Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Technical Note—A New Proof of the Optimality of the Shortest Remaining Processing Time Discipline

Donald R. Smith,

To cite this article:
Donald R. Smith, (1978) Technical Note—A New Proof of the Optimality of the Shortest Remaining Processing Time Discipline. Operations Research 26(1):197-199. https://doi.org/10.1287/opre.26.1.197

OPERATIONS RESEARCH
Vol. 26, No. 1, January-February 1978

0030-364X/78/2601-0197 $01.25

Technical Notes

A New Proof of the Optimality of the Shortest Remaining Processing Time Discipline

DONALD R. SMITH

Columbia University, New York, New York

(Received July 1, 1976)

We present a new proof of the fact that for a work-conserving queue, the queuing discipline that always serves a job with the shortest remaining processing time minimizes the number of jobs in the system. A key feature of the proof is a definition of work dominance, allowing comparison of two systems based on the remaining service times of jobs present. Work dominance is both necessary and sufficient for stochastic comparison of the number of jobs present under identical but arbitrary arrival processes.

SCHRAGE [1] proved that for a general work-conserving queuing system in which the arrival process is independent of the queuing discipline, the number of jobs in the system at any point in probability space was minimized by the SRPT discipline (shortest remaining processing time) over the class of possible queuing disciplines. We present an alternative proof based on a definition of dominance that allows one to say that the remaining service times of jobs in one system are "better" than the remaining service times of jobs in another system. Such a definition includes the requirement that "better" means fewer jobs in the system. The proof depends on the fact that for a fixed arrival process SRPT preserves dominance of one system over another system using any queuing discipline. As a bonus, we obtain the converse that if system A does not dominate system B, then for a specific arrival process and elapsed time the number of jobs in B will be less than the number of jobs in A under SRPT. Thus the definition of dominance is a good one.

1. DEFINITIONS

An infinite vector having non-negative, non-increasing components with finite sum will be called a work vector. The state of queue at any time can be specified by a work vector X having the following interpretation: X_1 is the maximum remaining service time in the system, X_2 is the next greatest remaining service time, and so forth, with duplications being repeated explicitly. The $(k+1)$-st and further coordinates will be the only coordinates equal to zero if and only if there are k jobs in the system. Thus the total work in the system is $\sum_{i=1}^{\infty} X_i$.

Given two work vectors X and Y, we say X work-dominates Y, written $X \rightarrow Y$ if $\sum_{i=k}^{\infty} X_i \leqq \sum_{i=k}^{\infty} Y_i \forall k$. In the previous definition, letting k be the number of nonzero coordinates in X, we see $Y_k > 0$ and thus there are at least k nonzero coordinates in Y. (If all coordinates of X are nonzero, this must also be true for Y.) Thus, from our previous interpretation of a work vector as a state, $X \rightarrow Y$ tells us that there are at least as many jobs in the Y system as in the X system.

2. RESULTS

LEMMA 1. *Suppose the state of the described queuing system under an arbitrary queuing discipline at time t is Y and at time $t+\Delta t$ is Y'. Furthermore, suppose no arrivals occur in the interval $(t, t+\Delta t]$. Then (a) $(Y_i - Y_i') \geqq 0$ and (b) $\sum_{i=1}^{\infty} (Y_i - Y_i') \leqq \Delta t$.*

Proof. To see this suppose $Y_i' > Y_i$. Then the job corresponding to Y_i' has to be a job corresponding to one of $Y_1, \cdots, Y_{i-1}$ and requiring more time than Y_i since no arrivals occurred. But $Y_1 \geqq \cdots \geqq Y_i$ tells us there is a total of i jobs requiring more time than Y_i originally, which is a contradiction. The second result follows from the fact that the work performed in the interval must be less than or equal to Δt.

LEMMA 2. *If $X \rightarrow Y$ and X' is obtained by subtracting Δt from the last nonzero coordinate of X (leaving this positive), then $X' \rightarrow Y'$ where Y' is any work vector satisfying (a) and (b) of Lemma 1.*

Proof. If $\sum_{i=k}^{\infty} X_i' = 0$, then $\sum_{i=k}^{\infty} X_i' \leqq \sum_{i=k}^{\infty} Y_i'$. If $\sum_{i=k}^{\infty} X_i' > 0$, then $\sum_{i=k}^{\infty} X_i' = \sum_{i=k}^{\infty} X_i - \Delta t \leqq \sum_{i=k}^{\infty} Y_i - \Delta t \leqq \sum_{i=k}^{\infty} Y_i'$.

LEMMA 3. *Suppose $X \rightarrow Y$ and X' is the state of a system obtained by adding a job of work V to the system whose state is X, and Y' is the state of a system obtained by adding a job of work V to the system whose state is Y. Then $X' \rightarrow Y'$.*

Proof. Suppose that V is the lth coordinate of X' and the mth coordinate of Y'.

$$l<k,\ m<k:\ \sum_{i=k}^{\infty} X_i' = \sum_{i=k-1}^{\infty} X_i \leqq \sum_{i=k-1}^{\infty} Y_i = \sum_{i=k}^{\infty} Y_i'.$$

$$l<k,\ m\geqq k:\ \sum_{i=k}^{\infty} X_i' = \sum_{i=k-1}^{\infty} X_i \leqq V + \sum_{i=k}^{\infty} X_i \leqq V + \sum_{i=k}^{\infty} Y_i = \sum_{i=k}^{\infty} Y_i'.$$

$$l\geqq k,\ m<k:\ \sum_{i=k}^{\infty} X_i' = V + \sum_{i=k}^{\infty} X_i \leqq \sum_{i=k-1}^{\infty} X_i \leqq \sum_{i=k-1}^{\infty} Y_i = \sum_{i=k}^{\infty} Y_i'.$$

$$l\geqq k,\ m\geqq k:\ \sum_{i=k}^{\infty} X_i' = V + \sum_{i=k}^{\infty} X_i \leqq V + \sum_{i=k}^{\infty} Y_i = \sum_{i=k}^{\infty} Y_i'.$$

These three lemmas easily give the result that SRPT is optimal.

Proof. Let the evolution of the state vector Y under some queuing discipline be on a space $(\Omega, \mathfrak{F}, P)$. For $\omega \in \Omega$ assume the arrival process is fixed and let X be the state vector under the SRPT discipline. We have easily that $X(\omega, 0) \to Y(\omega, 0)$; and Lemmas 2 and 3, together with the trivial dominance when $X=0$, imply that $X(\omega, t) \to Y(\omega, t) \forall t$. Thus the number of customers under SRPT is minimized for every ω, t.

The converse is also of interest. Suppose $X \not\to Y$. Then for a particular arrival process and elapsed time, under SRPT the system whose state was X will have more jobs present than the system whose state was Y.

Proof. We have a k for which $\sum_{i=k}^{\infty} X_i > \sum_{i=k}^{\infty} Y_i$. Let no arrivals occur during a time interval of length $\sum_{i=k}^{\infty} Y_i$. The system corresponding to Y will now have $k-1$ customers and the system corresponding to X will have at least k customers.

REFERENCE

1. L. Schrage, "A Proof of the Optimality of the Shortest Remaining Service Time Discipline," *Opns. Res.* **16**, 687–690 (1968).

Published in Operations Research on February 01, 1978 as DOI: 10.1287/opre.26.1.197.
This article has not been copyedited or formatted. The final version may differ from this version.