

SCHEDULING JOBS WITH STOCHASTICALLY ORDERED PROCESSING TIMES ON PARALLEL MACHINES TO MINIMIZE EXPECTED FLOWTIME

R. R. WEBER,* *University of Cambridge*
P. VARAIYA,** *University of California, Berkeley*
J. WALRAND,** *University of California, Berkeley*

Abstract

A number of jobs are to be processed using a number of identical machines which operate in parallel. The processing times of the jobs are stochastic, but have known distributions which are stochastically ordered. A reward $r(t)$ is acquired when a job is completed at time t . The function $r(t)$ is assumed to be convex and decreasing in t . It is shown that within the class of non-preemptive scheduling strategies the strategy SEPT maximizes the expected total reward. This strategy is one which whenever a machine becomes available starts processing the remaining job with the shortest expected processing time. In particular, for $r(t) = -t$, this strategy minimizes the expected flowtime.

STOCHASTIC SCHEDULING

1. Introduction

The processing times of n jobs are random variables which are stochastically ordered as $X_1 \leq_{st} \dots \leq_{st} X_n$. The jobs are to be processed *non-preemptively* using m machines which operate in parallel. Suppose that machine i becomes available at time $\tau_i \geq 0$. Let τ be the vector of times $(\tau_1, \tau_2, \dots, \tau_m)$ at which the machines become available. The aim is to find the scheduling strategy which, amongst non-preemptive strategies S , achieves the maximum reward,

$$R(\tau) = \sup_S E \left[\sum_{k=1}^n r(T_k) \right].$$

Here $r(t)$ is a convex, decreasing function of t , $(0 \leq t < \infty)$. T_k denotes the time at

Received 1 August 1984; revision received 19 June 1985.

* Postal address: Department of Engineering, Control and Management Systems Division, Mill Lane, Cambridge, CB2 1RX, UK.

** Postal address: Dept. of Electrical Engineering and Computer Sciences and Electronic Research Laboratory, University of California, Berkeley, CA 94720, USA.

Research supported by the Office of Naval Research Contract N00014-80-C-0507.

which job k is completed. Note that for $r(t) = -t$ the problem is one of minimizing the expected flowtime (sum of job completion times).

When the processing times are almost surely ordered, $X_1 \leq \dots \leq X_n$, then the optimal schedule is the one which as machines become available starts the jobs in the order $1, 2, \dots, n$ (see Conway et al. (1967)). This strategy is usually called SEPT (shortest expected processing time first). Glazebrook (1979) and Weiss and Pinedo (1980) have shown that SEPT minimizes expected flowtime when the jobs have processing times which are exponentially distributed with different means. Weber (1980), (1982) has shown that SEPT also minimizes expected flowtime for a more general model in which job i has a processing time distributed with distribution function $F_i(t) = \{F(t + t_i) - F(t)\} / \{1 - F(t)\}$, where $t_1 \geq \dots \geq t_n$ and $F(t)$ is a distribution function with an increasing hazard rate, $\rho(t) = f(t) / \{1 - F(t)\}$. In this model the jobs are essentially identical, but they have received different amounts of processing, t_1, \dots, t_n , prior to the start. In this paper we show that SEPT is optimal for a more general model, which encompasses those mentioned above. The only assumption is that the processing times are stochastically ordered.

2. Optimality of non-preemptive SEPT

Of particular interest are *list scheduling strategies*, which start processing the jobs in a predetermined order. Without confusion we can let $L = (k_1, \dots, k_n)$ denote both a permuted listing of $(1, 2, \dots, n)$ and the list scheduling strategy which starts jobs $1, \dots, n$ in the order k_1, \dots, k_n . Let $R(\tau; L)$ denote the expected reward obtained when the jobs in list L are processed according to list scheduling strategy L . Without loss of generality we suppose that $\tau_1 \leq \tau_2 \leq \dots \leq \tau_m$. For convenience we suppose that $r(t)$ is twice differentiable and that the processing times are continuous random variables which have density functions.

We approach the proof of the result through three lemmas. The first states that for a list scheduling strategy L the rate of change of the expected reward with respect to the starting time of any machine is just the expected reward obtained on that machine when the reward function is altered to $\dot{r}(t)$, the derivative of $r(t)$ with respect to t . Let $R_i(\tau; L)$ denote the expected reward obtained on machine i when the list scheduling strategy L is employed and the reward function is $\dot{r}(t)$. To define this without ambiguity, we adopt the convention that if machine i and one or more other machines become available at the same time then machine i is assigned a job (if any) from the list L only once all other machines becoming available at that time have been assigned jobs coming earlier in the list. In other words, the assignment is made by pretending that machine i becomes available at a slightly later time. Let $dR(\tau, L)/d\tau_i$ denote the right-hand derivative of $R(\tau, L)$ with respect to τ_i .

Lemma 1. For any list scheduling strategy L , $dR(\tau, L)/d\tau_i$ exists and

$$dR(\tau; L)/d\tau_i = R_i(\tau; L), \quad i = 1, \dots, m.$$

Proof. The proof is by induction on n . It is clearly true for $n = 0$. Suppose it is true for fewer than n jobs. Suppose $L = (i_1, i_2, \dots, i_n)$ (here not necessarily SEPT ordered). Let $f(t)$ be the density function of X_{i_1} and let $L_1 = (i_2, i_3, \dots, i_n)$. Then

$$R(\tau, L) = \int_0^\infty f(t)\{r(\tau_1 + t) + R(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\}dt.$$

Differentiating and using the inductive hypothesis,

$$dR(\tau; L)/d\tau_1 = \int_0^\infty f(t)\{\dot{r}(\tau_1 + t) + R_1(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\}dt = R_1(\tau; L).$$

Similarly,

$$dR(\tau, L)/d\tau_i = \int_0^\infty f(t)\{R_i(\tau_1 + t, \tau_2, \dots, \tau_m; L_1)\}dt = R_i(\tau; L), \quad i \neq 1.$$

This completes the inductive step and proof of the lemma.

The next lemma states that when the reward function is $\dot{r}(t)$ and the scheduling strategy is SEPT, then the expected reward obtained on a given machine is not decreased if that machine is made to start later and is not increased if any other machine is made to start later.

Lemma 2.

(a) Suppose L is the SEPT list $(1, 2, \dots, n)$. Then for $j \neq i$ and $n \geq 1$,

$$R_i(\tau; L) \text{ is non-decreasing in } \tau_i \text{ and non-increasing in } \tau_j.$$

(b) Suppose L^* is the list $(2, 3, \dots, n)$, omitting some $k \geq 2$. Then for $n \geq 2$,

$$E[R_1(\tau_1 + X_1, \tau_2 + X_k, \dots, \tau_m; L^*) - R_1(\tau_1 + X_k, \tau_2 + X_1, \dots, \tau_m; L^*)] \leq 0.$$

Proof. Again, the proof is by induction on n . Part (a) is trivial for $n = 1$; part (b) is trivial for $n = 2$. Suppose that the lemma is true when there are fewer than n jobs to process. We show that it is true when there are n jobs to process. The inductive step for (b) follows from the inductive hypothesis for (a) when there are $n - 2$ jobs to process and the fact that if a function $h(x_1, x_2) = R_1(\tau_1 + x_1, \tau_2 + x_2, \dots, \tau_m; L_2)$ is non-decreasing in x_1 and non-increasing in x_2 , then $E[h(X_1, X_2) - h(X_2, X_1)]$ is non-positive for $X_1 \leq_{st} X_2$.

To check the inductive step for (a) we begin by showing that $R_i(\tau, L)$ is non-decreasing in τ_i . Let $i = 1$ (without loss of generality) and suppose $\tau_2 \leq \dots \leq \tau_m$. Let $L_1 = (2, 3, \dots, n)$. Then for $\tau_1 < \tau_2$,

$$R_1(\tau; L) = E[\dot{r}(\tau_1 + X_1) + R_1(\tau_1 + X_1, \tau_2, \dots, \tau_m; L_1)].$$

It follows from $\dot{r}(t)$ non-decreasing and the inductive hypothesis that the expression over which the expectation is taken is non-decreasing in τ_1 . Thus within the region $\tau_1 < \tau_2$, it follows that $R_1(\tau; L)$ is non-decreasing in τ_1 . Similar observations apply in the region $\tau_1 > \tau_2$, where

$$R_1(\tau; L) = E[R_1(\tau_1, \tau_2 + X_1, \dots, \tau_m; L_1)].$$

It remains to consider the change in $R_1(\tau; L)$ as τ_1 passes through the value τ_2 . Suppose $\tau_2 = \dots = \tau_k < \tau_{k+1} \leq \dots \leq \tau_m$ and let L^* be the list $(2, 3, \dots, n)$, omitting job k . Then the change in $R_1(\tau; L)$ as τ_1 passes through the value τ_2 may be written as

$$R_1(\tau_2^+, \tau_2, \dots, \tau_m; L) - R_1(\tau_2^-, \tau_2, \dots, \tau_m; L).$$

For $k \leq n$ this change equals

$$E[\dot{r}(\tau_2 + X_k) + R_1(\tau_2 + X_k, \tau_2 + X_1, \tau_3, \dots, \tau_m; L^*) - \dot{r}(\tau_2 + X_1) - R_1(\tau_2 + X_1, \tau_2 + X_k, \tau_3, \dots, \tau_m; L^*)],$$

and for $k > n$ it equals

$$E[-\dot{r}(\tau_2 + X_1)].$$

In both these cases it follows from $\dot{r}(t)$ being negative and non-decreasing and the inductive hypothesis for part (b) that the expression over which the expectation is taken is non-negative. This completes the inductive step showing that $R_i(\tau; L)$ is non-decreasing in τ_i . Similar arguments (which we omit) establish that $R_i(\tau; L)$ is non-increasing in $\tau_j, j \neq i$.

Although it is not used in the proof of the theorem, the following interesting fact is an immediate corollary of Lemma 2.

Corollary. Suppose L is the SEPT list strategy. Then $R(\tau, L)$ is non-increasing and convex in each τ_i .

The final lemma states that when the reward function is $\dot{r}(t)$ then the expected reward obtained on machine 1 (the machine which starts first) is not greater when employing SEPT than when employing a strategy which schedules the shortest job as the first job on machine 2 (the machine which starts second) and schedules the remaining jobs according to SEPT.

Lemma 3. Suppose L is the SEPT list $(1, 2, \dots, n)$. Let L_1 be $(2, 3, \dots, n)$. Then for $n \geq 2$,

$$R_1(\tau; L) \leq E[R_1(\tau_1, \tau_2 + X_1, \dots, \tau_m; L_1)].$$

Proof. The proof is by induction on n . When $n = 2$ we have

$$R_1(\tau; L) \leq E[\dot{r}(\tau_1 + X_1)] \leq E[\dot{r}(\tau_1 + X_2)] = E[R_1(\tau_1, \tau_2 + X_1, \dots, \tau_m; (1))].$$

Suppose that the lemma is true when there are fewer than n jobs to process. Let L_2 be $(3, 4, \dots, n)$. If $\tau_1 = \tau_2$ then the lemma is true with equality. If $\tau_1 < \tau_2$ then we have

$$\begin{aligned} R_1(\tau; L) &= E[\dot{r}(\tau_1 + X_1) + R_1(\tau_1 + X_1, \tau_2, \dots, \tau_m; L_1)] \\ &\cong E[\dot{r}(\tau_1 + X_1) + R_1(\tau_1 + X_1, \tau_2 + X_2, \dots, \tau_m; L_2)] \\ &\leq E[\dot{r}(\tau_1 + X_2) + R_1(\tau_1 + X_2, \tau_2 + X_1, \dots, \tau_m; L_2)] \\ &= E[R_1(\tau_1, \tau_2 + X_1, \dots, \tau_m; L_1)]. \end{aligned}$$

The first inequality follows by the inductive hypothesis. The second inequality follows from $\dot{r}(t)$ non-decreasing and part (b) of Lemma 2. This completes the inductive step and proof of the lemma.

We are now ready to prove the theorem.

Theorem. Suppose n jobs have processing times which are stochastically ordered. Then the non-preemptive scheduling strategy SEPT maximizes the expected reward within the class of non-preemptive strategies. That is, when L is the SEPT ordered list, $L = (1, 2, \dots, n)$, then $R(\tau) = R(\tau; L)$.

Proof. The proof is by induction on n . The result is true trivially for $n = 1$. Suppose that the result is true when there are fewer than n jobs to process. Consider a scheduling strategy S , which begins by processing job k ($k > 1$) on machine 1 (the first machine to become available). By the inductive hypothesis it must be optimal to start job 1 next and then start the remaining jobs according to the SEPT list strategy L^* , where L^* is $(2, 3, \dots, n)$, omitting job k . Thus amongst strategies which start processing job k first, the best is the list strategy resulting from the concatenation of $(k, 1)$ and L^* , which we denote by $L^{k,1} = (k, 1) + L^*$. We shall shortly show that the list strategy $L^{1,k} = (1, k) + L^*$ is better than $L^{k,1}$. Assuming this, it follows by the inductive hypothesis that $L = (1, 2, \dots, n)$ is a better strategy than $L^{1,k}$, and the inductive step is complete. We need only show

$$\Delta = R(\tau; L^{1,k}) - R(\tau; L^{k,1}) \geq 0.$$

Let $R(\tau; S; c)$ be the expected reward using strategy S , conditional on $X_k = c$. We shall shortly show that the quantity defined as

$$\Delta(c) = R(\tau; L^{1,k}, c) - R(\tau; L^{k,1}; c)$$

is non-decreasing in c . Assuming this is so, we have

$$\Delta(X_k) \geq_{st} \Delta(\bar{X}_1),$$

where \bar{X}_1 is a random variable independent of X_1, X_2, \dots, X_n and identically distributed to X_1 . By taking the expected value we have

$$\Delta = E[\Delta(X_k)] \geq E[\Delta(\bar{X}_1)] = 0,$$

where the equality to 0 follows from the fact that \bar{X}_1 and X_1 are identically distributed. The theorem is therefore proved once we show that $\Delta(c)$ is non-decreasing in c . Now

$$\begin{aligned} R(\tau; L^{1,k}; c) &= E[1(X_1 < \tau_2 - \tau_1)\{r(\tau_1 + X_1) + r(\tau_1 + X_1 + c) \\ &\quad + R(\tau_1 + X_1 + c, \tau_2, \dots, \tau_m; L^*)\} \\ &\quad + 1(X_1 \geq \tau_2 - \tau_1)\{r(\tau_1 + X_1) + r(\tau_2 + c) \\ &\quad + R(\tau_1 + X_1, \tau_2 + c, \dots, \tau_m; L^*)\}] \end{aligned}$$

and

$$R(\tau; L^{k,1}; c) = r(\tau_1 + c) + R(\tau_1 + c, \tau_2, \dots, \tau_m; (1) + L^*).$$

Differentiation of the above gives

$$\begin{aligned} dR(\tau; L^{1,k}; c)/dc &= E[1(X_1 < \tau_2 - \tau_1)\{\dot{r}(\tau_1 + X_1 + c) \\ &\quad + R_1(\tau_1 + X_1 + c, \tau_2, \dots, \tau_m; L^*)\} \\ &\quad + 1(X_1 \geq \tau_2 - \tau_1)\{\dot{r}(\tau_2 + c) \\ &\quad + R_2(\tau_1 + X_1, \tau_2 + c, \dots, \tau_m; L^*)\}] \\ &= E[1(X_1 < \tau_2 - \tau_1)\{\dot{r}(\tau_1 + X_1 + c) \\ &\quad + R_1(\tau_1 + X_1 + c, \tau_2, \dots, \tau_m; L^*)\} \\ &\quad + 1(X_1 \geq \tau_2 - \tau_1)\{\dot{r}(\tau_2 + c) + R_1(\tau_2 + c, \tau_1 + X_1, \dots, \tau_m; L^*)\}] \end{aligned}$$

and

$$\begin{aligned} dR(\tau; L^{k,1}; c)/dc &= \dot{r}(\tau_1 + c) + R_1(\tau_1 + c, \tau_2, \dots, \tau_m; (1) + L^*) \\ &\leq E[\dot{r}(\tau_1 + c) + R_1(\tau_1 + c, \tau_2 + X_1, \dots, \tau_m; L^*)], \end{aligned}$$

where the inequality follows from Lemma 3. Thus

$$\begin{aligned} d\Delta(c)/dc &\geq E[1(X_1 < \tau_2 - \tau_1)\{\dot{r}(\tau_1 + X_1 + c) + R_1(\tau_1 + X_1 + c, \tau_2, \dots, \tau_m; L^*)\} \\ &\quad + 1(X_1 \geq \tau_2 - \tau_1)\{\dot{r}(\tau_2 + c) + R_1(\tau_2 + c, \tau_1 + X_1, \dots, \tau_m; L^*)\} \\ &\quad - \{\dot{r}(\tau_1 + c) + R_1(\tau_1 + c, \tau_2 + X_1, \dots, \tau_m; L^*)\}]. \end{aligned}$$

Using part (a) of Lemma 2, $\tau_1 \leq \tau_2$ and $\dot{r}(t)$ non-decreasing, it is easy to check that the expression over which the above expectation is taken is non-negative. This completes the proof of the theorem.

3. Discussion

We have shown that when the jobs have processing times which are stochastically ordered, then the non-preemptive SEPT strategy maximizes the

expected reward within the class of non-preemptive strategies. Examining the proof of Theorem 1, particularly Lemma 3, it can be seen that the result is still true for some models in which the reward obtained on completing each job differs from job to job. The reward obtained upon completing job i may be generalized to any convex, decreasing function $r_i(t)$, provided that when job i is stochastically shorter than job j the inequality $r_i(t) \leq r_j(t)$ holds for all t . We remark that (except for special models like those described in Section 1) the non-preemptive SEPT strategy does not maximize the expected reward within the class of preemptive strategies.

Acknowledgement

We should like to thank Mr Randy Cieslak and Professor Srikanta Kumar for contributing to discussions of the conjecture proved in this paper.

References

- CONWAY, R. W., MAXWELL, W. L. AND MILLER, L. W. (1967) *The Theory of Scheduling*. Addison-Wesley, Reading, MA.
- GLAZEBROOK, K. D. (1979) Scheduling tasks with exponential service times on parallel processors. *J. Appl. Prob.* **16**, 685–689.
- WEBER, R. R. (1980) Optimal Organization of Multi-server Systems. Ph.D. Dissertation. Cambridge University.
- WEBER, R. R. (1982) Scheduling jobs with stochastic processing requirements on parallel machines to minimize makespan or flowtime. *J. Appl. Prob.* **19**, 167–182.
- WEISS, G. AND PINEDO, M. (1980) Scheduling tasks with exponential service times on non-identical processors to minimize various cost functions. *J. Appl. Prob.* **17**, 187–202.