OPTIMAL JOB SCHEDULING OF $M/GI/1$ QUEUE
WITH FEEDBACK: THE DISCOUNTED CASE

Shoichi Nishimura
University of Tsukuba

Abstract   A non-preemptive $M/GI/1$ queue with several job classes is considered. At the completion of the
service time the multiple feedback occurs. The objective is to maximize the expected discounted reward with the
infinite horizon. Using the Harrison's method, the model is formulated as a bandit problem and its optimal policy
is characterized by the index rule. Next it is considered that the service time is decomposed into quantum of the
unit span. This model is a queueing system of a single feedback. Using the index rule, three type of optimal job
schedulings are discussed.

1. Introduction

We consider an $M/GI/1$ queue with several job classes. A single job is
served at a time and the served job is not interrupted by the other job, that
is, nonpreemptive. When the service is completed, the multiple feedback will
occur. The cost structure is a pure expected reward received at a decision
epoch. Our problem is to obtain the optimal job scheduling which maximizes
the expected discounted reward with the infinite horizon. Our model is formu-
lated as a bandit problem with Poisson inputs and the multiple feedback.

In a single server queueing system the optimal job scheduling which mini-
mizes the expected holding cost has been studied. The optimal policy is
characterized by a priority service discipline. When the service time is
known in a nonpreemptive queue "the shortest-processing-time-first" discipline
(SPTF) is optimal. When the expectation of the service time is known in a
nonpreemptive queue "the shortest-expected-processing-time-first" discipline
(SEPTF) is optimal. When the service time is known in a preemptive queue,

"the shortest-remaining-processing-time-first" discipline is optimal [16].

When the distributions of several job classes are known, the optimal priority is obtained by the index rule. For an average criterion Sevcik [17] studies no arrival case and Klimov [10, 11] studies an arrival case with the sengle feedback. A multiple feedback job scheduling is studied by Meilijson and Weiss [12]. For a discounted criterion Harrison [6, 7] studies a non-preemptive $M/G/1$ queue without the feedback.

A bandit problem is constructed by independent projects. We decide to work on one of these projects, we receive a reward and the next state of the project is determined by the transition probability. Gittins [4] proves that the optimal policy is given by a dynamic allocation index (DAI) and shows the several applications including job scheduling. Whittle [20, 21, 22] obtains the integral expression of the expected discounted reward and proves the optimality in both no arrival and arrival cases. Many applications of the bandit problem are discussed in [2, 5, 8, 13, 14, 19].

In Section 2, we introduce our model. In Section 3, the transform formula of the holding cost to the immediate reward is obtained. From this formula holding cost feedback problems can be represented by a bandit problem. In Section 4 for a fixed ordering set, the expected discounted reward is obtained. In Section 5, if the index is monotone decreasing with respect to the ordering set the integral representation of the expected reward is obtained and the optimality of DAI is proved. In Section 6, it is considered that the service time is decomposed into quantum of the unit span. This model is a queueing system of a single feedback. Using the index rule, three types of optimal job schedulings are discussed. For example, if the exiting probability is unimodal, the optimal policy is one of the multilevel processor-sharing scheduling algorithm.

2. Model

We consider an $M/GI/1$ queue whose arrival jobs are distinguished into several classes. These job classes are numbered as $k=1,2,...$ and the set of all job classes is denoted as K. The set K is finite or countably infinite. Let n_k be the number of class k jobs waiting in the queue including the job being served. The state of the system is represented by $s=(n_1,...,n_k,...)$ and the state space is $S = \{s=(n_1,...,n_k,...): \sum_{k \in K} n_k < \infty\}$. Suppose that arrivals of several class jobs are independent Poisson processes and the arrival rate of class k jobs is $\lambda_k (\Lambda_K = \sum_{k \in K} \lambda_k < \infty)$. Let $F_k(t)$ be the distribution function of the

service time X_k for class k jobs with the Laplace-Stieltjes transform (LST)
$\Psi_k(\beta) = \int_0^\infty e^{-\beta t} dF_k(t)$. We assume $\inf_k EX_k > 0$. For a fixed β, we simply denote
it as Ψ_k.

We assume that one job is served at a time and its service duration is
not interrupted by the other job, that is, the non-preemptive discipline.
For the state s, the set of available actions is $\{0\} \cup \{k : n_k > 0\}$, where the
action $k(>0)$ is to serve a class k job and 0 represents the idle action. At
the completion of the service, a multiple feedback occurs immediately. Let b_i
be the number of class i jobs being added to waiting jobs by feedback and
$b = (b_1, \ldots b_i, \ldots)$ be the multiple feedback vector. Denote by $P(b|k,x)$ the
probability of b under the condition that a class k job is served and its
service time is x. Assume that the expected number of feedback jobs is uni-
formly bounded, i.e. $\sum_b (\sum_i b_i) P(b|k,x) \leq B$. If $\sum_i b_i = 0$ or 1, this is a single
feedback case discussed in Section 6. We formulate our model as a semi-Markov
decision process. Assume that the decision epoch is the completion of the
service time, or the arrival time of a new job only when the idle action is
chosen. Let us put that for $k > 0$, $\delta_k = (0, \ldots, 0, 1, 0 \ldots)$ is the kth unit
vector and $\delta_0 = (0, \ldots, 0, \ldots)$ is the zero vector. If a class k job is served
and during this service time, the arrival job vector according to Poisson
processes is $a = (a_1, \ldots, a_k, \ldots)$, then the state of next decision epoch is
$s - \delta_k + a + b$. If the idle action is chosen and the first arrival job is in class
i, then the next state is $s + \delta_i$.

The cost structure in our model is a pure reward r_k, which is immediately
received at a decision epoch when a class k job is chosen to be served. In
Section 3, we discuss the reduction of a both reward and holding cost problem
to a pure reward problem. To simplify the notation we put $r_0 = 0$. Let $V_\pi(s)$
denote the expected β-discounted reward of an infinite horizon with the initial
state s, when a policy π is employed. Let $T(n)$ and $k(n)$ $(n=1,2,\ldots)$ be the nth
decision epoch and the nth action at $T(n)$, respectively. Then we have

$$(2.1) \qquad V_\pi(s) = E_\pi \left[\sum_{n=1}^\infty e^{-\beta T(n)} r_{k(n)} \mid s \right] .$$

Our problem is to obtain an optimal policy which maximizes $V_\pi(s)$. We will
prove that an optimal policy is characterized by the priority service rule
whose order is determined by an index.

Our model is a bandit problem whose job classes are countably infinite
or finite. Jobs arrive according to independent Poisson streams and at the
completion of process time the multiple feedback occurs. The criterion is to
maximize the expected discounted reward.

3. Reward and Holding Cost Case

In this section we prove the reduction of a both reward and holding cost problem to a pure reward problem. Such an approach to a stochastic job scheduling without the feedback is proved in Bell [1], Stidham and Prabhu [18] and Harrison [6]. Using the Harrison's argument we prove this in feedback case.

Suppose that r_k^* be the reward when the service of class k job is completed and h_k^* be the holding cost for each unit of time when a class k job stays in the system. As a natural assumption we put that both $\sup_k |r_k^*|$ and $\sup_k |h_k^*|$ are finite. The reduction of a general problem (r_k^* and h_k^*) to a pure reward problem (r_k) is proved.

We define right continuous step functions as follows.

$Q_k(t)$ = the number of class k jobs in a queue waiting or being served at time t.

$A_k(t)$ = the number of class k jobs that arrive by a Poisson stream with the rate λ_k before time t, not including initial jobs N_k.

$B_k(t)$ = the number of class k jobs that arrive by the multiple feedback before t.

$D_k(t)$ = the number of class k jobs whose service is completed before time t.
From these definitions

$$(3.1) \qquad Q_k(t) = N_k + A_k(t) + B_k(t) - D_k(t) \ .$$

From the renewal theorem it follows that for sufficiently large t,

$$(3.2) \qquad \frac{1}{t} E \sum_k D_k(t) \le \sup_k \frac{1}{EX_k} + 1 < \infty$$

and

$$(3.3) \qquad \frac{1}{t} E \sum_k B_k(t) \le [\sup_k \{1/EX_k\}+1] \sup_{k,x} \sum_b (\sum_i b_i) P(b|k, x) < \infty,$$

where X_k is the service time of a class k job. These inequalities imply that $E \sum_k D_k(t)$ and $E \sum_k B_k(t)$ are bounded by a linear function of time t. Under a policy π, the total expected discounted reward of infinite horizon with the initial state s in a general problem is given by

$$(3.4) \qquad V_\pi^*(s) = E_\pi [\sum_k r_k^* \int_0^\infty e^{-\beta t} dD_k(t) - \sum_k h_k^* \int_0^\infty e^{-\beta t} Q_k(t) dt] \ .$$

From the partial integration we have,

$$E\int_0^\infty e^{-\beta t}Q_k(t)dt = N_k/\beta + E[\int_0^\infty e^{-\beta t}A_k(t)dt - \beta^{-1}[e^{-\beta t}B_k(t)]_0^\infty$$

(3.5)

$$+ \beta^{-1}\int_0^\infty e^{-\beta t}dB_k(t) + \beta^{-1}[e^{-\beta t}D_k(t)]_0^\infty - \beta^{-1}\int_0^\infty e^{-\beta t}dD_k(t)].$$

From (3.2) and (3.3), the third and the fifth terms in the right-hand side are zero and the second term is $E\int_0^\infty e^{-\beta t}A_k(t)dt = \lambda_k/\beta^2$. Substituting (3.5) into (3.4) we get that

(3.6) $\qquad V_\pi^*(s) = E_\pi[\sum_k (r_k^*+h_k^*/\beta)\int_0^\infty e^{-\beta t}dD_k(t) - \sum_k (h_k^*/\beta)\int_0^\infty e^{-\beta t}dB_k(t)] - H$,

where $\qquad H = \sum_k (h_k^*/\beta)(N_k+\lambda_k/\beta)$.

We transform the reward r_k^* at the completion of service and the holding cost h_k^* to the pure immediate reward r_k at the decision epoch such that

(3.7)
$$r_k = \int_0^\infty e^{-\beta x}[(r_k^*+h_k^*/\beta) - \sum_b (\beta^{-1}\sum_i b_i h_i^*)P(b|k,x)]dF_k(x)$$

$$= \Psi_k(r_k^*+h_k^*/\beta) - \int_0^\infty e^{-\beta x}[\sum_b (\beta^{-1}\sum_i b_i h_i^*)P(b|k,x)]dF_k(x) ,$$

where the last term in (3.7) is the expected discounted holding cost of multiple feedback jobs incurred in advance. In a bandit problem a pure reward is assumed and the frozen projects contribute no reward. From this transformation the job scheduling which minimizes the expected linear holding cost, is formulated as a bandit problem. As a tax problem, similar discussion is given in [19].

(3.8)
$$V_\pi^*(s) = E_\pi[\sum_{n=1}^\infty e^{-\beta T(n)}r_{k(n)}] - H$$

$$= V_\pi(s) - H$$

Since H is independent of the policy π, using (3.7) the general problem and the pure reward problem are equivalent and the reduction is completed.

4. Ordering Set and Index

First we consider the priority service discipline whose order is determined by an ordering set Γ. For any i and $j(i \neq j)$ in K, one of (i,j) and (j,i) is in Γ. If $(i,j) \in \Gamma$ then we interpret that job i should be served before job j. Therefore Γ has a transitive property such that $(i,j) \in \Gamma$ and $(j,k) \in \Gamma$ imply $(i,k) \in \Gamma$. Let $Y_k = \{i: (i,k) \in \Gamma\}$ be the set of higher priority jobs i than k. And also we put $Z_k = Y_k \cup \{k\}$. If the set $J \subset K$ satisfies that $k \in J$ and

$(i,k) \in \Gamma$ imply $i \in J$, we say that J is a cut-off set of Γ. If J is a cut-off set, there is a cut-off point of ording Γ and any job whose priority is higher than the cut-off point is contained in J. It is trivial that both Y_k and Z_k are cut-off sets. In this section we fix the ordering Γ and obtain the job index.

Denote by $B(s,J)$ the first epoch when all jobs in J are cleared under the initial state s. Put $\xi(s,J) = E[e^{-\beta B(s,J)}]$ and in particular $\alpha(i,J) = \xi(\delta_i, J)$. Since each service time is independent we get $\xi(s,J) = \prod\limits_{i \in J} \alpha(i,J)^{n_i}$. Let $U(s,J)$ be the expected discounted reward during $B(s,J)$. For the empty set ϕ we put $\xi(s,\phi) = 1$ and $U(s,\phi) = 0$. For any $s \in S$ we introduces $s' = (n_1, \ldots, n_{k-1}, \infty, n_{k+1}, \ldots)$ and $\bar{s} = (0, \ldots, 0, \infty, 0, \ldots)$. Let us define the index of the class k job as

$$
\begin{aligned}
C_k &\equiv U(\bar{s}, Z_k) \\
&= U(\delta_k, Z_k)/(1-\alpha(k, Z_k)) \ .
\end{aligned}
$$
(4.1)

From the assumption $\inf_k EX_k > 0$, we have that $C \equiv \sup_k |r_k|/(1-\sup_k \psi_k) \geq \sup_k |C_k|$ or C_k is uniformly bounded. Using the index C_k, Harrison [6, 7] represents the discounted expected reward of an $M/GI/1$ job scheduling problem without feedback. His method can be applied to an multiple feedback problem. Suppose that the initial state is s'. First all jobs whose priority is higher than k including arrival jobs from Poisson streams and multiple feedback, are served. During this time interval $B(s', Y_k)$, the expected discounted reward is $U(s', Y_k) = U(s, Y_k)$ and the LST of the busy period is $\xi(s, Y_k) = \xi(s', Y_k)$. The next reward is C_k. Then the total expected discounted reward with the initial state s' is

$$U(s', Z_k) = U(s, Y_k) + \xi(s, Y_k)C_k \ .$$

Using the same discussion we also have

$$U(s', Z_k) = U(s, Z_k) + \xi(s, Z_k)C_k \ .$$

Finally we get

(4.2) $$U(s, Z_k) - U(s, Y_k) = (\xi(s, Y_k) - \xi(s, Z_k))C_k \ .$$

Lemma 1. For any cut-off set J, we have

(4.3) $$U(s, J) = \sum_{i \in J} (\xi(s, Y_i) - \xi(s, Z_i))C_i \ .$$

Proof: First we assume that J is finite. Put $J = \{i(1), \ldots, i(l)\}$ such that for all $u=1, \ldots, l$, $J_u = \{i(1), \ldots, i(u)\}$ are cut-off sets, that is,

$i(u)$ $(u=1, \ldots, l)$ is the service order according to Γ. Then we have

$$U(s, J) = \sum_{u=1}^{l} (\xi(s, J_{u-1}) - \xi(s, J_u))c_{i(u)}$$

$$= \sum_{i \in J} (\xi(s, Y_i) - \xi(s, Z_i))c_i \ .$$

Next suppose that $J = \{i(u); u=1,2,\ldots\}$ is an infinite cut-off set. But $i(u)$ $(u=1,2,\ldots)$ is not necessarily the same as the order of Γ. For any cut-off set J' we define $N(J')$ as the number of decisions during $B(s, J')$. For each sample path,

$$\sum_{u=1}^{l} (N(Z_{i(u)})^{\cdot} - N(Y_{i(u)})) \to N(J) \quad (\text{as } l \to \infty)$$

Since $|c_k|$ is uniformly bounded by C, the right hand side in (4.3) is absolutely convergent. It follows from the dominated convergence theorem we have

$$\sum_{u=1}^{l} (\xi(s, Y_{i(u)}) - \xi(s, Z_{i(u)}))c_{i(u)} = E[\sum_{u=1}^{l} \sum_{n=N(Y_{i(u)})+1}^{N(Z_{i(u)})} e^{-\beta T(n)} r_{k(n)}]$$

$$\to E[\sum_{n=1}^{N(J)} e^{-\beta T(n)} r_{k(n)}] \quad (\text{as } l \to \infty)$$

$$= U(s, J)$$

Thus the desired result is obtained.

Fix a cut-off set L and let π be a priority policy such that for any $i \in L$ the service order is determined by Γ and any job in class $i \in K-L$ is never served. Let $U*(0,L)$ denote the expected discounted reward under a policy π during the initial busy cycle when the initial state is empty. Let $\Lambda_L \equiv \sum_{k \in L} \lambda_k$ be the Poisson arrival rate of the set L. Then

$$U*(0,L) = \Lambda_L/(\Lambda_L+\beta) \times \sum_{k \in L} \lambda_k U(\delta_k, L)/\Lambda_L$$

$$= \sum_{k \in L} \lambda_k U(\delta_k, L)/(\Lambda_L+\beta) \ .$$

We define $\alpha(L) \equiv \sum_{k \in L} \lambda_k \alpha(k,L)/\Lambda_L$ as the LST of the busy period with the dummy variable β. Then the LST of the busy cycle is $\Lambda_L \alpha(L)/(\Lambda_L+\beta)$ and

$$V_\pi(0) = \sum_{n=0}^{\infty} [\Lambda_L \alpha(L)/(\Lambda_L+\beta)]^n U*(0, L)$$

$$= \sum_{k \in L} \lambda_k U(\delta_k, L)/(\beta+\Lambda_L-\Lambda_L \alpha(L)) \ .$$

Theorem 2. We have

$$V_\pi(s) = U(s, L) + \xi(s, L)V_\pi(0) ,$$

where $U(s, L) = \sum_{i \in L} (\xi(s, Y_i) - \xi(s, Z_i))C_i$ and

$$V_\pi(0) = \sum_{k \in L} \lambda_k U(\delta_k, L)/(\beta + \Lambda_L - \Lambda_L \alpha(L)) .$$

Using the policy improvement technique, the optimality is proved. For the initial state $s = (n_1, \ldots, n_k, \ldots)$, fix an action k such that $k=0$ or n_k is positive. Let π' be a policy such that if $k \in K$ then one job in class k is served and if $k=0$ than the idle action is chosen until a new arrival of any job in K occurs and thereafter according to the policy π all jobs in a cut-off set L are served and all jobs in $K-L$ are never served. As was defined under the policy π, for the policy π' we put

$B_k(s, J)$ = the initial busy period until the first class k job and all
                jobs in J are served.

$\xi_k(s, J) = E[e^{-\beta B_k(s, J)}]$,

and

$U_k(s, J)$ = the expected discounted reward during $B_k(s, J)$ under the policy
                π',

where $B_0(s, J)$ is equal to the first arrival time of any job in K plus the initial all jobs in J are cleared. Especially we put

(4.4) $\alpha_k(J) = \xi_k(\delta_k, J)$ $(k \in K)$

and

(4.5) $\alpha_0(J) = \xi_0(\delta_0, J) \equiv \{ \sum_{i \in J} \lambda_i \alpha_i(J) + \Lambda_K - \Lambda_J \}/(\beta + \Lambda_K)$.

If $J = \phi$, then $\alpha_k(\phi) = \Psi_k (k \neq 0)$ and $\alpha_0(\phi) = \Psi_0 \equiv \Lambda_K/(\beta + \Lambda_K)$. From these difinitions we have

(4.6) $\xi_k(s, J) = \begin{cases} \xi(s, J) & k \in J \\ \alpha_k(J)\xi(s, J) & k \in K-J \text{ or } k=0 \end{cases}$

Let a and b be a Poisson arrival vector and a multiple feedback vector, respectively. For $y = a+b$, let $\bar{p}(y|k, x)$ be the probability of y under the condition that the service time of a class k job is x. Then we get that for $k \in K$

$$U_k(s, L) = r_k + \int_0^\infty e^{-\beta x} \sum_y \bar{p}(y|k,x) U(s-\delta_k+y, L) dF_k(x)$$

(4.7)
$$= r_k + \int_0^\infty e^{-\beta x} [\sum_y \bar{p}(y|k,x) \sum_{i \in L} \{\xi(s-\delta_k+y, Y_i)$$
$$- \xi(s-\delta_k+y, Z_i)\} C_i] dF_k(x)$$

$$= r_k + \sum_{i \in L} \{\xi_k(s, Y_i) - \xi_k(s, Z_i)\} C_i \ .$$

It is easily proved that for $k=0$ the above equation is also satisfied. Moreover, we define $V_k(s)$ as the expected discounted reward for the infinite horizon problem under the policy π'.

Theorem 3. We have

$$V_k(s) = U_k(s, L) + \xi_k(s, L) V_\pi(0) \ ,$$

where

$$U_k(s, L) = r_k + \sum_{i \in L} \{\xi_k(s, Y_i) - \xi_k(s, Z_i)\} C_i \ .$$

Another expression of C_k defined by (4.1) is

(4.8)
$$C_k = U_k(\delta_k, Y_k)/(1-\alpha_k(Y_k))$$
$$= [r_k + \sum_{i \in Y_k} \{\alpha_k(Y_i)-\alpha_k(Z_i)\} C_i]/(1-\alpha_k(Y_k)) \ .$$

In this expression C_k is represented by higher indices of i than k.

5. Optimality

In this section we obtain an optimal policy, which is characterized by an index rule. At first we assume that the index C_i is monotone nonincreasing with respect to the ordering Γ. That is, $(i,j) \in \Gamma$ implies $C_i \geqq C_j$. Next, suppose that L is the largest cut-off set with the lowest index M ($M = \inf_{j \in L} C_j$ and $L = \{j: C_j \geqq M\}$).

For an infinite cut-off set J we prove the following Lemma.

Lemma 4. Suppose that the sequence of cut-off sets $\{J_l\}$ ($l=1,2,...$) is monotone increasing (or decreasing) and $\lim_{l \to \infty} J_l = J$. Then we have

(5.1)
$$\lim_{l \to \infty} \xi(s, J_l) = \xi(s, J) \ .$$

Proof: For an increasing sequence we prove this lemma. In the decreasing case the proof is the same as this. Since $\overset{\infty}{\underset{l=1}{U}} \{\omega: B(s, J_l) \leq t\} = \{\omega: B(s, J) \leq t\}$,

we have $\lim\limits_{l \to \infty} P\{B(s, J_l) \le t\} = P\{B(s, J) \le t\}$. It follows from the continuity of

LST that $\lim\limits_{l \to \infty} \xi(s, J_l) = \xi(s, J)$.

If C_i is monotone nonincreasing with respect to Γ, instead of a cut-off set J, we define*-function of the value m as

$$\xi^*(s, m) = \inf_{C_j \ge m} \xi(s, Z_j), \quad \alpha^*(k, m) = \inf_{C_j \ge m} \alpha(k, Z_j)$$

and

$$\xi_k^*(s, m) = \inf_{C_j \ge m} \xi_k(s, Z_j), \quad \alpha_k^*(m) = \inf_{C_j \ge m} \alpha_k(Z_j) .$$

For any m $(\sup_j C_j, C)$, $\xi^*(s, m) = \alpha^*(k, m) = 1$ and $\xi_k^*(s, m) = \alpha_k^*(m) = \Psi_k$. These four functions are left continuous nondecreasing step functions. For example, a jump of $\xi^*(s, m)$ at m is

$$\xi^*(s, m+) - \xi^*(s, m) = \sup_{C_j = m} \xi(s, Y_j) - \inf_{C_j = m} \xi(s, Z_j)$$

In Theorem 2 $U(s, L)$ is represented by the Stieltjes integral of the monotone function $u^*(s, m)$ and from partial integration we have

$$(5.2) \quad \begin{aligned} U(s, L) &= \int_M^C m \, d\xi^*(s, m) \\ &= C - \xi^*(s, M)M - \int_M^C \xi^*(s, m) dm . \end{aligned}$$

Whittle [21] modifies the bandit problem so as to allow the additional option of retiring with the reward M. If in our model the retiring option is employed with the reward M when all jobs in L is cleared, the expected discounted reward of the modified process is

$$U(s, L) + \xi^*(s, M)M = C - \int_M^C \prod_{i \in L} \alpha^*(s, m)^{n_i} dm$$

which is equivalent to (14) in [21].

For the policy π', a jump at m is

$$\xi_k^*(s, m+) - \xi_k^*(s, m) = \sup_{C_j = m} \xi_k(s, Y_j) - \inf_{C_j = m} \xi_k(s, Z_j)$$

and from (4.6)

$$\xi_k^*(s, m) = \begin{cases} \xi^*(s, m) & C_k \ge m \\ \alpha_k^*(m) \xi^*(s, m) & C_k < m \text{ or } k=0 \end{cases}$$

If $k \in L$, then $C \ge C_k \ge M$ and from Theorem 3.

$$U_k(s, L) = r_k + \int_M^C m \, d\xi_k^*(s, m)$$

(5.3)

$$= r_k + \Psi_k C - \xi^*(s, M)M - \int_{C_k}^C \alpha_k^*(m)\xi^*(s, m)dm - \int_M^{C_k} \xi^*(s, m)dm .$$

If k $K-L$ or $k=0$, then we also get

(5.4) $$U_k(s, L) = r_k + \Psi_k C - \alpha_k^*(M)\xi^*(s, M)M - \int_M^C \alpha_k^*(m)\xi^*(s, m)dm .$$

Lemma 5. If $n_k > 0$ and $k \epsilon L$ then

(5.5) $$V_\pi(s) - V_k(s) = -r_k + (1-\Psi_k)C - \int_{C_k}^C (1-\alpha_k^*(m))\xi^*(s, m)dm$$

and if $n_k > 0$ and $k \epsilon K-L$, or $k=0$ then

$$V_\pi(s) - V_k(s) = -r_k + (1-\Psi_k)C - (1-\alpha_k^*(M))\xi^*(s, M)(M-V_\pi(0))$$

$$- \int_M^C (1-\alpha_k^*(m))\xi^*(s, m)dm .$$

Proof: From Theorem 2 and 3, we have

$$V_\pi(s) - V_k(s) = U(s, L) - U_k(s, L) + (\xi(s, L) - \xi_k(s, L))V_\pi(0) .$$

Since for $k \epsilon L$, $\xi(s, L) = \xi_k(s, L)$ then it follow from (5.2) and (5.3) that
(5.5) is obtained. Also if k $K-L$ or $k=0$, then from (5.2) and (5.4) (5.6) is
obtained.

Generalizing the index C_k in (4.8), for $k \epsilon K-L$ or $k\equiv0$, we define $C_k(L)$ as

$$C_k(L) \equiv U_k(\delta_k, L)/(1-\alpha_k(L))$$

(5.7)

$$= \{r_k + \Psi_k C - \alpha_k^*(M)M - \int_M^C \alpha_k^*(m)dm\}/(1-\alpha_k^*(M)) .$$

Since from the definition $U_0(\delta_0, L) = \sum_{k \epsilon L} \lambda_k U(\delta_k, L)/(\beta+\Lambda_K)$ and $1-\alpha_0(L) = (\beta+\Lambda_L-\Lambda_L\alpha(L))/(\beta+\Lambda_K)$, then

(5.8) $$C_0(L) = \sum_{k \epsilon L} \lambda_k U(\delta_k, L)/(\beta+\Lambda_L-\Lambda_L\alpha(L))$$

$$= V_\pi(0) .$$

Theorem 6. Suppose that π is a priority policy with the ordering set Γ
and the cut-off set L such that for $k \epsilon L$, C_k is monotone nonincreasing with
respect to Γ and

$$M \equiv \inf_{k \epsilon L} C_k \geqq C_0(L) \geqq \sup_{k \epsilon K-L} C_k(L)$$

Then π is optimal.

Proof: We use the policy improvement technique. To prove this it is sufficient to show that for any fixed $s \neq \delta_0$ and any $k(n_k > 0$ or $k=0)$, $V_\pi(s) - V_k(s) \geq 0$. Since for $n_k > 0$, $\xi*(s, m) \leq \xi*(\delta_k, m)$ in Lemma 5, we have that for $k \in L$, $V_\pi(s) - V_k(s) \geq V_\pi(\delta_k) - V_k(\delta_k) = 0$. And since for $k \in K-L$ or $k=0$, $\xi*(\delta_k, m) = 1$ $(C \geq m \geq M)$, then

$$V_\pi(s) - V_k(s) \geq V_\pi(\delta_k) - V_k(\delta_k)$$

$$= -r_k + (1-\Psi_k)C - (1-\alpha_k*(M))(M-V_\pi(0)) - \int_M^C (1-\alpha_k*(m))dm$$

$$= (1-\alpha_k*(M))(C_0(L)-C_k(L)) \geq 0 \ ,$$

where the third equality is derived from (5.7) and (5.8). This completes the proof.

In order to contruct the optimal ordering set Γ and cut-off set L, the generalized index $C_i(L)$ is useful.

Proposition 7.   Suppose that i and $j \in (K-L) \cup \{0\}$ and $C_j(L) \leq C_i(L)$. Then

(5.9) $C_j(L) \leq C_j(L \cup i) \leq C_i(L)$

and

(5.10) $C_j(L) \leq C_i(L \cup j) \leq C_i(L)$

Proof: For $i \neq 0$ and $j \neq 0$ we get

$$(1-\alpha_j(L \cup i))C_j(L \cup i) = U_j(\delta_j, L) + (\alpha_j(L)-\alpha_j(L \cup i))C_i(L)$$

$$= (1-\alpha_j(L))C_j(L) + (\alpha_j(L)-\alpha_j(L \cup i))C_i(L)$$

$$= (1-\alpha_j(L \cup i))C_i(L) - (1-\alpha_j(L))(C_i(L)-C_j(L))$$

or

$$C_i(L) - C_j(L \cup i) = (C_i(L)-C_j(L))(1-\alpha_j(L))/(1-\alpha_j(L \cup i)) \ .$$

Since $0 \leq (1-\alpha_j(L))/(1-\alpha_j(L \cup i)) \leq 1$ we get (5.9). Exchanging the role of i and j we also have (5.10). If one of i and j is 0, we can prove this using the same discussion.

The optimal ordering Γ is constructed by a successive addition of the highest generalized index job until $k=0$ is chosen. For a finite set K the algorithm to obtain the optimal policy is shown [7]. In the next section we apply Prop. 7 to an $M/GI/1$ queue.

6. Application to M/GI/1 Queue with Decomposable Quantum Service Times

In this section we consider the following M/GI/1 queue. Customers arrive at a single-server station as a Poisson stream with the rate λ. Each arriving customer has a discrete service time $k=1, 2, \ldots$ with the probability g_k ($\sum_{k=1}^{\infty} g_k = 1$). We assume that the service time is decomposed into quantum of the unit span. Let a class k job be a customer who has served $k-1$ quanta and is waiting for kth quantum. The probability of a class k job to exit is $p_k = g_k / \sum_{i=k}^{\infty} g_i$ and its probability of the single feedback as a class $k+1$ job is $1-p_k$. For example, the CPU serves customers according to a round-robin discipline with the fixed quantum size 1. The problem is to obtain the optimal quantum to be served in the waiting customers.

The LST of the service time of a class k job is $\Psi_k = e^{-\beta} \equiv \Psi$. As a cost structure we assume that, $\gamma_k^* = 0$ and the holding cost for each unit time is constant. To simplify the notation we put $h_k^* = \beta e^{\beta}$. Then the immediate reward in (3, 7) is

$$\gamma_k = \Psi_k h_k^* / \beta - \Psi_k h_{k+1}^* (1-p_k) / \beta = p_k .$$

Our problem is to obtain the optimal job scheduling which minimizes the expected discounted holding cost.

At first we consider no arrival case and next we consider a Poisson arrival case. For $i<j$ let $\Psi_{i,j}$ be the LST of the service time from ith quantum to j. We recursively define as $\Psi_{i,i} = \Psi$ and

$$\Psi_{i,j} = p_i \Psi + (1-p_i) \Psi \Psi_{i+1,j} .$$

During this interval the expected discounted reward is $\gamma_{i,i} = p_i$ and

$$\gamma_{i,j} = p_i + (1-p_i) \Psi \gamma_{i+1,j} .$$

The service index from i to j is defined as

$$c_{i,j} = \gamma_{i,j} / (1-\psi_{i,j}) .$$

The index of a class i job without arrival is

(6.1) $\qquad c_i' \equiv \sup_{i \leq j} c_{i,j} .$

Gittins [4] shows the following results: If the exiting probability p_k is nonincreasing, then c_k' is nonincreasing and this is the deteriorating case. If p_k is nondecreasing, then c_k' is nondecreasing and this is the improving

case. Glazebrook [5] also prove this result when a single machine sometimes breaks down. In the case of a Poisson arrival we will obtain the optimal policy when p_k is monotone nonincreasing, nondecreasing or unimodal.

Case 1. Suppose that p_k is nonincreasing. Let $\Gamma_1 = \{(i,j): i<j\}$ be the deteriorating ordering set and $Y_k = \{1, \ldots, k-1\}$ be its cut-off set. Under the initial condition $s=\delta_1$, $\alpha(1, Y_k)$ is the LST of the busy period $B(\delta_1, Y_k)$ and $U(\delta_1, Y_k)$ is its expected discounted reward. Next for the initial condition $s=\delta_i$ $(i \in Y_k)$, $\alpha_i(Y_k)$ is the LST of the busy period $B_i(\delta_i, Y_k)$. During the unit interval, $e^{-\lambda}\lambda^a/a!$ is the probability that the number of arriving class 1 jobs is a. And

$$\alpha_i(Y_k) = \Psi \sum_{a=0}^{\infty} e^{-\lambda}\lambda^a \alpha(1, Y_k)^a/a!$$

$$= \Psi \exp\{-\lambda(1-\alpha(1, Y_k))\}$$

is independent of i. We also have

$$U_i(\delta_i, Y_k) = p_i + \Psi \sum_{a=1}^{\infty} \frac{e^{-\lambda}\lambda^a}{a!} U(\delta_1, Y_k)[1+\cdots+\alpha(1, Y_k)^{a-1}]$$

$$= p_i + \frac{U(\delta_1, Y_k)}{1-\alpha(1, Y_k)}[\Psi-\alpha_i(Y_k)] ,$$

where the second term of the last equation is independent of i. Then the generalized index

$$C_i(Y_k) = U_i(\delta_i, Y_k)/(1-\alpha_i(Y_k))$$

is a nonincreasing function of $i \in \{k, k+1, \ldots\}$. From Prop. 7, C_k is recursively determined as $C_k = C_k(Y_k)$. Therefore C_k is nonincreasing and from Theorem 6, Γ_1 is the optimal ordering set. This ordeded round-robin is said to feedback-to-lower-priority-queue-discipline (see Schrage [16] and Brown [3]) or this is the foreground-background (FB) scheduling algorithm with unit quantum (Kleinrock [9] p.172).

Applying Prop. 7 to $C_{i,j}$ we obtain the following.

Proposition 8. For any $i \leq j < k$ we have that

$$\text{if } C_{i,j} \leq C_{i+1,k} \text{ then } C_{i,j} \leq C_{i,k} \leq C_{j+1,k}$$

and

$$\text{if } C_{i,j} \geq C_{j+1,k} \text{ then } C_{i,j} \geq C_{i,k} \geq C_{j+1,k} .$$

Case 2. Suppose that p_k is nondecreasing. From Prop. 8, we have that $C_k = C_k' = C_{k,\infty}$ which is the monotone nondecreasing function of k and $\Gamma_2 = \{(i,j): i > j\}$ is the optimal ordering set. Using the queueing terminology the FCFS discipline is optimal.

Case 3. Suppose that p_k is unimodal and the peak of p_k is $p_{i_0} = \max_k p_k$. In other words p_k is nondecreasing in $k \varepsilon \{1, \ldots, i_0\}$ and nonincreasing in $k \varepsilon \{i_0, i_0+1, \ldots\}$. From Prop. 8, for any fixed i, C_{ij} is an unimodal function of j and then C_k' is also unimodal such that C_k' is increasing in $k \varepsilon \{1, \ldots, i_0\}$ and decreasing in $k \varepsilon \{i_0, i_0+1, \ldots\}$. In no arrival case $\Gamma_3 = \{(i,j): C_i' > C_j'$ or $(C_i' = C_j'$ and $i < j)\}$ is an optimal ordering set. Next we consider an arrival case. From (4.8) the index of class 1 jobs is not effected by an arrival and let $j(1)$ be such that $C_1' = C_{1,j(1)}$ then for $1 \leq k \leq j(1)$ the optimal index C_k is independent of Poisson arrival as $C_k = C_k'$. From (4.1) and (4.8) we have $C_1 = U(\delta_1, Z_1)/(1-\alpha(1, Z_1))$ where $Z_1 = \{1, \ldots, j(1)\}$. For $k \geq j(1)+1$ we put $Y_k = \{1, \ldots, k-1\}$ and as was shown in Case 1 we get

$$\alpha_k(Y_k) = \Psi \exp\{-\lambda(1-\alpha(1, Y_k))\}$$

$$U_k(\delta_k, Y_k) = p_k + \frac{U(\delta_1, Y_k)}{1-\alpha(1, Y_k)}[\Psi - \alpha_k(Y_k)]$$

and

$$C_k \equiv U_k(\delta_k, Y_k)/(1-\alpha_k(Y_k)) \leq C_1 .$$

Using the same discussion in Case 1, for $k \geq j(1)+1$ C_k is nonincreasing. It follows that for $k=1, \ldots, j(1)$ $C_k = C_k'$ is the unimodal index without arrival and for $k=j(1)+1, \ldots, C_k = C_k(Y_k)$ is the nonincreasing index of Case 1. As a whole the order of C_k is the same as the order of C_k' and an optimal ordering set is the same as Γ_3. For any initial state s there exists only one job from class 2 to $j(1)$ after sufficiently long time spent when Γ_3 is employed. At this situation from 1 to $j(1)$ the FCFS discipline is employed and if there is no job from 1 to $j(1)$, then from $j(1)+1$ unit quantum the FB discipline is employed. This mixed scheduling algorithm is one of the multilevel processor sharing algorithm discussed in Kleinrock [9] p.177.

References

[1] C. Bell: Characterization and Computation of Optimal Policies for Operating an *M/G*/1 Queueing System with Removal Server, *Opns. Res.*, 19 (1971) 208-218.

[2] D. A. Berry and B. Fristedt: *Bandit Problems*, Chapman and Hall (1985).

[3] T. Brown: Determination of the Conditional Response for Quantum Alloca-
     tion Algorithms, *J. ACM*, 23 (1982), 448-460.

[4] J. C. Gittins: Bandit Processes and Dynamic Allocation Indices, *J. Roy.
     Stat. Soc. Ser. B.*, 41 (1979), 148-177.

[5] K. D. Glazebrook; Scheduling Stochastic Jobs on a Single Machine Subject
     to Breakdown, *Nav. Res. Logist. Q.*, 31 (1984), 251-264.

[6] J. M. Harrison: A priority Queue with Discounted Linear Cost, *Opns. Res.*,
     23 (1975), 260-269.

[7] J. M. Harrison: Dynamic Scheduling of a Multiclass Queue: Discount
     Optimality, *Opns. Res.*, 23 (1975), 270-282.

[8] T. Hirayama: Nonpreemptive Scheduling of a Finite-Source Queue with Two
     Customer Classes, *J. Oper. Res. Soc. Japan*, 30 (1987), 200-217.

[9] L. Kleinrock: *Queueing Systems: Volume II Computer Applications*, Wiley
     (1976).

[10] G. P. Klimov: Time Sharing Service Systems I, *Theory Prob. Appl.*, 19
     (1974), 532-551.

[11] G. P. Klimov: Time Sharing Service Systems II, *Theory Prob. Appl.*, 23
     (1978), 314-321.

[12] I. Meilijson, and G. Weiss: Multiple Feedback at a Single-Server Station,
     Stochastic Process, Appl., 5 (1977), 195-205.

[13] Z. Rosbers: Process Scheduling in a Computer System, *IEEE Trans. on
     Comp.*, 34 (1985), 633-645.

[14] S. M. Ross: *Introduction to Stochastic Dynamic Programming*, Academic
     Press (1983).

[15] L. E. Schrage: The Queue $M/G/1$ with Feedback to Lower Priority Queues,
     Manage. Sci., 13 (1967), 466-474.

[16] L. E. Schrage: A Proof of the Optimality of the SRPT Discipline, *Opns.
     Res.*, 16 (1968), 687-190.

[17] K. C. Sevcik: Scheduling for Minimum Total Loss Using Service Time
     Distributions, *J. ACM*, 21 (1974), 66-75.

[18] S. Stidham Jr. and N. U. Prabhu: Optimal Control of Queueing Theory,
     Lecture Notes in Economics and Mathematical Systems, 98 (1973), Springer,
     263-294.

[19] P. P. Varaiya, J. C. Walrand and C. Buyukkoc: Extensions of the Multi-
     armed Bandit Problem: The Discounted Case, *IEEE Auto. Cont.*, 30 (1985),
     426-439.

[20] P. Whittle: Multi-armed Bandits and the Gittins Index, *J. Roy. Statist.
     Soc., Ser. B.*, 42 (1980), 143-149.

[21] P. Whittle: Arm-acquiring Bandits, *Ann. Prob.* 9 (1981), 284-292.

[22] P. Whittle: *Optimization over time*, vol.1, Wiley, 1982.

Shoichi NISHIMURA: Department of
    Institute of Socio-Economic Planning,
    University of Tsukuba, Tsukuba,
    Ibaraki, 305, Japan.