v | V \geq v + 1), \quad v = 0, 1, 2, \dots$$

Definition 2. Positive integer-valued random variable V is *new worse than used in expectation* (N.W.U.E.) if

$$E(V) \leq E(V - v | V \geq v + 1), \quad v = 0, 1, 2, \dots$$

The following result follows fairly simply from Theorem 3.14 in Chapter 6 of Barlow and Proschan (1975). In Lemma 1 ' \leq_{ST} ' denotes stochastic ordering.

Lemma 1. Suppose that there exists a N.B.U.E. random variable V_1 and a N.W.U.E. random variable W_1 such that $V_1 \leq_{ST} U_1 \leq_{ST} W_1$ then

$$\begin{aligned} t/E(V_1) &\geq m_V(t) \geq m_U(t) \\ &\geq m_W(t) \geq t/E(W_1), \quad t \geq 0. \end{aligned}$$

Geometric random variables are both N.B.U.E. and N.W.U.E. and hence Lemma 1 relates to the situation where the uptimes are stochastically bounded above and below by such variates. The following definitions relate to the 'closest' such bounding variables obtainable.

Definition 3. The *upper geometric rate* $\bar{\eta}$ of uptime U_1 is defined as $\bar{\eta} = \inf_{\eta} \{\eta > 0; V_1 \leq_{ST} U_1 \text{ where } V_1 \text{ is geometric with probability } \eta\}$.

Definition 4. The *lower geometric rate* $\underline{\eta}$ of uptime U_1 is defined as $\underline{\eta} = \sup_{\eta} \{\eta > 0; U_1 \leq_{ST} W_1 \text{ where } W_1 \text{ is geometric with probability } \eta\}$.

It is not difficult to show algebraically that if $\eta(\cdot)$ is the completion rate function of uptime U_1 then

$$\begin{aligned} \sup_j \eta(j) &\geq \bar{\eta} = \sup_j \left[1 - \left(\prod_{k=1}^j \{1 - \eta(k)\} \right)^{1/j} \right] \\ &\geq \sup_j \left\{ \sum_{k=1}^j \eta(k) \cdot j^{-1} \right\} \\ &\geq \inf_j \left\{ \sum_{k=1}^j \eta(k) \cdot j^{-1} \right\} \\ &\geq \inf_j \left[1 - \left(\prod_{k=1}^j \{1 - \eta(k)\} \right)^{1/j} \right] \\ &= \underline{\eta} \geq \inf_j \eta(j). \end{aligned} \tag{5}$$

We are now in a position to use the above bounds on the uptimes and the related bounds on renewal functions to give the evaluation of the 'no breakdowns' optimal policy π_γ we seek. Let π be a strategy for our stochastic scheduling problem with breakdowns, which is stationary with regard to the values of past downtimes. Optimal policy π^* plainly has this property. Under π , job j is completed at time $F_j(\pi)$. Denote by $F_j^0(\pi)$ the implied completion time of j under policy π when there are no breakdowns (i.e. U_1 is infinite). The assumptions concerning independence and finiteness of moments imply via a simple conditioning argument that

$$\begin{aligned} C(\pi) &= E \left\{ \sum_{j=1}^N w_j F_j(\pi) \right\} \\ &= E \left[\sum_{j=1}^N w_j (F_j^0(\pi) + m_U \{F_j^0(\pi)\} \mu_D) \right]. \end{aligned} \tag{6}$$

Theorem 1 embodies an evaluation of π_γ in terms of a measure of the extent to which the uptimes fail to be geometric.

Theorem 1.

(a) $C(\pi_\gamma) - C(\pi^*)$

$$\leq (\bar{\eta} - \underline{\eta}) \mu_D E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right],$$

(b) $\{C(\pi_\gamma) - C(\pi^*)\} \{C(\pi^*)\}^{-1}$

$$\leq \{(\bar{\eta} - \underline{\eta}) \mu_D\} \{1 + \underline{\eta} \mu_D\}^{-1}$$

$$\leq \{\bar{\eta}(\underline{\eta})^{-1}\} - 1.$$

Proof. From Lemma 1, Definition 3 and (6) it follows that

$$\begin{aligned} C(\pi_\gamma) &= E \left[\sum_{j=1}^N w_j (F_j^0(\pi_\gamma) + m_U \{F_j^0(\pi_\gamma)\} \mu_D) \right] \\ &\leq (1 + \bar{\eta} \mu_D) E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right]. \end{aligned}$$

Similarly we have that

$$\begin{aligned} C(\pi^*) &\geq (1 + \underline{\eta} \mu_D) E \left[\sum_{j=1}^N w_j F_j^0(\pi^*) \right] \\ &\geq (1 + \underline{\eta} \mu_D) E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right], \end{aligned}$$

the latter inequality following from the optimality of π_γ for the no-breakdowns case. Inequalities (a) and (b) now follow simply.

Note that since $(\bar{\eta} - \underline{\eta})$ and $\{\bar{\eta}(\underline{\eta})^{-1}\} - 1$ are both natural measures of the extent to which uptimes fail to be geometric then in Theorem 1 we have achieved our stated objective—namely the evaluation of π_γ in terms of such measures. It follows fairly simply from Lemma 1 that we can do rather better than Theorem 1 in the case of N.B.U.E. and N.W.U.E. uptime distributions. This is of considerable practical importance since all of the standard discrete distributions have monotonic completion rate functions and hence are either N.B.U.E. or N.W.U.E. Theorem 2 has a proof which is a simple elaboration of that of Theorem 1 and hence it will not be given.

Theorem 2. (i) If uptime U_1 is N.B.U.E. then

(a) $C(\pi_\gamma) - C(\pi^*)$

$$\leq (\mu_1^{-1} - \underline{\eta}) \mu_D E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right],$$

(b) $\{C(\pi_\gamma) - C(\pi^*)\} \{C(\pi^*)\}^{-1}$

$$\leq \{(\mu_1^{-1} - \underline{\eta}) \mu_D\} \{1 + \underline{\eta} \mu_D\}^{-1}$$

$$\leq (\underline{\eta} \mu_1)^{-1} - 1.$$

(ii) If uptime U_1 is N.W.U.E. then

(c) $C(\pi_\gamma) - C(\pi^*)$

$$\leq (\bar{\eta} - \mu_1^{-1})\mu_D E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right],$$

(d) $\{C(\pi_\gamma) - C(\pi^*)\} \{C(\pi^*)\}^{-1}$

$$\leq \{(\bar{\eta} - \mu_1^{-1})\mu_D\} \{1 + \mu_1^{-1}\mu_D\}^{-1}$$

$$\leq \bar{\eta}\mu_1 - 1.$$

3. Bounds based on Lorden's inequality

The bounds described here extend Theorem 2.5 in Birge, Frenk, Mittenthal and Rinnooy Kan (1987) to the case with random processing times. They originate from Lorden's inequality for the renewal function which, under the conditions assumed here, states that

$$\frac{t}{\mu_1} - 1 \leq m_U(t) \leq \frac{t}{\mu_1} + \frac{\mu_2}{\mu_1^2} - 1, \quad t \geq 0. \quad (7)$$

Theorem 3.

$$C(\pi_\gamma) - C(\pi^*) \leq \mu_2 \mu_D \mu_1^{-2} \left(\sum_{j=1}^N w_j \right).$$

Proof. From the right-hand side of (7), together with (6), it follows that

$$C(\pi_\gamma) = E \left[\sum_{j=1}^N w_j (F_j^0(\pi_\gamma) + m_U \{F_j^0(\pi_\gamma)\} \mu_D) \right]$$

$$\leq E \left[\sum_{j=1}^N w_j \left(F_j^0(\pi_\gamma) + \mu_D \left\{ \frac{F_j^0(\pi_\gamma)}{\mu_1} + \frac{\mu_2}{\mu_1^2} - 1 \right\} \right) \right]$$

$$= (1 + \mu_D \mu_1^{-1}) E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right]$$

$$+ (\mu_2 \mu_D \mu_1^{-2} - \mu_D) \left(\sum_{j=1}^N w_j \right).$$

Similarly, from the left-hand side of (7), we have that

$$C(\pi^*) \geq (1 + \mu_D \mu_1^{-1}) E \left[\sum_{j=1}^N w_j F_j^0(\pi^*) \right]$$

$$- \mu_D \left(\sum_{j=1}^N w_j \right)$$

$$\geq (1 + \mu_D \mu_1^{-1}) E \left[\sum_{j=1}^N w_j F_j^0(\pi_\gamma) \right]$$

$$- \mu_D \left(\sum_{j=1}^N w_j \right),$$

the latter inequality following from the optimality of π_γ for the no-breakdowns case. The result now follows.

4. Comparing the bounds

Each of the bounds given in Theorems 1 to 3 may be applicable in a practical problem, depending on the available information and the problem context. Theorem 1 is most applicable when the time between machine failures follows the common empirical observation of a 'bathtub' distribution that has an upper bounding geometric probability either during infant mortality or after long service and has a lower bounding geometric probability at some time between these extremes. Theorem 2 applies when either the infancy or aging effect is dominant so that monotone completion rates result. Theorem 3 applies best when only two moments of the distribution are known or when the value $E[\sum_{j=1}^N w_j F_j^0(\pi_\gamma)]$ is large in comparison with $\sum_{j=1}^N w_j$.

As an example, suppose the uptime follows a negative binomial distribution on the positive integers. This distribution is N.B.U.E. and has readily available upper and lower bounding geometric rates. If this distribution corresponds to the number of failures (plus one) until $r = 2$ successes with probability $p = \frac{1}{2}$ of success, then $\mu_1 = 3$, $\mu_2 = 13$, $\bar{\eta} = \frac{1}{2}$ and $\eta = \frac{1}{4}$. The bound from Theorem 1 on the relative error from ignoring breakdowns is 1.0. The N.B.U.E. information allows us to halve this bound to 0.5 using Theorem 2. If we assume that $\mu_D = \sum_{j=1}^N w_j = 1$, $E[\sum_{j=1}^N w_j F_j^0(\pi_\gamma)] = 10$, then the relative error using Theorem 3 is 0.15.

References

- R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York, 1975.
- J.R. Birge, J.B.G. Frenk, J. Mittenthal and A.H.G. Rinnooy Kan, "Single machine scheduling subject to stochastic breakdowns" (unpublished m.s.), 1987.
- J.C. Gittins, "Bandit processes and dynamic allocation indices" (with discussion), *J. Roy. Statist. Soc. B* 41, 148-177 (1979).
- K.D. Glazebrook, "Scheduling stochastic jobs on a single machine subject to breakdowns", *Naval Res. Logist. Quart.* 31, 251-264 (1984).
- K.D. Glazebrook, "Evaluating the effects of machine breakdowns in stochastic scheduling problems", *Naval Res. Logist. Quart.* 34, 319-335 (1987).
- S.M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, CA, 1970.