

## ON A CONJECTURE BY COFFMAN, FLATTO, AND WRIGHT ON STOCHASTIC MACHINE MINIMIZATION\*

NAH-OAK SONG<sup>†</sup> AND DEMOSTHENIS TENEKETZIS<sup>‡</sup>

**Abstract.** We investigate a conjecture stated by Coffman, Flatto, and Wright within the context of a stochastic machine minimization problem with a hard deadline. We prove that the conjecture is true.

**Key words.** stochastic allocation, hard deadlines, machine minimization, Bellman equation, time thresholds

**AMS subject classifications.** 93E20, 93E03, 90C39

**PII.** S0097539797324643

**1. Introduction.** Problems of optimal stochastic allocation of machines under waiting-time constraints have recently received considerable attention, as they are important in the design of computer and communication networks and in stochastic real-time scheduling problems (see [1], [2], [3]). From a theoretical point of view these problems are complementary to makespan minimization problems (see, for example, [4], [6], [7], [8]).

The general class of problems of optimal stochastic allocation of machines under waiting-time constraints (also known as stochastic machine minimization problems) can be formulated as follows: There are  $N$  jobs with processing times  $T_1, T_2, \dots, T_N$ , deterministic or random with known distributions. Job waiting times are bounded by a time  $W$  that is independent of  $T_1, T_2, \dots, T_N$ , and may be deterministic or random with a known distribution. There is an unlimited number of machines (processors) initially available to process these jobs. At time 0, a timer is started with initial value  $W$  and job scheduling begins. When the timer expires, all jobs not running at that time and still waiting to be processed are assigned to available machines. The objective is to determine, within the class of nonpreemptive policies, a policy that minimizes the expected cost, with cost defined as the number of distinct machines used throughout the schedule.

Several cases within the above class of problems have been considered so far in the literature [1], [2], [3]. A problem of practical importance is the hard deadline case where  $W$  is a constant. A partial analysis of this problem has been presented in [1], where the structure of an optimal policy was proved under a conjecture that remains so far an open problem. In this paper we prove that the conjecture stated in [1] is true; consequently, the structure of the optimal policy proposed in [1] is correct.

The paper is organized as follows: the conjecture stated in [1] is precisely formulated in section 2; an outline of the conjecture's proof is given in section 3 and

---

\*Received by the editors July 23, 1997; accepted for publication (in revised form) January 5, 2000; published electronically June 27, 2000. This research was supported in part by NSF grant NCR-9204419, AFOSR grant F49620-96-1-0028, ARO grant DAAH04-96-1-0377, and NSF grant ECS-9979347.

<http://www.siam.org/journals/sicomp/30-2/32464.html>

<sup>†</sup>Samsung Electronics Co., Ltd, 11th Floor, Samsung Plaza Bldg., 263, Seohyeon-Dong, Bundang-Gu, Sungnam-Si, Kyungki-Do, Korea 463-050 (sno@telecom.samsung.co.kr). This author's work was also supported in part by a Barbour Scholarship from the University of Michigan.

<sup>‡</sup>Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109-2122 (teneketzis@eecs.umich.edu).

the technical details of the proof appear in section 4. Discussion of the results and suggestions for further research appear in section 5.

**2. The conjecture.** Coffman, Flatto, and Wright [1] considered a stochastic machine minimization problem (described in section 1) where the job processing times  $T_1, T_2, \dots, T_N$  are independent samples of an exponentially distributed random variable  $T$  with  $E(T) = 1$ , and the job waiting times are bounded by a constant  $W$ .

For this particular stochastic machine minimization problem it is possible to define a Markov process on the set of states  $(n, k, s)$ , where  $n$  is the number of unfinished jobs,  $k$  is the number of jobs currently assigned to processors, and  $s$  is the time remaining on the timer (see [1]). Let  $V(n, k, s)$  denote the expected cost incurred by an optimal allocation policy when the initial state is  $(n, k, s)$ , and define

$$(1) \quad V(n, s) = \min_{1 \leq k \leq n} V(n, k, s)$$

as the expected cost incurred by an optimal allocation policy when there are  $n$  jobs to be processed with  $s$  units of time remaining on the timer, and there is an unlimited number of machines initially available.  $V(n, k, s)$  can be computed by the Bellman equation,

$$(2) \quad V(n, k, s) = \inf_{0 \leq t \leq s} \left\{ e^{-k(s-t)} V(n, k+1, t) + \int_0^{s-t} k e^{-ku} V(n-1, k, s-u) du \right\}, \quad k < n,$$

$$(3) \quad V(n, n, s) = n.$$

Coffman, Flatto, and Wright [1] proved that  $V(n, k, s)$  has the following properties.

LEMMA 2.1. (i)  $V(n, k, 0) = n$ ,  $V(n, k, \infty) = k$ . (ii)  $V(n, k, s)$  is nondecreasing in  $k$  for fixed  $n$  and  $s$ , strictly increasing in  $n$  for fixed  $k$  and  $s$ , and strictly decreasing in  $s$  for fixed  $n$  and  $k < n$ .

Furthermore, the following assertion was made in [1].

CONJECTURE. For each  $n > 1$ , there exist nonnegative numbers  $l_{nk}$ ,  $1 \leq k \leq n-1$ , such that

$$(4) \quad V(n, k+1, s) > V(n, k, s) \quad \text{for } s > l_{nk},$$

$$(5) \quad V(n, k+1, s) = V(n, k, s) \quad \text{for } s \leq l_{nk},$$

$$(6) \quad l_{n(n-1)} \leq l_{n(n-2)} \leq \dots \leq l_{n1}.$$

Based on Lemma 2.1 and the above conjecture, Coffman, Flatto, and Wright proved the following result in [1].

THEOREM 2.2. For all  $k$ ,  $1 \leq k \leq n-2$ ,

$$(7) \quad l_{(n-1)k} < l_{nk}.$$

The above theorem and the conjecture imply the following structure of an optimal policy.

**THEOREM 2.3** (see [1, Theorem 5.3]). *Let the initial state have  $n > 1$  jobs to be scheduled, an unlimited number of machines available, and a clock time  $s$  such that  $l_{nk} < s \leq l_{n(k-1)}$  for some  $1 \leq k \leq n - 1$ . Then assign jobs to  $k$  machines and start running jobs in state  $(n, k, s)$ . Continue running jobs in any given state until one of the following three events occurs: (1) The clock expires; in this case assign the remaining  $n - k \geq 0$  waiting jobs to new machines. (2) A machine completes a job in some state  $(n', k', s')$ ,  $k' < n'$ ; in this case replenish the machine and continue in state  $(n' - 1, k', s')$ . (3) A state  $(n', k', s')$ ,  $k' < n'$ , is reached in which the remaining time has reduced to  $s' = l_{n'k'}$ ; in this case assign a waiting job to a new machine and continue in state  $(n', k' + 1, s')$ .*

In the remainder of this paper, we prove that the above stated conjecture is true, thus completing the proof of the optimality of the policy described in Theorem 2.3. We proceed as follows: First we briefly outline the main ideas of the proof and then we present all the technical details.

**3. Outline of the proof of the conjecture.** We establish, via the Bellman equation, the existence of “time thresholds”  $l_{nk}, n = 1, 2, \dots$  and  $k = 1, 2, \dots, n - 1$  that have the following features:

- (F1) When the initial state is  $(n, k, s)$ ,  $s > l_{nk}$ , then along sample path realizations where there are no job completions it is optimal not to assign any new (i.e., previously unused) machines to jobs waiting to be processed as long as the time remaining in the timer is strictly larger than  $l_{nk}$ . (Along the same sample path realizations it is optimal to assign a new machine to a job waiting to be processed when the time remaining in the timer is equal to  $l_{nk}$ .)
- (F2) If the initial state is  $(n, k, t)$  and  $t \leq l_{nk}$ , then it is optimal to assign right away at least one new machine to a job waiting to be processed.
- (F3) For any  $n, n = 1, 2, \dots$  and  $k = 1, 2, \dots, n - 1$  the “time thresholds”  $l_{nk}$  are ordered as follows:

$$l_{n(n-1)} \leq l_{n(n-2)} \leq \dots \leq l_{n1}.$$

To establish (F1) we use the Bellman equation and consider an initial state  $(n, k, s)$  with the following characteristic: The earliest time  $l_{nks}$  (according to the Bellman equation) at which it is optimal to add a new machine along sample path realizations where there are no job completions is such that  $l_{nks} < s$ . We prove that for all states  $(n, k, s')$  such that  $s' > l_{nks}$  we have  $l_{nks} = l_{nks'} := l_{nk}$ . This result and the definition of  $l_{nks}$  (see (8)–(9)) lead to the proof of the first part of the conjecture (see (4)).

To establish (F2) we consider an instance where the initial state is  $(n, k, s)$ ,  $s > l_{nk}$ , and an optimal allocation policy satisfying the Bellman equation is used. We show that along sample path realizations where there are no job completions until  $t$  units of time remain on the timer ( $t < l_{nk}$  by assumption) it is optimal to have at least  $k + 1$  jobs under processing at  $t$ . This implies that if the initial state is  $(n, k, t)$  and an optimal allocation policy (satisfying the Bellman equation) is used, at least one new machine must be allocated to a job waiting to be processed at  $t$ . This feature together with a property of the cost function  $V(n, k, s)$ , described by Lemma 2.1(ii), leads to the proof of the second part of the conjecture (Eq. (5)).

To establish (F3) we use the first and second parts of the conjecture (already proved), a property of the cost function  $V(n, k, s)$  described by Lemma 2.1 (ii), the Bellman equation, and a contradiction argument. Feature (F3) of the “time thresholds”  $l_{nk}$  describes the last part of the conjecture (see (6)).

Thus, the validity of the conjecture is established.

**4. Proof of the conjecture.** Consider the Bellman equation and for each state  $(n, k, s)$  define

$$A_{n,k}^s := \left\{ l : l = \arg \inf_{0 \leq t \leq s} \left[ e^{-k(s-t)} V(n, k+1, t) + \int_0^{s-t} k e^{-ku} V(n-1, k, s-u) du \right] \right\} \tag{8}$$

and

$$l_{nks} := \max \{ l : l \in A_{n,k}^s \}. \tag{9}$$

Let  $\pi^*$  denote the allocation policy that satisfies the Bellman equation and has the following characteristic: Along sample paths that originate at any state  $(n, k, s)$ , with  $n > k$ ,  $s > l_{nks}$ , and have no job completions until  $l_{nks}$  units of time remain on the timer,  $\pi^*$  adds a new machine at that point.

*Proof of (4).* Consider an initial state  $(n, k, s)$  such that  $s > l_{nks}$ . We prove that for all  $s' > l_{nks}$  we have

$$l_{nks} = l_{nks'} := l_{nk}. \tag{10}$$

This fact together with the definition of  $l_{nks}$  leads to the proof of the first part of the conjecture, namely, (4). To prove (10) we proceed in two steps.

*Step (i).* Take  $s'$  such that  $l_{nks} < s' < s$ . We prove that  $A_{n,k}^s = A_{n,k}^{s'}$ . Pick  $l \in A_{n,k}^{s'}$ ; then

$$V(n, k, s') = e^{-k(s'-l)} V(n, k+1, l) + \int_0^{s'-l} k e^{-ku} V(n-1, k, s'-u) du. \tag{11}$$

Furthermore, since  $s > s' > l_{nks}$ ,

$$V(n, k, s) = e^{-k(s-s')} V(n, k, s') + \int_0^{s-s'} k e^{-ku} V(n-1, k, s-u) du. \tag{12}$$

Substituting (11) into (12), we get

$$V(n, k, s) = e^{-k(s-s')} \left[ e^{-k(s'-l)} V(n, k+1, l) + \int_0^{s'-l} k e^{-ku} V(n-1, k, s'-u) du \right] + \int_0^{s-s'} k e^{-ku} V(n-1, k, s-u) du. \tag{13}$$

Letting

$$v = s - s' + u \tag{14}$$

in the second term of (13), we obtain

$$V(n, k, s) = e^{-k(s-l)} V(n, k+1, l) + \int_0^{s-l} k e^{-kv} V(n-1, k, s-v) dv. \tag{15}$$

Equation (15) proves that  $l \in A_{n,k}^s$ . Hence,

$$A_{n,k}^{s'} \subset A_{n,k}^s. \tag{16}$$

Next, take  $l \in A_{n,k}^s$ ; suppose that  $l \notin A_{n,k}^{s'}$ . Then, since by assumption  $s > s' > l_{nks} \geq l$ , it follows that

$$\begin{aligned}
 V(n, k, s) &= e^{-k(s-s')}V(n, k, s') + \int_0^{s-s'} ke^{-ku}V(n-1, k, s-u)du \\
 &< e^{-k(s-s')} \left[ e^{-k(s'-l)}V(n, k+1, l) + \int_0^{s'-l} ke^{-ku}V(n-1, k, s'-u)du \right] \\
 &\quad + \int_0^{s-s'} ke^{-ku}V(n-1, k, s-u)du \\
 (17) \quad &= e^{-k(s-l)}V(n, k+1, l) + \int_0^{s-l} ke^{-ku}V(n-1, k, s-u)du;
 \end{aligned}$$

the inequality in (17) results because  $l \notin A_{n,k}^{s'}$ . According to (17),  $l \notin A_{n,k}^s$  and this is a contradiction. Hence,  $l \in A_{n,k}^{s'}$ ; therefore,

$$(18) \quad A_{n,k}^s \subset A_{n,k}^{s'}$$

From (16) and (18), we conclude that

$$(19) \quad A_{n,k}^s = A_{n,k}^{s'} \quad \text{for all } s' \text{ such that } l_{nks} < s' < s.$$

*Step (ii).* Consider  $s'$  such that  $l_{nks} < s < s'$ . We prove that there is no  $\hat{\ell} \in A_{n,k}^{s'}$  such that  $\hat{\ell} \geq s$ . The proof is by contradiction.

Suppose there exists  $\hat{\ell} \in A_{n,k}^{s'}$  such that  $\hat{\ell} \geq s$ ; then  $l_{nks'} \geq s$ . Start at  $(n, k, s')$  and use the optimal allocation policy  $\pi^*$ . Let  $\Omega'$  be the set of sample paths that start at  $(n, k, s')$  and along which there are no job completions until  $s$  units of time remain on the timer. Since  $l_{nks'} \geq s$ , the expected cost incurred by the optimal policy  $\pi^*$  along  $\Omega'$  is  $V(n, k + \hat{k}, s)$ , where  $\hat{k} \geq 1$ . Furthermore, since  $\hat{k}$  machines are added by the optimal policy  $\pi^*$  along  $\Omega'$ , we conclude that when there are  $n$  unfinished jobs and  $s$  units of time remain on the timer it is optimal to use  $k + \hat{k}$  machines ( $\hat{k} \geq 1$ ). That is,

$$(20) \quad V(n, s) = V(n, k + \hat{k}, s).$$

On the other hand, since  $s > l_{nks}$ , (1), Lemma 2.1(ii), the Bellman equation, and the definition of  $l_{nks}$  imply that

$$(21) \quad V(n, s) < V(n, k + 1, s) \leq V(n, k + \hat{k}, s),$$

which, in turn, implies that when there are  $n$  unfinished jobs and  $s$  units of time remain on the timer it is optimal to use less than  $k + 1$  machines. Thus, under the assumption that there exists  $\hat{\ell} \in A_{n,k}^{s'}$  such that  $\hat{\ell} \geq s$ , a contradiction (see (20) and (21)) is reached. Consequently, every  $\hat{\ell} \in A_{n,k}^{s'}$  satisfies  $\hat{\ell} < s$ . Furthermore, since  $s' > s > l_{nks'}$ , by arguments similar to those of step (i), we conclude that

$$(22) \quad A_{n,k}^{s'} = A_{n,k}^s \quad \text{for all } s' \text{ such that } l_{nks} < s < s'.$$

Hence, for all states  $(n, k, s), (n, k, s')$  such that  $s > l_{nks}, s' > l_{nks}$ , we have

$$(23) \quad l_{nks} = l_{nks'} := l_{nk}.$$

From (2), (3), (8), (9), (19), (22), and (23), we conclude that there exists  $l_{nk}$  such that

$$(24) \quad V(n, k, s) < V(n, k + 1, s) \quad \text{for all } s > l_{nk}.$$

*Proof of (5).* Consider any  $t \leq l_{nk}$ . We want to prove that for such a  $t$ ,

$$(25) \quad V(n, k, t) = V(n, k + 1, t).$$

To do this we formalize the idea outlining the proof of (F2) in section 3. Take  $(n, k, s)$ ,  $s > l_{nk}$ , as the initial state and use the allocation policy  $\pi^*$ . Let  $\bar{\Omega}$  be the set of sample paths starting at  $(n, k, s)$  along which there are no job completions until  $t$  time units remain on the timer. Then, because of (2), (8), (9), and the definition of policy  $\pi^*$ , the expected cost incurred by the optimal policy  $\pi^*$  along  $\bar{\Omega}$  is  $V(n, k + r, s)$  for some  $r \geq 1$ . Moreover, since  $r$  machines are added by the optimal policy  $\pi^*$  along  $\bar{\Omega}$ , it follows that when there are  $n$  unfinished jobs and  $t$  units of time remain on the timer it is optimal to use  $k + r$  machines ( $r \geq 1$ ). Hence,

$$(26) \quad V(n, k + r, t) = V(n, t) \leq V(n, k, t).$$

On the other hand, by Lemma 2.1(ii)

$$(27) \quad V(n, k + r, t) \geq V(n, k, t).$$

From (26) and (27) we conclude that

$$(28) \quad V(n, k, t) = V(n, k + r, t).$$

From (28) and Lemma 2.1(ii) we obtain

$$(29) \quad V(n, k, t) = V(n, k + 1, t) \quad \text{for } t \leq l_{nk}.$$

*Proof of (6).* We use (4) and (5) and formalize the idea outlining the proof of (F3) in section 3. Suppose that  $l_{nk} > l_{n(k-1)}$ . Consider  $t$  such that  $l_{nk} > t > l_{n(k-1)}$  and assume that there are  $n$  jobs to be processed at  $t$ . Then  $t < l_{nk}$  and (5) imply that

$$(30) \quad V(n, k, t) = V(n, k + 1, t).$$

From (1), (30), Lemma 2.1(ii), and the Bellman equation it follows that it is optimal to use at least  $k + 1$  machines when there are  $n$  jobs to be processed with  $t$  units of time remaining on the timer. On the other hand,  $t > l_{n(k-1)}$  and (4) imply that

$$(31) \quad V(n, k, t) > V(n, k - 1, t).$$

From (1), (31), Lemma 2.1(ii), and the Bellman equation we conclude that it is optimal to use less than  $k$  machines when there are  $n$  jobs to be processed with  $t$  units of time left on the timer. Thus, under the assumption  $l_{nk} > l_{n(k-1)}$  we reach a contradiction. Consequently,  $l_{nk} \leq l_{n(k-1)}$ . By arguments similar to the above we obtain the remaining inequalities in (6).

The proof of the conjecture is now complete.  $\square$

**5. Discussion.** The proof of the conjecture stated in [1] and described in section 2 completes the proof of optimality of the policy described in Theorem 2.3. Theorem 2.3 describes the nature of an optimal policy for the stochastic machine minimization problem formulated in [1] and described in sections 1 and 2 of this paper. According to this optimal policy allocation decisions are made at job completions, when the timer expires, or when the “time thresholds”  $l_{nk}$  are reached. These “time thresholds” can be computed for all  $n, k, k = 1, 2, \dots, n - 1$  by (8) and (9). The computation of  $l_{nk}$  is a challenging and formidable task that will not be further pursued in this paper.

The stochastic machine minimization problems investigated by Coffman, Flatto, and Wright in [1], [2], [3] consider only the scheduling of jobs (tasks, projects) that are initially available (i.e., they are available at time 0) in the system. Stochastic machine minimization problems with arrivals are interesting extensions of the problems considered in [1], [2], [3], as they arise in wireless communication networks and in automated target recognition systems (see [5]). A stochastic machine minimization problem with Poisson arrivals, exponential service times, and hard deadlines has been investigated in [5].

A class of interesting technical questions is to determine whether the structure of the optimal policies described in [1], [2], [3] remains unaltered when arrivals are included in the problem formulation.

**Acknowledgments.** The authors are grateful to the anonymous reviewers, whose comments helped to significantly improve the presentation of the results of this paper.

#### REFERENCES

- [1] E. G. COFFMAN, JR., L. FLATTO, AND P. E. WRIGHT, *Optimal stochastic allocation of machines under waiting-time constraints*, SIAM J. Comput., 22 (1993), pp. 332–348.
- [2] E. G. COFFMAN, JR., L. FLATTO, AND P. E. WRIGHT, *Stochastic machine minimization with constant service times*, Math. Oper. Res., 18 (1993), pp. 300–316.
- [3] E. G. COFFMAN, JR., L. FLATTO, B. POONEN, AND P. E. WRIGHT, *The processor minimization problem with independent waiting-time constraints*, Theoret. Comput. Sci., 125 (1994), pp. 3–16.
- [4] M. PINEDO AND L. SCHRAGE, *Stochastic shop scheduling: A survey*, in Deterministic and Stochastic Scheduling, M. Dempster, J. K. Lenstra, and A. Rinoooy-Kan, eds., D. Reidel, Dordrecht, The Netherlands, 1982.
- [5] N. SONG AND D. TENEKETZIS, *The Stochastic Machine Minimization Problem with Hard Deadlines and Arrivals*, Control Group Report CGR-96-13, Department of EECS, University of Michigan, 1996 (under revision).
- [6] R. R. WEBER, *Scheduling jobs with stochastic processing requirements on parallel machines to minimize makespan or flowtime*, J. Appl. Probab., 19 (1982), pp. 167–182.
- [7] G. WEISS, *Multiserver stochastic scheduling*, in Deterministic and Stochastic Scheduling, M. Dempster, J. K. Lenstra, and A. Rinoooy-Kan, eds., D. Reidel, Dordrecht, The Netherlands, 1982.
- [8] G. WEISS AND M. PINEDO, *Scheduling tasks with exponential services times on nonidentical processors to minimize various cost functions*, J. Appl. Probab., 17 (1980), pp. 187–202.