

No. 318

OPTIMAL JOB SCHEDULING WITH FEEDBACK:  
THE DISCOUNTED CASE

by  
Shoichi NISHIMURA

October 1986

# OPTIMAL JOB SCHEDULING WITH FEEDBACK : THE DISCOUNTED CASE

Shoichi NISHIMURA

Institute of Socio Economic Planning

The University of Tsukuba, Ibaraki, Japan

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 [6, 7], the model is formulated as a bandit problem and its optimal policy is characterized by the index rule. The application to a discrete approximation of a preemptive  $M/GI/1$  queue is discussed.

job scheduling \* bandit problem \* multiple feedback \* discounted cost \* index rule

## 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 formulated as a bandit problem with Poisson inputs and the multiple feedback.

In a single server queueing system the optimal job scheduling which minimizes 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 simple feedback. A multiple feedback job scheduling is studied by Meilijson and Weiss [12]. For a discounted criterion Harrison [6, 7] studies a nonpreemptive  $M/G/1$  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 1, we introduce our model. In section 2, 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 3 for a fixed ordering set, the expected discounted reward is obtained. In section 4, 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 5, we apply our results to a discrete approximation of a preemptive  $M/GI/1$  queue. In three cases the optimal policy is obtained. For example, if the exiting probability is unimodal, the optimal policy is one of the multilevel processor-sharing scheduling algorithm.

## 1. Model

We consider an  $M/GI/1$  queue whose arrival jobs are distinguished into several classes. These job classes are numbered as  $k=1, 2, \dots$  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  $s=(n_1, \dots, n_k, \dots)$  and the state space is  $S=\{s=(n_1, \dots, n_k, \dots) : \sum_{k \in K} n_k < \infty\}$ . 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  $\beta$ , we simply denote it as  $\Psi_k$ .

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  $\{0\} \cup \{k : n_k > 0\}$ , the action  $k (> 0)$  is to serve a class  $k$  job and 0 represents the idle action. At the completion of the service time, a multiple feedback occurs. Let  $P(b|k, x)$  be the conditional probability of the multiple feedback vector  $b=(b_1, \dots, b_i, \dots)$  when a class  $k$  job is served and its service time is  $x$ . Assume that the expected number of feedback jobs is uniformly bounded, that is, for all  $k$  and  $x$ ,  $\sum_i (\sum_j b_j) P(b|k, x) \leq \bar{B}$ . 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  $k > 0$ ,  $\delta_k=(0, \dots, 0, 1, 0, \dots)$  is the  $k$ th unit vector and  $\delta_0=(0, \dots, 0, \dots)$  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  $a=(a_1, \dots, a_k, \dots)$ , 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 2, 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  $\beta$ -discounted reward of an infinite horizon with the initial state  $s$ , when a policy  $\pi$  is employed. Let  $T(n)$  and  $k(n) (n=1, 2, \dots)$  be the  $n$ th decision epoch and the  $n$ th action at  $T(n)$ , respectively. Then we have

$$(1.1) \quad V_\pi(s) = E_\pi \left[ \sum_{n=1}^{\infty} e^{-\beta T(n)} r_{k(n)} | s \right].$$

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.

## 2. 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]. From this reduction technique, the both reward and holding cost problem is formulated as a bandit problem (see, Varaiya, Walrand, and Buyukkoc [19]). 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  $\lambda_k$  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 time  $t$ .

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

From these definitions

$$(2.1) \quad Q_k(t) = N_k + A_k(t) + B_k(t) - D_k(t).$$

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

$$(2.2) \quad \frac{1}{t} E \sum_k D_k(t) \leq \sup_k \frac{1}{EX_k} + 1 < \infty$$

and

$$(2.3) \quad \frac{1}{t} E \sum_k D_k(t) \leq [\sup_k \{1/EX_k\} + 1] \sup_{k,x} \sum_b \left( \sum_1 b_i \right) P(b|k, x) < \infty,$$

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  $\pi$ , the total expected discounted reward of infinite horizon with the initial state  $s$  in a general problem is given by

$$(2.4) \quad V_\pi^*(s) = E_\pi \left[ \sum_k r_k^* \int_0^\infty e^{-\beta t} dD_k(t) - \sum_k h_k^* \int_0^\infty e^{-\beta t} Q_k(t) dt \right].$$

From the partial integration we have.

$$(2.5) \quad E \int_0^\infty e^{-\beta t} Q_k(t) dt = N_k/\beta + E \left[ \int_0^\infty e^{-\beta t} A_k(t) dt - \beta^{-1} \left[ e^{-\beta t} B_k(t) \right]_0^\infty + \beta^{-1} \int_0^\infty e^{-\beta t} dB_k(t) \right. \\ \left. + \beta^{-1} \left[ e^{-\beta t} D_k(t) \right]_0^\infty - \beta^{-1} \int_0^\infty e^{-\beta t} dD_k(t) \right]$$

From (2.2) and (2.3), 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 (2.5) into (2.4) we get that

$$(2.6) \quad V_{\pi}^*(s) = E_{\pi} \left[ \sum_k (r_k^* + h_k^* / \beta) \int_0^{\infty} e^{-\beta t} dD_k(t) - \sum_k (h_k^* / \beta) \int_0^{\infty} e^{-\beta t} dB_k(t) \right] - H,$$

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

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

$$(2.7) \quad r_k = \int_0^{\infty} e^{-\beta x} \left[ (r_k^* + h_k^* / \beta) - \sum_b (\beta^{-1} \sum_i b_i h_i^*) P(b|k, x) \right] dF_k(x) \\ = \Psi_k(r_k^* + h_k^* / \beta) - \int_0^{\infty} e^{-\beta x} \left[ \sum_b (\beta^{-1} \sum_i b_i h_i^*) P(b|k, x) \right] dF_k(x),$$

where the last term in (2.7) is the expected discounted holding cost of multiple feedback jobs incurred in advance. Using the expression of the decision epoch in (1.1) we have

$$(2.8) \quad V_{\pi}^*(s) = E_{\pi} \left[ \sum_{n=1}^{\infty} e^{-\beta T(n)} r_{k(n)} \right] - H \\ = V_{\pi}(s) - H$$

Since  $H$  is independent of the policy  $\pi$ , using (2.7) the general problem and the pure reward problem are equivalent and the reduction is completed.

### 3. Ordering set and index

First we consider the priority service discipline whose order is determined by a ordering set  $\Gamma$ . For any  $i$  and  $j$  ( $i \neq j$ ) in  $K$ , one of  $(i, j)$  and  $(j, i)$  is in  $\Gamma$ . If  $(i, j) \in \Gamma$  then we interpret that job  $i$  should be served before job  $j$ . Therefore  $\Gamma$  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  $\Gamma$ . If  $J$  is a cut-off set, there is a cut-off point of ordering  $\Gamma$  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  $\Gamma$  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  $\xi(s, J) = \prod_{i \in J} \alpha(i, J)^{n_i}$ . Let  $U(s, J)$  be the expected discounted reward during  $B(s, J)$ . For the empty set  $\phi$  we put  $\xi(s, \phi) = 1$  and  $U(s, \phi) = 0$ . For any  $s \in A$  we introduce  $s' = (n_1, \dots, n_{k-1}, \infty, n_{k+1}, \dots)$

and  $\bar{s} = (0, \dots, 0, \infty, 0, \dots)$ . Let us define the index of the class  $k$  job as

$$(3.1) \quad C_k \equiv U(\bar{s}, Z_k) \\ = U(\delta_k, Z_k) / (1 - \alpha(k, Z_k)).$$

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. Suppose that the initial state is  $s$ . According to  $\Gamma$ , all jobs whose priority is higher than  $k$ , including arrival jobs from Poisson streams and by the multiple feedback, are served. During this time interval  $B(s, Y_k)$  the expected discounted reward is  $U(s, Y_k)$  and the LST of this expected discounted time interval is  $\xi(s, Y_k)$ . After that  $C_k$  is obtained. 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

$$(3.2) \quad U(s, Z_k) - U(s, Y_k) = (\xi(s, Y_k) - \xi(s, Z_k)) C_k.$$

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

$$(3.3) \quad U(s, J) = \sum_{i \in J} (\xi(s, Y_i) - \xi(s, Z_i)) C_i.$$

PROOF. First we assume that  $J$  is finite. Put  $J = \{i(1), \dots, i(l)\}$  such that for all  $u = 1, \dots, l$   $J_u = \{i(1), \dots, i(u)\}$  are cut-off sets, that is,  $i(u)$  ( $u = 1, \dots, l$ ) is the service order according to  $\Gamma$ . 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  $J = \{i(u) : u = 1, 2, \dots\}$  is an infinite cut-off set. But  $i(u)$  ( $u = 1, 2, \dots$ ) is not necessarily the same as the order of  $\Gamma$ . For any cut-off set  $J'$  we define  $N(J')$  as the number of decisions during  $B(s, J')$ . For each sample path,

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

Since  $|C_k|$  is uniformly bounded by  $C$ , the right hand side in (3.3) is absolutely convergent. It follows from the dominated convergence theorem we have

$$\begin{aligned} \sum_{u=1}^l \left( \xi(s, Y_{i(u)}) - \xi(s, Z_{i(u)}) \right) C_{i(u)} &= E \left[ \sum_{u=1}^l \sum_{n=N(Y_{i(u)})+1}^{N(Z_{i(u)})} e^{-\beta T(n)} r_{k(n)} \right] \\ &\longrightarrow E \left[ \sum_{n=1}^{N(J)} e^{-\beta T(n)} r_{k(n)} \right] \quad (\text{as } l \rightarrow \infty) \\ &= U(s, J) \end{aligned}$$

Thus the desired result is obtained.

Fix a cut-off set  $L$  and let  $\pi$  be a priority policy such that for any  $i \in L$  the service order is determined by  $\Gamma$  and any job in class  $i \in K - L$  is never served. Let  $U^*(0, L)$  denote the expected discounted reward under a policy  $\pi$  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

$$\begin{aligned} U^*(0, L) &= \Lambda_L / (\Lambda_L + \beta) \times \sum_{k \in L} \lambda_k U(\delta_k, L) / \Lambda_L \\ &= \sum_{k \in L} \lambda_k U(\delta_k, L) / (\Lambda_L + \beta). \end{aligned}$$

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  $\beta$ . Then the LST of the busy cycle is  $\Lambda_L \alpha(L) / (\Lambda_L + \beta)$  and

$$\begin{aligned} V_\pi(0) &= \sum_{n=0}^{\infty} \left[ \Lambda_L \alpha(L) / (\Lambda_L + \beta) \right]^n U^*(0, L) \\ &= \sum_{k \in L} \lambda_k U(\delta_k, L) / (\beta + \Lambda_L - \Lambda_L \alpha(L)). \end{aligned}$$

**THEOREM 2.** We have

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

where  $U(s, L) = \sum_{i \in L} \left( \xi(s, Y_i) - \xi(s, Z_i) \right) C_i$  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  $s = (n_1, \dots, n_k, \dots)$ , fix an action  $k$  such that  $k=0$  or  $n_k$  is positive. Let  $\pi'$  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  $\pi$  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  $\pi$ , for the policy  $\pi'$  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  $\pi'$ ,

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

$$(3.4) \quad \alpha_k(J) = \xi_k(\delta_k, J) \quad (k \in K)$$

and

$$(3.5) \quad \alpha_0(J) = \xi_0(\delta_0, J) \equiv \left\{ \sum_{i \in J} \lambda_i \alpha_i(J) + \Lambda_K - \Lambda_J \right\} / (\beta + \Lambda_K).$$

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 definitions we have

$$(3.6) \quad \xi_k(s, J) = \begin{cases} \xi_k(s, J) & k \in J \\ \alpha_k(J) \xi_k(s, J) & k \in K - J \text{ or } k = 0 \end{cases}$$

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  $k \in K$

$$(3.7) \quad \begin{aligned} U_k(s, L) &= r_k + \int_0^\infty e^{-\beta x} \sum_y \bar{p}(y|k, x) U(s - \delta_k + y, L) dF_k(x) \\ &= r_k + \int_0^\infty e^{-\beta x} \left[ \sum_y \bar{p}(y|k, x) \sum_{i \in L} \{ \xi_k(s - \delta_k + y, Y_i) - \xi_k(s - \delta_k + y, Z_i) \} C_i \right] dF_k(x) \\ &= r_k + \sum_{i \in L} \{ \xi_k(s, Y_i) - \xi_k(s, Z_i) \} C_i. \end{aligned}$$

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  $\pi'$ .

**THEOREM 3.** We have

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

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 (3.1) is

$$(3.8) \quad C_k = U_k(\delta_k, Y_k) / (1 - \alpha_k(Y_k)) \\ = \left[ r_k + \sum_{i \in Y_k} \{\alpha_k(Y_i) - \alpha_k(Z_i)\} C_i \right] / (1 - \alpha_k(Y_k)).$$

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

#### 4. 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  $\Gamma$ . That is,  $(i, j) \in \Gamma$  implies  $C_i \geq C_j$ . Next, suppose that  $L$  is the largest cut-off set with the lowest index  $M$  ( $M = \inf_{j \in L} C_j$  and  $L = \{j : C_j \geq M\}$ ).

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

LEMMA 4. Suppose that the sequence of cut-off sets  $\{J_i\}$  ( $i=1, 2, \dots$ ) is monotone increasing (or decreasing) and  $\lim_{i \rightarrow \infty} J_i = J$ . Then we have

$$\lim_{i \rightarrow \infty} \xi(s, J_i) = \xi(s, J).$$

PROOF. For an increasing sequence we prove this lemma. In the decreasing case the proof is the same as this. Since  $\bigcup_{i=1}^{\infty} \{\omega : B(s, J_i) \leq t\} = \{\omega : B(s, J) \leq t\}$ , we have  $\lim_{i \rightarrow \infty} P\{B(s, J_i) \leq t\} = P\{B(s, J) \leq t\}$ . It flows from the continuity of LST that  $\lim_{i \rightarrow \infty} \xi(s, J_i) = \xi(s, J)$ .

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

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

and

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

For any  $m \in (\sup_j C_j, C)$ ,  $\xi^*(s, m) = \alpha^*(k, m) = 1$  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  $\xi^{*}(s, m)$  and from partial integration we have

$$(4.2) \quad U(s, L) = \int_M^C m d\xi^{*}(s, m) \\ = C - \xi^{*}(s, M)M - \int_M^C \xi^{*}(s, m) dm.$$

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

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

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

For the policy  $\pi'$ , a jump at  $m$  is

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

and from (3.6)

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

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

$$(4.3) \quad U_k(s, L) = r_k + \int_M^C m d\xi_k^{*}(s, m) \\ = r_k + \Psi_k C - \xi^{*}(s, M)M - \int_{C_k}^C \alpha_k^{*}(m) \xi^{*}(s, m) dm - \int_M^{C_k} \xi^{*}(s, m) dm.$$

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

$$(4.4) \quad U_k(s, L) = r_k + \Psi_k C - \alpha_k^{*}(M) \xi^{*}(s, M)M - \int_M^C \alpha_k^{*}(m) \xi^{*}(s, m) dm.$$

LEMMA 5. If  $n_k > 0$  and  $k \in L$  then

$$(4.5) \quad V_{\pi}(s) - V_k(s) = -r_k + (1 - \Psi_k)C - \int_{C_k}^C (1 - \alpha_k^{*}(m)) \xi^{*}(s, m) dm$$

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

$$(4.6) \quad V_{\pi}(s) - V_k(s) = -r_k + (1 - \Psi_k)C - (1 - \alpha_k^*(M))\xi^*(s, M)(M - V_{\pi}(0)) \\ - \int_M^C (1 - \alpha_k^*(m))\xi^*(s, m)dm.$$

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  $k \in L$ ,  $\xi(s, L) = \xi_k(s, L)$  then it follow from (4.2) and (4.3) that (4.5) is obtained. Also if  $k \in K - L$  or  $k = 0$ , then from (4.2) and (4.4) (4.6) is obtained.

Generalizing the index  $C_k$  in (3.8), for  $k \in K - L$  or  $k = 0$ , we define  $C_k(L)$  as

$$(4.7) \quad C_k(L) \equiv U_k(\delta_k, L) / (1 - \alpha_k(L)) \\ = \{r_k + \Psi_k C - \alpha_k^*(M)M - \int_M^C \alpha_k^*(m)dm\} / (1 - \alpha_k^*(M)).$$

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

$$(4.8) \quad C_0(L) = \sum_{k \in L} \lambda_k U(\delta_k, L) / (\beta + \Lambda_L - \Lambda_L \alpha(L)) \\ = V_{\pi}(0).$$

THEOREM 6. Suppose that  $\pi$  is a priority policy with the ordering set  $\Gamma$  and the cut-off set  $L$  such that for  $k \in L$ ,  $C_k$  is monotone nonincreasing with respect to  $\Gamma$  and

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

Then  $\pi$  is optimal.

PROOF. We use the policy improvement technique. To prove this it is sufficient to show that for any fixed  $s = \delta_0$  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

$$V_{\pi}(s) - V_k(s) \geq V_{\pi}(\delta_k) - V_k(\delta_k) \\ = -r_k + (1 - \Psi_k)C - (1 - \alpha_k^*(M))(M - V_{\pi}(0)) - \int_M^C (1 - \alpha_k^*(m))dm$$

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

where the third equality is derived from (4.7) and (4.8). This completes the proof.

In order to construct the optimal ordering set  $\Gamma$  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

$$(4.9) \quad C_j(L) \leq C_j(L \cup i) \leq C_i(L)$$

and

$$(4.10) \quad C_j(L) \leq C_i(L \cup j) \leq C_i(L)$$

PROOF For  $i \neq 0$  and  $j \neq 0$  we get

$$\begin{aligned} (1-a_j(L \cup i))C_j(L \cup i) &= U_j(\delta_j, L) + (a_j(L) - a_j(L \cup i))C_i(L) \\ &= (1-a_j(L))C_j(L) + (a_j(L) - a_j(L \cup i))C_i(L) \\ &= (1-a_j(L \cup i))C_i(L) - (1-a_j(L))(C_i(L) - C_j(L)) \end{aligned}$$

or

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

Since  $0 \leq (1-a_j(L))/(1-a_j(L \cup i)) \leq 1$  we get (4.9). Exchanging the role of  $i$  and  $j$  we also have (4.10). If one of  $i$  and  $j$  is 0, we can prove this using the same discussion.

The optimal ordering  $\Gamma$  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.

### 5. Approximation of preemptive $M/GI/1$ queue

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  $\lambda$ . Each arriving customer has a discrete service time  $k=1, 2, \dots$  with the probability  $g_k \left( \sum_{k=1}^{\infty} g_k = 1 \right)$ . Let a class  $k$  job be a customer who has served  $k-1$  quanta and is waiting for  $k$ th quantum. The conditional probability of a class  $k$  job exiting 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  $1-p_k$ . This model is a discrete approximation of a preemptive  $M/GI/1$  queue. 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 (2,7) 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  $i$ th 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

$$(5.1) \quad C_i' \equiv \sup_{i \leq j} C_{i,j}.$$

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  $Y_k = \{1, \dots, k-1\}$  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

$$\begin{aligned} U_i(\delta_i, Y_k) &= p_i + \Psi \sum_{a=0}^{\infty} \frac{\dot{e}^{-\lambda} \lambda^a}{a!} U(\delta_1, Y_k) [1 + \dots + \alpha(1, Y_k)^{a-1}] \\ &= p_i + \frac{U(\delta_1, Y_k)}{1 - \alpha(1, Y_k)} [\Psi - \alpha_i(Y_k)], \end{aligned}$$

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  $i \in \{k, k+1, \dots\}$ . From Prop. 7,  $C_k$  is recursively determined as  $C_k = C_k(Y_k)$ . Therefore  $C_k$  is nonincreasing and from Theorem 6,  $\Gamma_1$  is the optimal ordering set. This ordered 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  $k \in \{1, \dots, i_0\}$  and nonincreasing in  $k \in \{i_0, i_0+1, \dots\}$ . From Prop. 8, for any fixed  $i$ ,  $C_{i,j}$  is an unimodal function of  $j$  and then  $C_k'$  is also unimodal such that  $C_k'$  is increasing in  $k \in \{1, \dots, i_0\}$  and decreasing in  $k \in \{i_0, i_0+1, \dots\}$ . In no arrival case  $\Gamma_3 = \{(i, j) : C_i' > C_j' \text{ or } (C_i' = C_j' \text{ and } i < j)\}$  is an optimal ordering set. Next we consider an arrival case. From (3,8) 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 (3,1) and (3,8) we have  $C_1 = U(\delta_1, Z_1) /$

$(1 - \alpha(1, Z_1))$  where  $Z_1 = \{1, \dots, j(1)\}$ . For  $k \geq j(1) + 1$  we put  $Y_k = \{1, \dots, k-1\}$  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  $k = 1, \dots, j(1)$   $C_k = C_k'$  is the unimodal index without arrival and for  $k = j(1) + 1, \dots$   $C_k = C_k(Y_k)$  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  $\Gamma_3$ . For any initial state  $s$  there exists only one job from class 2 to  $j(1)$  after sufficiently long time spent when  $\Gamma_3$  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  $M/G/1$  queuing system with removable 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 allocation 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, Dynamic Scheduling of a finite-source  $M/M/1$  queue with two customer classes (submitted).
- [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. Rosberg, 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 multiarmed 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.