Optimality of index policies for stochastic scheduling with switching penalties
— **Source link** ⬈

Mark P. Van Oyen, Dimitrios G. Pandelis, Demosthenis Teneketzis

Related papers:

- Extensions of the multiarmed bandit problem: The discounted case

- Restless bandits: activity allocation in a changing world

- Multi-Armed Bandits and the Gittins Index

- Multi-Armed Bandit Allocation Indices.

- Optimal stochastic scheduling of forest networks with switching penalties

Optimality of Index Policies for Stochastic Scheduling with Switching Penalties
Author(s): Mark P. Van Oyen, Dimitrios G. Pandelis, Demosthenis Teneketzis
Source: *Journal of Applied Probability,* Vol. 29, No. 4, (Dec., 1992), pp. 957-966
Published by: Applied Probability Trust
Stable URL: http://www.jstor.org/stable/3214727
Accessed: 04/04/2008 12:08

OPTIMALITY OF INDEX POLICIES FOR STOCHASTIC SCHEDULING WITH SWITCHING PENALTIES

MARK P. VAN OYEN,
DIMITRIOS G. PANDELIS AND
DEMOSTHENIS TENEKETZIS,* *University of Michigan*

Abstract

We investigate the impact of switching penalties on the nature of optimal scheduling policies for systems of parallel queues without arrivals. We study two types of switching penalties incurred when switching between queues: lump sum costs and time delays. Under the assumption that the service periods of jobs in a given queue possess the same distribution, we derive an index rule that defines an optimal policy. For switching penalties that depend on the particular nodes involved in a switch, we show that although an index rule is not optimal in general, there is an exhaustive service policy that is optimal.

OPTIMAL CONTROL OF QUEUES; SWITCHING COST; SWITCHING TIME; MULTI-ARMED BANDITS; COUPLING

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 93 E20
SECONDARY 90 B35

1. Introduction

We consider a system of N parallel queues and a single server attending them. There are no arrivals: we are concerned with the scheduling of the jobs initially present in the system. We seek a non-preemptive, non-idling scheduling policy which minimizes the sum of an expected discounted holding cost and an expected discounted switching penalty (switching cost or switching delay) incurred whenever the server moves from one queue to another.

The novel feature of this work is the determination of *optimal* stochastic scheduling policies when penalties are incurred for changes in resource allocation. Whereas stochastic scheduling problems without switching cost have received much attention (see for example Baras et al. (1985); Buyukkoc et al. (1985); Dempster et al. (1982); Gittins (1989); Klimov (1974), (1978); Lai and Ying (1988); Nain (1989); Nain et al. (1989); Varaiya et al. (1985); Walrand (1988); and references therein), there is little work on optimal scheduling and resource allocation with switching penalties (see Agrawal

et al. (1988), (1990); Glazebrook (1980); Gupta et al. (1987); and Hofri and Ross (1987)). The reason, it appears, is the difficulty of such problems. The inclusion of switching penalties changes the fundamental features of even simple scheduling problems, and most of the simple and intuitively appealing results reported in the first set of references cited above no longer hold. Until now, most of the research on stochastic scheduling and resource allocation with switching penalties has concentrated on performance evaluation of *ad hoc* policies for systems with switching delays (see for example Baker and Rubin (1987), Browne and Yechiali (1989), Eisenberg (1972), Ferguson and Aminetzah (1985), Murata and Takagi (1986), Sykes (1970), and Takagi (1986)).

From an applications perspective, queueing models such as the one considered in this paper capture the essence of fundamental issues encountered in manufacturing, computer, and communication networks as well as the scheduling of research projects (see Dempster et al. (1982); Gittins (1989); Glazebrook (1980); Hofri and Ross (1987); Takagi (1986); and Walrand (1988) and the references therein). One feature of many applications that is notably absent from our formulation is that of job arrivals. The problem with arrivals and switching penalties is difficult and researchers have only *partially* characterized optimal policies for *special cases of two parallel queues* (see Gupta et al. (1987) and Hofri and Ross (1987)). We derive *optimal* scheduling policies for a model of N parallel queues with switching penalties, linear holding costs, general service distributions, and no arrivals. Thus, our work differs from that of Gupta et al. (1987) and Hofri and Ross (1987). It is also different from Glazebrook (1980) who identifies certain conditions on the service distribution for which existing algorithms (see Glazebrook and Gittins (1981) and Sidney (1975)) can be used to solve stochastic scheduling problems with precedence constraints and switching costs associated with individual jobs.

This paper is organized as follows. We formulate the problem precisely in Section 2. In Section 3, we prove that only exhaustive policies can be optimal. For switching penalties that do not depend on the particular transition, we show in Section 4 that optimal policies are characterized by indices and calculate these indices for the problem with switching cost and with switching delay.

2. Problem formulation

We consider a system of N queues (nodes) and a single server; alternatively, one may think of a single queue and server with N classes of jobs. We are concerned only with the scheduling of the jobs initially present in the system. The N queues are differentiated by service distribution and holding cost. The lengths of the service periods for all jobs are mutually independent random variables. The service period of a job of type n ($n = 1, 2, \cdots, N$) possesses a general distribution with mean $1/\mu_n$ (assume $0 < 1/\mu_n < \infty$). For every job of type n, an instantaneous holding cost c_n is incurred until the instant the job leaves the system.

In addition to holding costs, we assess a penalty for switching. Two optimization problems are considered; one treats lump sum switching costs and the other switching delays. These penalties are defined as follows. For the *switching cost* problem, a

switching cost, $K_{ij} > 0$, is incurred at each instant the server completes a job of node i and then serves a job of node j. We require that for all $i, j, l \in \{1, 2, \cdots, N\}$,

$$(2.1) \qquad K_{ij} < K_{il} + K_{lj}.$$

For the *switching delay* problem, a random length of time, D_{ij}, is required to switch from queue i to j; thus, the holding cost incurred during the switching delay is the implicit penalty. The successive delays resulting from a transition from node i to j are assumed to be positive i.i.d. random variables with a general distribution and finite mean. For a given parameter $\alpha \geqq 0$, it is also assumed that for all $i, j, l \in \{1, 2, \cdots, N\}$,

$$(2.2) \qquad E\{\exp(-\alpha D_{ij})\} > E\{\exp(-\alpha(D_{il} + D_{lj}))\}.$$

Note that (2.2) is implied by the assumption that D_{ij} is stochastically smaller than $D_{il} + D_{lj}$.

The class of admissible strategies is taken to be the class G of non-idling, non-preemptive, and non-anticipative scheduling policies. (The switching delay is considered part of the act of service, not idling.) With $\mathbb{R}^+$ ($\mathbb{Z}^+$) denoting the non-negative reals (integers), let $\{X_n(t): t \in \mathbb{R}^+\}$ be the right-continuous queue length process of node $n \in \{1, 2, \cdots, N\}$ at time t (we include the node n customer, if any, in service at time t). We denote the vector of initial queue lengths by $x = (x_1, x_2, \cdots, x_N) = X(0^-) \in (\mathbb{Z}^+)^N$. Let $n^g(t)$ be a right-continuous process which describes the location of the server at time t under policy g. For both the switching cost problem and the switching delay problem, the objective is to characterize a policy which minimizes an infinite horizon expected discounted cost due to holding and switching. Specifically, let $\Gamma_{ij}^g = \{t \in \mathbb{R}^+: n^g(t^-) = i, n^g(t) = j\}$ be the set of random switching instants under policy g corresponding to a transition from i to j; we see that $\cup_{i,j=1}^N \Gamma_{ij}^g$ contains at least $N - 1$ elements and at most $(\Sigma_n x_n - 1)$. For the problem of switching costs, the expected α-discounted cost $J(g)$ due to policy $g \in G$ is defined as

$$(2.3) \qquad J(g) \triangleq E\left\{\int_0^\infty e^{-\alpha t} \sum_{n=1}^N c_n X_n^g(t)dt + \sum_{i,j=1}^N \sum_{t \in \Gamma_{ij}^g} K_{ij} e^{-\alpha t} \mid X(0^-) = x\right\}.$$

The objective is to characterize a policy $g^* \in G$ that minimizes $J(g)$. In the case of switching delay, the cost to be minimized is

$$(2.4) \qquad \bar{J}(g) \triangleq E\left\{\int_0^\infty \exp(-\alpha t) \sum_{n=1}^N c_n X_n^g(t)dt \mid X(0^-) = x\right\}.$$

Without loss of optimality (see Ross (1983)), the minimization of both problems can be restricted to the class $G^{PM} \subset G$ of pure Markov policies.

3. Optimality of exhaustive policies

We begin with a fundamental structural property characterizing an optimal policy for either problem. First we need the following.

Definition 1. A policy g is said to be *exhaustive* if according to g the server never leaves a node before completing the service of all the customers at that node.

Theorem 1. Only exhaustive policies can be optimal.

Proof. The following arguments apply to either the case of switching cost or switching delay.

Suppose g $(g \in G^{PM})$ is a policy that is not exhaustive. Then, there exists at least one queue, say q, such that after g begins service in q it moves to another node before completing the service of all customers present in q, and returns to q later on. Suppose that according to g it takes p visits to node q to serve all of its customers. Then for $i = 1, 2, \cdots, p$, let: (1) q^i be the number of customers served during the ith visit to node q; (2) s_q^i be the time when g starts serving node q for the ith time; (3) f_q^i be the time when g leaves node q for the ith time; (4) $\hat{s}_q^i$ $(\hat{f}_q^i)$ denote the last service completion epoch under g before s_q^i (f_q^i); and (5) $\tilde{s}_q^i$ $(\tilde{f}_q^i)$ denote the first service completion epoch under g after s_q^i (f_q^i). We construct two modifications of g, $g^{l(1)} \in G^{PM}$ and $g^{r(1)} \in G^{PM}$, which are illustrated in Figure 1.
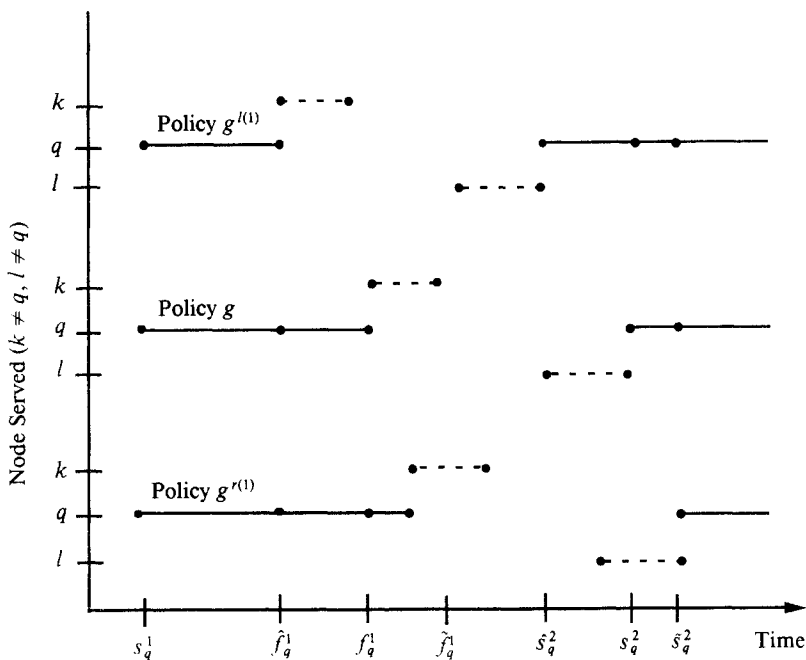
Figure 1. Illustration of policies $g^{r(1)}$ and $g^{l(1)}$

Consider the policy $g^{l(1)}$ with the following characteristics: (1) $g^{l(1)}$ is the same as g up to time $\hat{f}_q^1$; (2) from $\hat{f}_q^1$ until $s_q^2 - (f_q^1 - \hat{f}_q^1)$, $g^{l(1)}$ follows the same schedule as g does from f_q^1 up to s_q^2; (3) at $s_q^2 - (f_q^1 - \hat{f}_q^1)$, $g^{l(1)}$ switches to node q; and (4) from time s_q^2 on, g and $g^{l(1)}$ are the same.

Similarly, consider the policy $g^{r(1)}$ with the following characteristics: (1) $g^{r(1)}$ is the same as g until f_q^1; (2) at f_q^1, $g^{r(1)}$ remains at node q to serve one additional customer; (3) from $f_q^1 + (\hat{s}_q^2 - s_q^2)$ up to $\hat{s}_q^2$, $g^{r(1)}$ follows the same schedule as g does from f_q^1 up to s_q^2; and (4) from $\hat{s}_q^2$ on, $g^{r(1)}$ and g are the same.

We now compare policies $g^{l(1)}$ and $g^{r(1)}$ to g. Noting that the service periods are mutually independent and denoting by $J(\gamma)$ the expected discounted cost incurred by any policy γ, we obtain

$$(3.1) \qquad J(g^{l(1)}) - J(g) = E\{\exp(-\alpha \hat{f}_q^1)\} E\{\mathscr{L}^{l(1)}\},$$

$$(3.2) \qquad J(g) - J(g^{r(1)}) = E\{\exp(-\alpha f_q^1)\} E\{\mathscr{L}^{r(1)}\}.$$

The random variable $\mathscr{L}^{l(1)}$ represents the difference in holding costs and switching penalties associated with the control actions of $g^{l(1)}$ and g between $\hat{f}_q^1$ and s_q^2 after time is shifted $\hat{f}_q^1$ units to the left. Similarly, the term $\mathscr{L}^{r(1)}$ reflects the difference in holding costs and switching penalties associated with the control actions of $g^{r(1)}$ and g between f_q^1 and $\hat{s}_q^2$ when time is shifted f_q^1 units to the left. Since service periods at each node are i.i.d., $\mathscr{L}^{l(1)}$ is equal to $\mathscr{L}^{r(1)}$ in distribution. Consequently, $E\{\mathscr{L}^{l(1)}\} = E\{\mathscr{L}^{r(1)}\} \triangleq A \in \mathbb{R}$ and

$$(3.3) \qquad J(g^{l(1)}) - J(g) = d_1 A,$$

$$(3.4) \qquad J(g) - J(g^{r(1)}) = b_1 A,$$

where d_1 and b_1 are positive constants. We consider two cases: $A \geqq 0$ and $A < 0$.

Case I: $A \geqq 0$. Then, $g^{r(1)}$ performs better than g. Based on $g^{r(1)}$, define a new policy $g^{r(2)}$ by the same modification that produced $g^{r(1)}$ from g. Then,

$$(3.5) \qquad J(g^{r(1)}) - J(g^{r(2)}) = b_2 A$$

for some positive constant b_2; and since $A \geqq 0$, $g^{r(2)}$ performs better than $g^{r(1)}$. Proceeding in this way, define policies $g^{r(3)}, g^{r(4)}, \cdots, g^{r(q^2-1)}, g^{r(q^2)}$. Then for all $j = 1, 2, \cdots, q^2 - 1$, there are constants b_j such that

$$(3.6) \qquad J(g^{r(j-1)}) - J(g^{r(j)}) = b_j A,$$

$$(3.7) \qquad J(g^{r(q^2-1)}) - J(g^{r(q^2)}) \geqq b_{q^2}(A + e)$$

for some $e > 0$, because $g^{r(q^2)}$ requires at least one less switch than $g^{r(q^2-1)}$. Suppose that $g^{r(q^2-1)}$ switches from node i to q at time $\hat{f}_q^2$ and then from q to j at f_q^2. If $i \neq j$, either condition (2.1) or (2.2) guarantees that the resulting transition from i to j under $g^{r(q^2)}$ at time f_q^2 represents a strictly positive cost saving over $g^{r(q^2-1)}$. Consequently, when $A \geqq 0$, we can improve the performance of policy g by adopting $g^{r(q^2)}$ which has the following characteristics: (1) its first visit to node q occurs at s_q^1; (2) the number of customers served during its first visit to q is $q^1 + q^2$; and (3) after f_q^2, $g^{r(q^2)}$ is the same as g.

Case II: $A < 0$. In this case, $g^{l(1)}$ performs better than g. Based on $g^{l(1)}$, define a new policy $g^{l(2)}$ by the same modification that produces $g^{l(1)}$ from g. Then,

$$(3.8) \qquad J(g^{l(2)}) - J(g^{l(1)}) = d_2 A$$

for some $d_2 > 0$, and since $A < 0$, $g^{l(2)}$ performs better than $g^{l(1)}$. Proceeding in this way, define policies $g^{l(3)}, g^{l(4)}, \cdots, g^{l(q^1-1)}, g^{l(q^1)}$. Then for all $j = 1, 2, \cdots, q^1 - 1$, there are positive constants d_j such that

$$(3.9) \qquad\qquad J(g^{l(j)}) - J(g^{l(j-1)}) = d_j A,$$

$$(3.10) \qquad\qquad J(g^{l(q^1)}) - J(g^{l(q^1-1)}) < d_{q^1} A.$$

Note that either (2.1) or (2.2) implies (3.10), since $g^{l(q^1)}$ saves at least one switch over $g^{l(q^1-1)}$. Therefore, when $A < 0$, we can improve the performance of policy g by adopting $g^{l(q^1)}$ which has the following properties: (1) its first visit to node q occurs at $s_q^2 - \Sigma_{i=1}^{q^1} \sigma_q(i)$, where $\sigma_q(i)$ is the service duration of the ith customer served in node q; (2) the number of customers served during its first visit to node q is $q^1 + q^2$; and (3) after s_q^2, $g^{l(q^1)}$ is the same as g.

In either of the above cases, we can improve performance by a policy which during its first visit to node q serves $q^1 + q^2$ customers. Applying $(p-1)$ times the whole argument so far, we continually improve system performance and thereby produce a policy which serves node q exhaustively.

The same procedure can be applied to other nodes that are not exhaustively served to prove that only exhaustive policies can be optimal.

As a consequence of Theorem 1, an optimal policy prescribes exactly $N - 1$ switches. Theorem 1 provides a reduction in the class of policies which are candidate solutions to the optimization problem. The number of policies in G^{PM} is large and depends on the initial queue lengths. Theorem 1 limits the search to the set G^E of exhaustive pure Markov policies. Since G^E contains $N!$ policies, it would be interesting to determine further structure.

4. Optimality of index policies

In this section, we prove that index rules are optimal for the problems formulated in Section 2 under the following additional assumption: for all $i, j \in \{1, 2, \cdots, N\}$, the switching costs are equal,

$$(4.1) \qquad\qquad K_{ij} = K,$$

and the switching delays are all drawn from a common distribution so that

$$(4.2) \qquad\qquad D_{ij} = D,$$

with successive switching delays being independent.

In the case of zero switching cost and no switching delay, it is known (Proposition 8.2.2 of Walrand (1988)) that the policy that serves jobs in decreasing order of $c_i \mu_i$ minimizes the total expected holding cost. Such a $c\mu$-rule that serves each queue in an exhaustive manner incurs $(N - 1)$ switches. Since every policy requires at least $(N - 1)$ switches, it follows that in the case of switching cost and $\alpha = 0$, it is optimal to serve the nodes exhaustively in decreasing order of $c_i \mu_i$. Thus, we concentrate on the scheduling

problem of switching delays with discount parameter $\alpha \geq 0$ and on the switching cost problem with $\alpha > 0$.

Our development is based on the following ideas. We first show that minimizing the cost of a policy as defined by (2.3) or (2.4) is equivalent to maximizing a reward obtained by serving the jobs in the system. After obtaining explicit expressions for the reward earned under a given exhaustive policy, we show that for each queue it is possible to define a measure of reward rate that incorporates the holding cost savings and switching penalty associated with serving that queue. Thus, we convert the problem of maximizing the reward into a multi-armed bandit problem. For this problem we know from Gittins (1989), Chapter 3, that an optimal policy is characterized by an index rule. We derive explicit formulas for the indices of the queues for the switching cost problem and the switching delay problem.

We develop expressions for expected discounted reward in the case that $\alpha > 0$ for both the switching cost problem and the switching delay problem. For the switching cost problem with $\alpha > 0$ and $g \in G^E$, we can rewrite (2.3) as

(4.3)
$$J(g) = \sum_{n=1}^{N} c_n x_n \alpha^{-1} - K$$

$$- \sum_{n=1}^{N} E\left\{ \int_0^\infty \exp(-\alpha t) c_n Y_n^g(t) dt - K \exp(-\alpha t^g(n)) \right\},$$

where $(x_1, x_2, \cdots, x_N)$ are the initial queue lengths; $\{Y_n^g(t), t \geq 0\}$ is the process describing the number of departures from node n through time t; and $t^g(n)$ denotes the random time at which g initiates the service of node n. Minimization of $J(g)$ is equivalent to the maximization of the reward $R(g)$ over the class of exhaustive policies G^E, where

(4.4) $$R(g) \triangleq \sum_{n=1}^{N} E\left\{ \int_0^\infty \exp(-\alpha t) c_n Y_n^g(t) dt - K \exp(-\alpha t^g(n)) \right\}.$$

We introduce some notation to write $R(g)$ more explicitly. Let the time at which the kth departure of $\{Y_n^g(t), t \geq 0\}$ occurs be denoted by $t^g(n) + f_{n,k}$, where $f_{n,k} \triangleq \sum_{j=1}^{k} \sigma_n(j)$ and $\sigma_n(j)$ denotes the length of the jth service period in node n. With $S_n \triangleq E\{\exp(-\alpha \sigma_n(1))\}$, we note that the i.i.d. nature of service periods in node n yields $E\{\exp(-\alpha f_{n,k})\} = S_n^k$. Let $C_n(x_n)$ denote the holding cost saved by serving x_n jobs in n beginning at time 0:

(4.5) $$C_n(x_n) = E\left\{ \sum_{k=1}^{x_n} \int_{f_{n,k}}^\infty c_n \exp(-\alpha t) dt \right\}$$

(4.6) $$= c_n \alpha^{-1} S_n (1 - S_n)^{-1} (1 - S_n^{x_n}).$$

Then, we can write the reward of policy g as

(4.7) $$R(g) = \sum_{n=1}^{N} E\{\exp(-\alpha t^g(n))\}(C_n(x_n) - K).$$

The analysis of the problem with switching delay closely follows that above. For $g \in G^E$ and $\alpha > 0$, the reward is

$$(4.8) \qquad \bar{R}(g) = \sum_{n=1}^{N} E\{\exp(-\alpha t^g(n))\} C_n(x_n).$$

The time g begins node n, $t^g(n)$, is taken to include any switching delays incurred prior to processing node n.

Based on (4.7) and (4.8), the effects of holding cost savings and switching penalties can be combined to yield an equivalent constant reward rate: the ratio of expected discounted reward earned serving a queue to the expected discounted time required to effect the switch and to clear the queue. With $\alpha > 0$, the reward rates for the switching cost problem and switching delay problem are given by

$$(4.9) \qquad v_n = \frac{C_n(x_n) - K}{E\left\{ \int_0^{\tau_n} \exp(-\alpha t) dt \right\}} = c_n S_n (1 - S_n)^{-1} - \alpha K (1 - S_n^{x_n})^{-1},$$

$$(4.10) \qquad \bar{v}_n = \frac{C_n(x_n)}{E\left\{ \int_0^{\tau_n + D} \exp(-\alpha t) dt \right\}} = \frac{c_n S_n (1 - S_n)^{-1} (1 - S_n^{x_n})}{1 - S_n^{x_n} E\{\exp(-\alpha D)\}},$$

respectively, where $\tau_n = f_{n,x_n}$ is the total processing time of x_n jobs in node n. The original problem with switching cost (switching delay) is now equivalent to a standard multi-armed bandit problem, where the constant reward rate of arm n ($n = 1, 2, \cdots, N$) is given by (4.9) ((4.10)). This constant reward rate also gives the dynamic allocation index of arm n. Thus, Theorem 3.6 of Gittins (1989) justifies the following result.

Theorem 2. An index rule is optimal: serve the queues exhaustively in decreasing order of the dynamic allocation indices v_n for the switching cost problem and $\bar{v}_n$ for the switching delay problem.

Following the method discussed in Section 3.6 of Gittins (1989), it is straightforward to prove that an index policy is optimal for the problem with switching delay and $\alpha = 0$. The index of node n is given by

$$(4.11) \qquad \bar{v}_n = c_n \mu_n (x_n \mu_n^{-1}) / (x_n \mu_n^{-1} + E\{D\}).$$

This rule represents the familiar $c\mu$ index multiplied by the fraction of time over which useful work is performed.

For fixed α, the index of each node reveals the tradeoff between the initial queue length, the switching penalty, and the expected discounted reward rate (i.e. $cS(1 - S)^{-1}$) received by serving a given queue. The longer the queue, the less is the penalty incurred due to switching. Hence, higher priority should be given to queues with a sufficiently large number of customers and high $cS(1 - S)^{-1}$, so that the overall reward received by serving the queue is maximized. If $K = 0$ and $D = 0$, Theorem 2 states a generalized version of the $c\mu$-rule. On the other hand, if the holding costs and service distributions

are the same for all queues, the above discussion indicates that the optimal policy should proceed in order of decreasing queue length.

We conclude by noting that for $\alpha > 0$ and exponential service distributions with rate μ_n for node n, (4.9) and (4.10) give

$$(4.12) \qquad v_n = c_n \mu_n \alpha^{-1} - \frac{\alpha K}{1 - (\mu_n/(\alpha + \mu_n))^{x_n}},$$

$$(4.13) \qquad \bar{v}_n = c_n \mu_n \alpha^{-1} \frac{1 - (\mu_n/(\alpha + \mu_n))^{x_n}}{1 - (\mu_n/(\alpha + \mu_n))^{x_n} E\{\exp(-\alpha D)\}}.$$

References

[1] AGRAWAL, R., HEGDE, M. AND TENEKETZIS, D. (1988) Asymptotically efficient adaptive allocation rules for the multi-armed bandit problem with switching cost. *IEEE Trans. Autom. Control* **33**, 899–906.

[2] AGRAWAL, R., HEGDE, M. AND TENEKETZIS, D. (1990) Multi-armed bandit problems with multiple plays and switching cost. *Stoch. Stoch. Rep.* **29**, 437–459.

[3] BAKER, J. E. AND RUBIN, I. (1987) Polling with a general-service order table. *IEEE Trans. Comm.* **35**, 283–288.

[4] BARAS, J. S., MA, D.-J. AND MAKOWSKI, A. M. (1985) K competing queues with geometric service requirements and linear costs: the μc-rule is always optimal. *Systems Control Lett.* **6**, 173–180.

[5] BROWNE, S. AND YECHIALI, U. (1989) Dynamic priority rules for cyclic-type queues. *Adv. Appl. Prob.* **21**, 432–450.

[6] BUYUKKOC, C., VARAIYA, P. AND WALRAND, J. (1985) The $c\mu$-rule revisited, *Adv. Appl. Prob.* **17**, 237–238.

[7] DEMPSTER, M. A. H., LENSTRA, J. K. AND RINNOOY KAN, A. M. G. (1982) *Deterministic and Stochastic Scheduling.* D. Reidel, Dordrecht.

[8] EISENBERG, M. (1972) Queues with periodic service and changeover time. *Operat. Res.* **20**, 440–451.

[9] FERGUSON, M. J. AND AMINETZAH, Y. J. (1985) Exact results for nonsymmetric token ring systems. *IEEE Trans. Comm.* **33**, 223–231.

[10] GITTINS, J. C. (1979) Bandit processes and dynamic allocation indices. *J. R. Statist. Soc.* B **41**, 148–177.

[11] GITTINS, J. C. (1989) *Multi-armed Bandit Allocation Indices.* Wiley, New York.

[12] GITTINS, J. C. AND JONES, D. M. (1974) A dynamic allocation index for the sequential design of experiments. *Progress in Statistics*, ed. J. Gani et al., pp. 241–266. North-Holland, Amsterdam.

[13] GLAZEBROOK, K. D. (1980) On stochastic scheduling with precedence relations and switching cost. *J. Appl. Prob.* **17**, 1016–1024.

[14] GLAZEBROOK, K. D. AND GITTINS, J. C. (1981) On single-machine scheduling with precedence relations and linear or discounted costs. *Operat. Res.* **29**, 161–173.

[15] GUPTA, D., GERCHAK, Y. AND BUZACOTT, J. A. (1987) On optimal priority rules for queues with switchover costs. Preprint, Dept. of Management Sciences, University of Waterloo.

[16] HARRISON, J. M. (1975) Dynamic scheduling of a multi-class queue: discount optimality. *Operat. Res.* **23**, 270–282.

[17] HOFRI, M. AND ROSS, K. W. (1987) On the optimal control of two queues with server setup times and its analysis. *SIAM J. Comput.* **16**, 399–420.

[18] KLIMOV, G. P. (1974) Time sharing service systems I. *Theory Prob. Appl.* **19**, 532–551.

[19] KLIMOV, G. P. (1978) Time sharing service systems II. *Theory Prob. Appl.* **23**, 314–321.

[20] LAI, T. L. AND YING, Z. (1988) Open bandit processes and optimal scheduling of queueing networks. *Adv. Appl. Prob.* **20**, 447–472.

[21] MONMA, C. L. AND POTTS, C. N. (1989) On the complexity of scheduling with batch setup times. *Operat. Res.* **37**, 798–804.

[22] MURATA, M. AND TAKAGI, H. (1986) Mean waiting times in nonpreemptive priority $M/G/1$ queues with server switchover time. In *Teletraffic Analysis and Computer Performance Evaluation*, ed. O. J. Boxma, J. W. Cohen, and H. C. Tijms, pp. 395–407, Elsevier, Amsterdam.

[23] NAIN, P. (1989) Interchange arguments for classical scheduling problems in queues. *Sys. Control Lett.* **12**, 177–184.

[24] NAIN, P., TSOUCAS, P. AND WALRAND, J. (1989) Interchange arguments in stochastic scheduling. *J. Appl. Prob.* **27**, 815–826.

[25] ROSS, S. (1983) *Introduction to Stochastic Dynamic Programming*. Academic Press, New York.

[26] SIDNEY, J. B. (1975) Decomposition algorithms for single-machine sequencing with precedence relations and deferral costs. *Operat. Res.* **23**, 283–298.

[27] SYKES, J. S. (1970) Simplified analysis of an alternating priority queueing model with set up times. *Operat. Res.* **18**, 1182–1192.

[28] TAKAGI, H. (1986) *Analysis of Polling Systems*. MIT Press, Cambridge, MA.

[29] VARAIYA, P., WALRAND, J. AND BUYUKKOC, C. (1985) Extensions of the multi-armed bandit problem. *IEEE Trans. Autom. Control* **30**, 426–439.

[30] WALRAND, J. (1988) *An Introduction to Queueing Networks*. Prentice Hall, Englewood Cliffs, NJ.

[31] WHITTLE, P. (1980) Multi-armed bandits and the Gittins index. *J. R. Statist. Soc.* B **42**, 143–149.

[32] WHITTLE, P. (1981) Arm-acquiring bandits. *Ann. Prob.* **9**, 284–292.