

ANALYSIS OF A TWO-CLASS PRIORITY QUEUE WITH BERNOULLI SCHEDULES

Tsuyoshi Katayama Yoshitaka Takahashi
Toyama Prefectural University NTT Laboratories

(Received May 13, 1991)

Abstract A Bernoulli schedule is random service discipline for a multi-class priority queueing system which operates as follows: If queue i ($1 \leq i \leq N$) is not empty just after servicing a message in its queue, a message in queue i is served again with probability p_i , and the highest class message present in the system is served with probability $1 - p_i$, where $0 \leq p_i \leq 1$. This paper presents an analysis of a two-class priority queue ($M/G/1$ type queue) with Bernoulli schedules of parameter ($p_1 = 1, 0 \leq p_2 \leq 1$) and class-dependent set-up times. The generating functions of joint queue-length distributions and the Laplace-Stieltjes transforms of waiting time distributions are determined. A closed-form expression with infinite summations is obtained for the mean waiting times.

1. Introduction

There have been some analytical studies of priority queueing systems with controllable parameters, e. g. Refs [3, 7, 9, 10, 19]. These priority schedules have some advantages over the ordinary priority schedules without parameters [11], since these enable us to control efficiently service qualities in telecommunication systems, e. g. delay time, loss probability etc. Recently, the Broadband Integrated Services Digital Network (B-ISDN) has been extensively studied according to the increasing demands of telecommunication services. Owing to multiple grades of service requirements, it needs essentially flexible call processings to achieve performance objectives in the network nodes such as switching systems, communication processor units, terminal instruments etc. From such a point of view, these flexible priority schedules with controllable parameters are effective for the performance optimization and applicable to call processings and routing schemes in the telecommunication systems.

We thus introduce a priority queue with Bernoulli schedules [14]. This schedule is described as follows: Let there be N classes of messages (or calls), where messages with a smaller class number have a priority (nonpreemptive) over messages with a greater class number. This schedule discipline is parameterized by a vector (p_1, p_2, \dots, p_N) , where $0 \leq p_i \leq 1, i = 1, 2, \dots, N$. For the moment, suppose that a message of class- i is served, $i = 1, 2, \dots, N$. At service completion of the message, if there are messages of class- i and messages of higher class- j ($j < i$), a class- i message will be served next with probability p_i and a message of highest class present in the system will be served with probability $1 - p_i$. If there is no message of higher class- j ($j < i$), a message of the highest class present in the system will be served with probability one. It is then plausible to be defined as $p_1 = 1$. The class- i service with $p_i = 1$ is equivalent to the exhaustive service discipline. If $p_i = 0$ for $i = 2, 3, \dots, N$, then this schedule reduces to the ordinary nonpreemptive (head-of-the-line) priority discipline, whereas if $p_i = 1$ for $i = 2, 3, \dots, N$, it reduces to the exhaustive service priority discipline described in [26, 29]. In the case of $N = 2$ (i. e. $p_1 = p_2 = 1$), it also reduces to the alternating priority discipline [11, 23].

The multi-class priority queue with Bernoulli schedules (p_1, p_2, \dots, p_N) has a potential applicability to call processing with multi-class tasks (service requests) in switching systems since the processor unit under the ordinary nonpreemptive (or preemptive) priority discipline is to devote more of its power to some tasks of higher priorities. In such situations, the Bernoulli schedule can be used by simply choosing an appropriate parameter $p_i, i = 2, 3, \dots, N$ for each class task.

In this paper, we consider a two-class priority queue with Bernoulli schedules $(p_1 = 1, p_2 = p)$. The analysis of a multi-class priority queue with Bernoulli schedules is closely related to the one of a cyclic-service polling model with Bernoulli schedules of (p_1, p_2, \dots, p_N) . There are fruitful results for polling models. See [1, 4, 5, 18, 24, 28]. However, there are not so many results for priority models.

The rest of this paper is organized as follows: Section 2 describes a two-class priority queueing model with Bernoulli schedules and class-dependent set-up times. Section 3 is devoted to the derivation of queue-length generating functions and the Laplace-Stieltjes transform (LST) of waiting time distributions. Section 4 derives the mean waiting time expressions with infinite summations by using iterative schemes. Section 5 gives some comments on the multi-class priority queue with Bernoulli schedules. Section 6 gives concluding remarks. Throughout this paper, we shall use queueing terms instead of technical terms of call processing in telecommunication systems.

2. Queueing Model with Set-up Times

This section presents a two-class priority queueing system with Bernoulli schedules having set-up times in detail and introduces some notations. The queueing system consists of two-parallel queues, Q_1 and Q_2 , for customers of class-1 and class-2, respectively. Each queue Q_i with an infinite capacity waiting room has a service counter for class- i customers $S_i, i = 1, 2$. Customers of class- i arrive in a Poisson process at rate $\lambda_i, i = 1, 2$. Service time τ_i for class- i customers has a general distribution $H_i(t)$ with finite first and second moments h_i and $h_i^{(2)}$. Customers in Q_1 and Q_2 are served by a single-server in accordance with the Bernoulli schedules $(p_1 = 1, p_2 = p)$. That is, at the service completion of a class-2 customer, if there are customers of class-1 and class-2, a class-2 customer is served next with probability p and a class-1 customer is served with probability $\bar{p} := 1 - p$. If there is no customer of class-1, a class-2 customer will be served. Starting once a service of class-1 customers, the server serves all customers in Q_1 until it becomes empty. The server requires set-up time v_i having a general distribution $S_i(t)$ with finite first and second moments s_i and $s_i^{(2)}$, before starting a busy period with an initial customer of class- $i, i = 1, 2$. Even if class-1 customers arrive during the set-up time v_2 , the server starts first servicing of the class-2 customer as soon as its set-up time terminates. (This slightly differs from the priority discipline in [25] under which the first service in a busy period is given to a class-1 customer in the above case. The set-up time here is also different from the orientation time investigated in [11] since the latter is not required before starting the first service in a busy period if a message of an oriented class arrives during an idle period [25].) Customers within a class are served in each queue on the first-in-first-out (FIFO) discipline. The LSTs of $H_i(t)$ and $S_i(t)$ are denoted by $H_i^*(s)$, and $S_i^*(s), i = 1, 2$. The switch-over time needed to switch service from one class to another is assumed to be zero.

For simplicity, the following notation is introduced:

$$\begin{aligned}
 \lambda &:= \lambda_1 + \lambda_2 & r_i &:= \lambda_i / \lambda & i &= 1, 2 \\
 h &:= r_1 h_1 + r_2 h_2 & h^{(2)} &:= r_1 h_1^{(2)} + r_2 h_2^{(2)} \\
 s &:= r_1 s_1 + r_2 s_2 & s^{(2)} &:= r_1 s_1^{(2)} + r_2 s_2^{(2)}
 \end{aligned} \tag{1.a}$$

and

$$\rho_i := \lambda_i h_i \quad i = 1, 2 \quad \rho := \rho_1 + \rho_2. \tag{1.b}$$

The total traffic intensity is assumed to be less than unity ($\rho < 1$) to ensure stability. We denote by $q_i(m, n,)$ or $r_i(m, n)$ the probability that m customers arrive at Q_1 and n customers arrive at Q_2 during a service time $\tau_i, i = 1, 2$, or during a set-up time $v_i, i = 1, 2$, respectively. Denoting by $Q_i(x, y) := \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} q_i(m, n) x^m y^n$ and $R_i(x, y)$ the corresponding generating functions for $q_i(m, n)$ and $r_i(m, n)$, we have

$$Q_i(x, y) = H_i^* \{ \lambda_1(1-x) + \lambda_2(1-y) \} \tag{2.a}$$

$$R_i(x, y) = S_i^* \{ \lambda_1(1-x) + \lambda_2(1-y) \} \quad i = 1, 2. \tag{2.b}$$

Remark 2.1 Welch [31] gives a necessary and sufficient condition, $\rho < 1$, for stability in the ordinary $M/G/1$ queue with an exceptional service time. With respect to the condition of system stability, the present model is equivalent to his model with arrival rate λ and the *LST*s of service time and exceptional service time, $r_1 H_1^*(s) + r_2 H_2^*(s)$ and $r_1 S_1^*(s) + r_2 S_2^*(s)$. □

3. Queueing Analysis

In this section, we will derive queue-length generating functions and *LST*s of waiting time distributions. We firstly determine a set of generating functions for the joint queue-length distribution at departure epochs of customers from each service counter. We introduce:

$\pi_k(i, j)$: the steady-state probability that i customers are waiting in Q_1 and j customers are waiting in Q_2 just after a class- k customer has completed service at the service counter $S_k, k = 1, 2, \quad i, j = 0, 1, 2, \dots$

and

$$\Pi_k(x, y) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \pi_k(i, j) x^i y^j \quad |x|, |y| \leq 1 \quad k = 1, 2. \tag{3}$$

Then, considering the events that occur during two successive service completion points at the service counters S_1 and S_2 , we have the following balance equations for the joint queue-length distribution $\{\pi_k(i, j)\}$:

$$\begin{aligned} \pi_1(i, j) = & \sum_{m=0}^i \sum_{n=0}^j \pi_1(i-m+1, j-n) q_1(m, n) + \sum_{m=0}^i \sum_{n=0}^{j-1} \pi_2(i-m+1, j-n) \bar{p} q_1(m, n) \\ & + \sum_{m=0}^i \pi_2(i-m+1, 0) q_1(m, j) + \pi_0 r_1 \sum_{m=0}^i \sum_{n=0}^j r_1(i-m, j-n) q_1(m, n) \quad i, j \geq 0 \end{aligned} \tag{4.a}$$

$$\begin{aligned} \pi_2(i, j) = & \sum_{m=0}^{i-1} \sum_{n=0}^j \pi_2(i-m, j-n+1) p q_2(m, n) + \sum_{n=0}^j \pi_2(0, j-n+1) q_2(i, n) \\ & + \sum_{n=0}^j \pi_1(0, j-n+1) q_2(i, n) + \pi_0 r_2 \sum_{m=0}^i \sum_{n=0}^j r_2(i-m, j-n) q_2(m, n) \quad i, j \geq 0 \end{aligned}$$

where

$$\pi_0 := \pi_1(0, 0) + \pi_2(0, 0). \tag{4.b}$$

Then, from these equations, we get the following functional relationship:

$$\Pi_1(x, y) = [\{\Pi_1(x, y) - \Pi_1(0, y)\} + \{F_2(x, y) - F_2(0, y)\} + \pi_0 r_1 x R_1(x, y)] Q_1(x, y) \cdot \frac{1}{x} \tag{5.a}$$

$$\begin{aligned} \Pi_2(x, y) = & [\{(\Pi_2(x, y) - \Pi_2(x, 0)) - (\Pi_2(0, y) - \Pi_2(0, 0))\} p \\ & + \{\Pi_2(0, y) - \Pi_2(0, 0)\} + \{\Pi_1(0, y) - \Pi_1(0, 0)\} + \pi_0 r_2 y R_2(x, y)] Q_2(x, y) \cdot \frac{1}{y} \end{aligned} \tag{5.b}$$

where

$$F_2(x, y) := \{\Pi_2(x, y) - \Pi_2(x, 0)\} \bar{p} + \Pi_2(x, 0). \tag{5.c}$$

(These functional relationships can be also derived by the method of collective marks).

Hence, by arrangement of (5.a), (5.b) and (5.c), we get

$$\begin{aligned} \Pi_1(x, y) = & \frac{Q_1(x, y)}{x - Q_1(x, y)} \left[\frac{y - Q_2(x, y)}{y - pQ_2(x, y)} \{\psi(x) - \varphi(y)\} \right. \\ & \left. + \{r_1 x R_1(x, y) - \frac{\bar{p}\{1 - r_2 y R_2(x, y)\} Q_2(x, y)}{y - pQ_2(x, y)}\} \pi_0 \right] \end{aligned} \tag{6.a}$$

$$\Pi_2(x, y) = \frac{Q_2(x, y)}{y - pQ_2(x, y)} [\varphi(y) - \psi(x) - \{1 - r_2 y R_2(x, y)\} \pi_0] \tag{6.b}$$

where

$$\begin{aligned} \psi(x) & := p\Pi_2(x, 0) \\ \varphi(y) & := \bar{p}\Pi_2(0, y) + \Pi_1(0, y) + p\Pi_2(0, 0). \end{aligned} \tag{6.c}$$

These equations are the starting point for our analysis. It is necessary to determine the unknown probability π_0 and unknown functions $\psi(x)$ and $\varphi(y)$ on the right-hand side of (6.a) and (6.b).

(1) Determination of π_0

The unknown probability π_0 should be determined by the normalization condition,

$$\Pi_1(1, 1) + \Pi_2(1, 1) = 1. \tag{7}$$

From (7) and the relations derived by letting $x, y \rightarrow 1$ in (6.a) and (6.b), we have

$$\pi_0 = \frac{1 - \rho}{1 + \lambda s} \tag{8.a}$$

and

$$\Pi_i(1, 1) = r_i \quad i = 1, 2. \tag{8.b}$$

A probabilistic interpretation for (8.a) is given by using the decomposition property for the queueing model with vacation: The probability p_0 that there is no customer in an N -class service queueing system with set-up times at an arbitrary time is given by

$$\begin{aligned} p_0 & = (1 - \rho) \cdot \frac{1/\lambda}{1/\lambda + \sum_{i=1}^N r_i s_i} \\ & = \frac{1 - \rho}{1 + \lambda s} \end{aligned} \tag{8.c}$$

where the first factor on the right-hand side of (8.c), $1 - \rho$, stands for the probability that the server is idle in the queueing system without set-up times, and the second factor is the probability that a randomly chosen sampling falls in the idle period at a non-serving interval, i. e. a server idle period plus a set-up time, in the queueing system with set-up times [13]. Further, it is derived from the property of the $M/G/1$ type queue [26] that $p_0 = \pi_0$. Hence, we get (8.a).

(2) Determination of $\psi(x)$ and $\varphi(y)$

On the right-hand side of (6.a),

$$x - Q_1(x, y) = x - H_1^* \{ \lambda_1(1 - x) + \lambda_2(1 - y) \} = 0 \tag{9.a}$$

has exactly one root, say $x = \delta(y)$, in the unit circle $|x| \leq 1$ under the condition: $\rho_1 \leq 1$ and $|y| \leq 1$. Similarly, in (6.a) and (6.b),

$$y - pQ_2(x, y) = y - pH_2^* \{ \lambda_1(1 - x) + \lambda_2(1 - y) \} = 0 \tag{9.b}$$

has one root, $y = \sigma(x)$, in the unit circle $|y| \leq 1$ under the condition: $\rho_2 \leq 1$ and $|x|, p \leq 1$. Explicit expressions of $\delta(y)$ and $\sigma(x)$ are given by (14.b) known as Takács' lemma [22]. From the regularity of $\Pi_1(x, y)$ and $\Pi_2(x, y)$, the numerators on the right-hand side of (6.a) and (6.b) should be equal to zero for $x = \delta(y)$ and $y = \sigma(x)$, respectively. Thus, we obtain a simultaneous functional equation for two unknowns, $\psi(x)$ and $\varphi(y)$, to be solved for determination of $\Pi_k(x, y)$, $k = 1, 2$:

$$\psi(\delta(y)) - \varphi(y) = \frac{\bar{p} \{ 1 - r_2 y R_2(\delta(y), y) \} Q_2(\delta(y), y) - r_1 \delta(y) R_1(\delta(y), y) \{ y - pQ_2(\delta(y), y) \}}{y - Q_2(\delta(y), y)} \pi_0 \tag{10.a}$$

$$\varphi(\sigma(x)) - \psi(x) = \{ 1 - r_2 \sigma(x) R_2(x, \sigma(x)) \} \pi_0. \tag{10.b}$$

Here, eliminating $\psi(\delta(y))$ from (10.a) and (10.b) after setting $x = \delta(y)$ in (10.b), we obtain a non-homogeneous linear functional equation for $\varphi(y)$,

$$\varphi[f(y)] - \varphi(y) = \pi_0 g(y) \tag{11.a}$$

where

$$f(y) := \sigma \{ \delta(y) \} \tag{11.b}$$

$$g(y) := 1 - r_2 \sigma \{ \delta(y) \} R_2[\delta(y), \sigma \{ \delta(y) \}] + \frac{\bar{p} \{ 1 - r_2 y R_2(\delta(y), y) \} Q_2(\delta(y), y) - r_1 \delta(y) R_1(\delta(y), y) \{ y - pQ_2(\delta(y), y) \}}{y - Q_2(\delta(y), y)}.$$

Using an iterative scheme, e. g. [12, 23], $\varphi(y)$ is determined as follows: First, let us introduce a sequence of $\{y_i\}$ and as function defined by

$$\begin{aligned} y_0 &= y \quad 0 \leq y \leq 1 \\ y_i &= f(y_{i-1}) \quad i = 1, 2, 3, \dots \end{aligned} \tag{12.a}$$

and

$$g\{y_i | y_0 = y\} := g(y_i) \quad i = 0, 1, 2, \dots \tag{12.b}$$

Then, it follows from (11.a) that

$$\varphi(y_{i+1}) - \varphi(y_i) = \pi_0 g(y_i) \quad i = 0, 1, 2, \dots \tag{13.a}$$

Using this relation repeatedly, we have

$$\varphi(y) = \eta - \pi_0 \sum_{i=0}^{\infty} g\{y_i|y_0 = y\} \tag{13.b}$$

where η is a constant which is independent of the sequence of $\{y_i\}$. Fig. 1 shows the iterative procedure for determining a sequence of $\{y_i\}$. It is shown that $\varphi(y_i)$ for $i \rightarrow \infty$ approaches to $\varphi(\xi) = \eta$ under the condition $\rho < 1$ (i. e. $\rho_1, \rho_2 \leq 1$ and $\pi_0 \neq 0$), where ξ is a root of the equation $y = \sigma\{\delta(y)\}$ for $0 \leq y \leq 1$. The convergence of the infinite sum in (13.b) is also shown (see Appendix).

From a boundary condition, $\varphi(0) = \pi_0$, which is derived by putting $y = 0$ in (6.c), the constant η can be determined as

$$\eta = \pi_0 [1 + \sum_{i=0}^{\infty} g\{y_i|y_0 = 0\}]. \tag{13.c}$$

Thus, the unknown function $\varphi(y)$ has been determined, and $\psi(x)$ is obtained from (10.b). In this way, the generating functions $\Pi_k(x, y)$, $k = 1, 2$ have been completely determined. We thus obtain the following results.

Theorem 1. The generating functions $\Pi_k(x, y)$, $k = 1, 2$ for the joint queue-length distribution $\{\pi_k(i, j)\}$ are given by (6.a) and (6.b), where $\varphi(y)$ and $\psi(x)$ are expressed by:

$$\varphi(y) = \pi_0 [1 + G(0) - G(y)] \tag{14.a}$$

$$\psi(x) = \pi_0 [r_2 \sigma(x) R_2\{x, \sigma(x)\} + G(0) - G\{\sigma(x)\}]. \tag{14.b}$$

Here

$$G(y) := \sum_{i=0}^{\infty} g\{y_i|y_0 = y\} \tag{14.c}$$

and

$$\begin{aligned} \delta(y) &= \sum_{j=1}^{\infty} \frac{\lambda_1^{j-1}}{j!} \int_0^{\infty} \exp\{-(\lambda - \lambda_2 y)t\} t^{j-1} dH_1^{(j)}(t) \\ \sigma(x) &= \sum_{j=1}^{\infty} \frac{p^j}{j!} \int_0^{\infty} \exp\{-(\lambda - \lambda_1 x)t\} (\lambda_2 t)^{j-1} dH_2^{(j)}(t) \end{aligned} \tag{14.d}$$

where $H_k^{(j)}(t)$ denotes the j -th iterated convolution of $H_k(t)$ with itself, $k = 1, 2$. □

In the case of exponential service time distributions with $LSTs H_i^*(s) = \mu_i / (s + \mu_i)$, $i = 1, 2$, the explicit forms without summation symbols of $\delta(y)$ and $\sigma(x)$ of (14.d) are given by:

$$\delta(y) = \frac{1}{2\lambda_1} [\lambda_1 + \mu_1 + \lambda_2(1 - y) - \sqrt{\{\lambda_1 + \mu_1 + \lambda_2(1 - y)\}^2 - 4\lambda_1\mu_1}] \tag{15.a}$$

$$\sigma(x) = \frac{1}{2\lambda_2} [\lambda_2 + \mu_2 + \lambda_1(1 - x) - \sqrt{\{\lambda_2 + \mu_2 + \lambda_1(1 - x)\}^2 - 4\lambda_2\mu_2}]. \tag{15.b}$$

We denote by $W_i^*(s)$ the LST of distribution function for the waiting time of a class- i customer $W_i(t)$, $i = 1, 2$. From the usual argument that the number of class- i customers at

the departure time of a class- i customer is equal to the number of class- i customers that arrive during its sojourn time because of the *FIFO* service discipline, we have

$$\frac{\Pi_1(x, 1)}{\Pi_1(1, 1)} = W_1^* \{ \lambda_1(1 - x) \} H_1^* \{ \lambda_1(1 - x) \} \tag{16.a}$$

and

$$\frac{\Pi_2(1, y)}{\Pi_2(1, 1)} = W_2^* \{ \lambda_2(1 - y) \} H_2^* \{ \lambda_2(1 - y) \}. \tag{16.b}$$

Using *Theorem 1*, Eqs (2), (8.b) and these relationships, we obtain the following result.

Theorem 2. The *LSTs* of waiting time distribution for class- i customer, $i = 1, 2$ are given by:

$$W_1^*(s) = \frac{\lambda}{s - \lambda_1 \{ 1 - H_1^*(s) \}} \left[\frac{1 - H_2^*(s)}{1 - p H_2^*(s)} \{ \varphi(1) - \psi(1 - s/\lambda_1) \} - \{ r_1(1 - s/\lambda_1) S_1^*(s) - \frac{\bar{p} \{ 1 - r_2 S_2^*(s) \} H_2^*(s)}{1 - p H_2^*(s)} \} \pi_0 \right] \tag{17.a}$$

$$W_2^*(s) = \frac{\lambda}{s - \lambda_2 \{ 1 - p H_2^*(s) \}} \left[\psi(1) - \varphi(1 - s/\lambda_2) + \{ 1 - r_2(1 - s/\lambda_2) S_2^*(s) \} \pi_0 \right]. \tag{17.b}$$

□

We denote by $W_i^c(0), i = 1, 2$ the probability (often called delay probability) that a class- i customer must wait on its arrival at $Q_i, i = 1, 2$. $W_i^c(0)$ can be found by applying the Abelian theorem to (17.a) and (17.b) after dividing by s both the denominator and the numerator on the right-hand side of (17.a) and (17.b), i. e.,

$$W_i^c(0) := \lim_{t \rightarrow 0} (1 - W_i(t)) = 1 - \lim_{s \rightarrow \infty} W_i^*(s) \quad i = 1, 2. \tag{18}$$

Hence, we get the following result.

Corollary 1. Delay probabilities of class- i customers, $W_i^c(0), i = 1, 2$ are given by:

$$\begin{aligned} W_1^c(0) &= W_2^c(0) = \rho & \text{for } s_1 = s_2 = 0 \\ W_1^c(0) &= W_2^c(0) = 1 & \text{for } s_1, s_2 > 0 \end{aligned}$$

and

$$\begin{aligned} W_1^c(0) &= \frac{\rho + \lambda_2 s_2}{1 + \lambda_2 s_2} & W_2^c(0) &= 1 & \text{for } s_1 = 0, s_2 > 0 \\ W_1^c(0) &= 1 & W_2^c(0) &= \frac{\rho + \lambda_1 s_1}{1 + \lambda_1 s_1} & \text{for } s_1 > 0, s_2 = 0. \end{aligned} \tag{19}$$

□

It is seen that delay probabilities $W_i^c(0), i = 1, 2$ are independent of the parameter p . This is closely related to that π_0 in (8.a) is also independent of p .

4. Mean Waiting Times

In this section, we derive explicit forms for mean waiting times. The mean waiting times of class- i customers can be obtained by putting $s = 0$ after differentiating $-W_i^*(s), i = 1, 2$

in *Theorem 2*. However, we need to apply L'Hospital's rule to (17.a) and (17.b) in the case of $p = 1$. Indeed, the denominators on the right-hand side of (17.a) and (17.b), i. e. $s - \lambda_i \{1 - H_i^*(s)\}$, $i = 1, 2$ tend to zero as $s \rightarrow 0$. Denoting by $E(W_i)$, $i = 1, 2$, the mean waiting times for class- i customers, we get the explicit expressions for $E(W_i)$, $i = 1, 2$, where the following abbreviations are used for a differentiable generating function $R_i(x, y)$,

$$R'_{ix}(a, b) := \left[\frac{\partial}{\partial x} R_i(x, y)\right]_{x=a, y=b} \quad R''_{ixy}(a, b) := \left[\frac{\partial^2}{\partial x \partial y} R_i(x, y)\right]_{x=a, y=b}. \tag{20}$$

Theorem 3. The mean waiting times of class- i customers, $E(W_i)$, $i = 1, 2$ are given by:

$$E(W_1) = \frac{1}{1 - \rho_1} [(\rho_1 - 1 + \pi_0)h_1 + \pi_0\{(1 + \rho_1)s_1 + \rho_2s_2\} + (\bar{p}h_1 + ph_2)(\rho_2 + \pi_0\lambda_2s_2)/(1 - p)] \\ + \frac{1}{2(1 - \rho_1)} [\lambda_1h_1^{(2)} + \lambda_2h_2^{(2)} + \pi_0(\lambda_1s_1^{(2)} + \lambda_2s_2^{(2)}) - 2h_2/(1 - p) \cdot \psi'(1)/r_1] \\ \text{for } 1 > p \geq 0 \tag{21.a}$$

$$= \frac{1}{2(1 - \rho_1)^2} \{[(1 + \lambda_1s_1)\pi_0 + \psi'(1)/r_1]\lambda_1h_1^{(2)} \\ + (1 - \rho_1)\{(2s_1 + \lambda_1s_1^{(2)})\pi_0 + \psi''(1)/(\lambda_1r_1)\}\} \text{ for } p = 1 \tag{21.b}$$

$$E(W_2) = \frac{1}{\lambda_2(1 - p)} [p\rho_2 - 1 + \pi_0(1 + \lambda_2s_2) + \varphi'(1)/r_2] \text{ for } 1 > p \geq 0 \tag{22.a}$$

$$= \frac{1}{2(1 - \rho_2)^2} \{[(1 + \lambda_2s_2)\pi_0 + \varphi'(1)/r_2]\lambda_2h_2^{(2)} \\ + (1 - \rho_2)\{(2s_2 + \lambda_2s_2^{(2)})\pi_0 + \varphi''(1)/(\lambda_2r_2)\}\} \text{ for } p = 1. \tag{22.b}$$

The differential terms on the right-hand side of (21.b) are given by:

$$\psi'(1) = \pi_0[r_2\sigma(1)\{R'_{2x}(1, \sigma(1)) + \sigma'(1)R'_{2y}(1, \sigma(1))\} + r_2\sigma'(1)R_2(1, \sigma(1)) - \sigma'(1)G'(\sigma(1))] \\ \psi''(1) = \pi_0[r_2\{\sigma''(1)R_2(1, \sigma(1)) + 2\sigma'(1)R'_{2x}(1, \sigma(1)) + 2(\sigma'(1))^2R'_{2y}(1, \sigma(1)) \\ + \sigma(1)\{R''_{2xx}(1, \sigma(1)) + 2\sigma'(1)R''_{2xy}(1, \sigma(1)) + (\sigma'(1))^2R''_{2yy}(1, \sigma(1)) \\ + \sigma''(1)R'_{2y}(1, \sigma(1))\}\} - \sigma''(1)G'(\sigma(1)) - (\sigma'(1))^2G''(\sigma(1))] \\ \varphi'(1) = -\pi_0G'(1) \\ \varphi''(1) = -\pi_0G''(1) \tag{23.a}$$

where

$$G'(y) = \sum_{i=0}^{\infty} g'(y_i|y_0 = y) \prod_{j=0}^{i-1} f'(y_j) \\ G''(y) = \sum_{i=0}^{\infty} g''(y_i|y_0 = y) \left(\prod_{j=0}^{i-1} f'(y_j)\right)^2 \\ + \sum_{i=0}^{\infty} g'(y_i|y_0 = y) \prod_{j=0}^{i-1} f'(y_j) \cdot \left\{\sum_{j=0}^{i-1} \frac{f''(y_j)}{f'(y_j)} \prod_{k=0}^{j-1} f'(y_k)\right\}. \tag{23.b}$$

(We assume that null products are equal to one and empty sums are equal to zero). □

The convergences of the infinite sums in (23.b) are also shown by the same argument as Appendix. The n -th (≥ 2) moments of waiting time distributions can be also derived from *Theorem 2* though the expressions are considerably complicated since those contain an iterative form function with infinite summation $G(y)$ defined by (14.c).

Next, we will consider two special cases of nonpreemptive priority service ($p = 0$) and alternating priority service ($p = 1$). Putting $p = 0$ in (6.c), we have $\psi(x) = 0$ for all x in $|x| \leq 1$. Hence, the explicit forms without infinite summation of $\varphi'(1)$ for $p = 0$ can be obtained by differentiating (9.a) and applying L'Hospital's rule to the right-hand side of (10.a). We thus get the explicit expressions without infinite summation for the mean waiting times. For $0 < p < 1$, however, mean waiting times are expressed by the form with infinite summation using iterative sequences as the busy period [20].

Corollary 2. The mean waiting times of a class- i customer, $E(W_i), i = 1, 2$ are given by: for $p = 0$ (the nonpreemptive priority service)

$$E(W_1) = \frac{1}{2(1 - \rho_1)} \{ \lambda_1 h_1^{(2)} + \lambda_2 h_2^{(2)} \} + \frac{\pi_0}{2(1 - \rho_1)} [2(s_1 + \rho_2 s_2) + \lambda_1 s_1^{(2)} + \lambda_2 s_2^{(2)}] \quad (24.a)$$

$$E(W_2) = \frac{\lambda_1 h_1^{(2)} + \lambda_2 h_2^{(2)}}{2(1 - \rho_1)(1 - \rho)} + \frac{\pi_0 [2(\rho_1 s_1 + \rho_2 s_2) + 2(1 - \rho_1)(1 - \rho)s_2 + \lambda_1 s_1^{(2)} + \lambda_2 s_2^{(2)}]}{2(1 - \rho_1)(1 - \rho)} \quad (24.b)$$

and for $p = 1$ (the alternating priority service)

$$E(W_1) = \frac{\lambda_1}{2(1 - \rho_1)} \{ h_1^{(2)} + \pi_0 s_1^{(2)} \} + \frac{\pi_0}{1 - \rho_1} [s_1 + \frac{\rho_1 \rho_2^2 s_1 + (1 - \rho_1)^2 \rho_2 s_2}{(1 - \rho)(1 - \rho + 2\rho_1 \rho_2)} + \frac{\rho_2^2 \lambda_1 h_1^{(2)} + (1 - \rho_1)^2 \lambda_2 h_2^{(2)} + \pi_0 \{ \rho_2^2 \lambda_1 s_1^{(2)} + (1 - \rho_1)^2 \lambda_2 s_2^{(2)} \}}{2(1 - \rho_1)(1 - \rho)(1 - \rho + 2\rho_1 \rho_2)}] \quad (25.a)$$

$$E(W_2) = \frac{\lambda_2}{2(1 - \rho_2)} \{ h_2^{(2)} + \pi_0 s_2^{(2)} \} + \frac{\pi_0}{1 - \rho_2} [s_2 + \frac{\rho_1 (1 - \rho_2)^2 s_1 + \rho_1^2 \rho_2 s_2}{(1 - \rho)(1 - \rho + 2\rho_1 \rho_2)} + \frac{(1 - \rho_2)^2 \lambda_1 h_1^{(2)} + \rho_1^2 \lambda_2 h_2^{(2)} + \pi_0 \{ (1 - \rho_2)^2 \lambda_1 s_1^{(2)} + \rho_1^2 \lambda_2 s_2^{(2)} \}}{2(1 - \rho_2)(1 - \rho)(1 - \rho + 2\rho_1 \rho_2)}]. \quad (25.b)$$

□

It is seen from *Theorem 3* and *Corollary 2* that the mean waiting times satisfy the following conservation law for an N -class service queueing system with set-up times [13, 26],

$$\sum_{i=1}^N \rho_i E(W_i) / \rho = \frac{\lambda h^{(2)}}{2(1 - \rho)} + \frac{\pi_0}{2\rho(1 - \rho)} \{ 2 \sum_{i=1}^N \rho_i s_i + \rho \lambda s^{(2)} \}. \quad (26)$$

Note that the quantity of $X := \delta(1)\psi'(1) - \varphi'(1)$, which is used to calculate the left-hand side of (26) in the case $0 < p < 1$, can be obtained by differentiating both sides of (10.a), though we have no explicit expression without infinite summation for $\psi'(1)$ and $\varphi'(1)$.

Remark 4.1 It is seen that (24.a) and (25.b) do not agree with $E(W_i), i = 1, 2$ given by the formula (49) in [25] due to a slight difference of the service discipline for customers of

class-1 who arrive during a set-up time v_2 . Eqs (25.a) and (25.b) yield a generalization of the result by Takács [23] to the case with set-up times [29].

□

5. Some Comments

This section gives some comments on a multi-class priority queue with Bernoulli parameter $(p_1, p_2, p_3, \dots, p_N)$. By using similar notations to those in Section 3, the same argument as Section 3 leads to the following functional relationships for $\Pi_k(z_1, z_2, \dots, z_N), k = 1, 2, \dots, N$ corresponding to (6.a) and (6.b):

$$\Pi_k(z_1, z_2, \dots, z_N) = \frac{Q_k(z_1, z_2, \dots, z_N)}{z_k - p_k Q_k(z_1, z_2, \dots, z_N)} [\Psi_k(z_k, z_{k+1}, \dots, z_N) - \Psi_k(\delta_k, z_{k+1}, \dots, z_N) + \pi_0 r_k \{z_k R_k(z_1, z_2, \dots, z_N) - \delta_k R_k(z_1, \dots, z_{k-1}, \delta_k, z_{k+1}, \dots, z_N)\}] \tag{27.a}$$

where

$$\delta_k := \delta_k(z_1, z_2, \dots, z_{k-1}, z_{k+1}, \dots, z_N) : \text{one root in the unit circle } |z_k| \leq 1 \text{ for the equation } z_k - p_k Q_k(z_1, z_2, \dots, z_N) = 0. \tag{27.b}$$

The unknown functions $\Psi_k(z_k, z_{k+1}, \dots, z_N), k = 1, 2, \dots, N$ on the right-hand side of (27.a) should be determined from a set of functional equations [15, 29]:

$$\begin{aligned} & \sum_{k=i}^N \frac{z_k - Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)}{z_k - p_k Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)} \{ \psi_k(z_k, z_{k+1}, \dots, z_N) - \psi_k(\delta_k^*, z_{k+1}, \dots, z_N) \} \\ &= \left\{ \sum_{k=1}^{i-1} r_k u_{i,k} R_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_k, \dots, z_N) - 1 \right\} \pi_0 \\ &+ \sum_{k=i}^N \pi_0 r_k z_k R_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N) \left\{ 1 - \frac{z_k - Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)}{z_k - p_k Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)} \right\} \\ &+ \sum_{k=i}^N \pi_0 r_k \delta_k^* R_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_{k-1}, \delta_k^*, z_{k+1}, \dots, z_N) \\ &\quad \frac{z_k - Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)}{z_k - p_k Q_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_N)} \end{aligned} \tag{28.a}$$

for $i = 1, 2, 3, \dots, N$

where a set of $\{u_{i,k}; k = 1, 2, \dots, i - 1\}$ is defined by

$$u_{i,k} - H_k^* \left\{ \lambda \left(1 - \sum_{j=1}^{i-1} r_j u_{i,j} - \sum_{j=i}^N r_j z_j \right) \right\} = 0 \text{ for } i = 2, 3, \dots, N \text{ and } k = 1, 2, \dots, i - 1 \tag{28.b}$$

and

$$\delta_k^* := \delta_k(u_{i,1}, \dots, u_{i,i-1}, z_i, \dots, z_{k-1}, z_{k+1}, \dots, z_N). \tag{28.c}$$

(we assume that a null sum is equal to zero and that $\delta_1^* := \delta_1(z_2, z_3, \dots, z_N)$).

Equation (28.a) is a simultaneous functional equation for $\Psi_k(z_k, z_{k+1}, \dots, z_N)$ to be solved for determination of the generating functions $\Pi_k(z_1, z_2, \dots, z_N), k = 1, 2, \dots, N$. The

solution of (28.a) can be obtained successively from these equations starting from $k = N$ and by using the iterative schemes used in Section 3, however, a closed-form solution for an arbitrary number N can be hardly obtained, which is more complicated than the cases of the non-preemptive priority queue without set-up times [11, 30] and the exhaustive priority queue with set-up times [26]. That is, it is not impossible to solve Eq. (28.a), formally. This differs from the cyclic-service polling model with Bernoulli parameters (p_1, p_2, \dots, p_N) , for which any analytical method has not yet been given except for the case of $N = 2$ as described in [2, 5].

6. Concluding Remarks

We have presented a priority queueing model with Bernoulli schedules having set-up times. For the two-class priority queueing model, we have derived generating functions for joint queue-length distributions, *LSTs* for waiting time distributions and explicit forms with infinite summations for mean waiting times. We need rather the marginal distribution of queue-length for each priority class than the joint queue-length distribution. The *LSTs* of waiting time distributions might be derived by different methods. However, it is suggested from the analysis presented here that these *LSTs* and mean waiting times are expressed by the form with infinite summations using iterative schemes.

Further research will be extended to a generalization of arrival processes (e. g. correlated arrivals, batch arrivals and so on [27]) and to the optimization of switching ruoes [2, 3, 17, 32] and Bernoulli parameters. For example, using the results obtained here, it may be possible to determine an optimal probability $p = p^*$ so as to minimize a cost function defined by a linear combination of individual mean waiting times by help of computer programs. However, we have to find some effective iteration procedures so as to minimize inevitable truncation errors due to the infinite summations and infinite products.

Acknowledgements. The authors acknowledge many constructive comments by the referees.

References

- [1] O. J. Boxma: "Models of two queues: a few new views," *Teletraffic Analysis and Computer Performance Evaluation*, Eds. O. J. Boxma, J. W. Cohen and H. C. Times, pp. 75–98, North-Holland (1986).
- [2] O. J. Boxma: "Analysis and optimization of polling systems," *Queueing, Performance and Control in ATM (ITC-13)*, Eds. J. W. Cohen and C. D. Pack, pp. 173–183, North-Holland (1991).
- [3] S. Browne and U. Yechiali: "Dynamic priority rules for cyclic-type queues," *Adv. Appl. Prob.*, 21, 2, pp. 432–450 (1989).
- [4] J. W. Cohen and O. J. Boxma: "The $M/G/1$ queue with alternating service formulated as a Riemann-Hilbert problem," *Performance '81, Proc. 8th Int. Symp.*, Ed. F. J. Klysta, pp. 181–199 (1983).
- [5] J. W. Cohen and O. J. Boxma: "*Boundary Value Problems in Queueing System Analysis*," Mathematics Studies 79, Amsterdam, North-Holland (1983).
- [6] B. T. Doshi: "A note on stochastic decomposition in a $GI/G/1$ queue with vacations or set-up times," *J. Appl. Prob.*, 22, 2, pp. 419–428(1985).
- [7] S. S. Fratini: "Analysis of dynamic priority queue," *Commun. Statist.-Stochastic Models*, 6, 3, pp. 415–444(1990).
- [8] S. W. Fuhrmann and A. Moon: "Queues served in cyclic order with an arbitrary start-up distribution," *Naval Research Logistics*, 37, 1, pp. 123–133(1990).
- [9] M. Hofri: "Two queues and one server with threshold switching," *Teletraffic Analysis*

- and *Computer Performance Evaluation*, Eds. O. J. Boxma, J. W. Cohen and H. C. Tijms, Elsevier Science Publisher B. V., North-Holland (1986).
- [10] M. Hofri and K. W. Ross: "On the optimal control of two queues with server setup times and its analysis," *SIAM J. Comput.*, 16, 2, pp. 399–420 (1987).
- [11] N. K. Jaiswal: "*Priority Queues*," Academic Press, New York (1968).
- [12] T. Katayama: "A cyclic-service tandem queueing model with parallel queues in the first stage," *Commun. Statist.-Stochastic Models*, 4, 3, pp. 421–443 (1988).
- [13] T. Katayama: "A note on conservation laws for a multi-class service queueing system with setup times," Proceedings of symposium on queueing models of information and communications networks, Ed. M. Mori et al., Hamamatsu, Japan, pp. 201–212 (1991).
- [14] J. Keilson and L. D. Servi: "Oscillating random walk models for $GI/G/1$ vacation systems with Bernoulli schedules," *J. Appl. Prob.*, 23, pp. 790–802 (1986).
- [15] H. Kesten and J. Th. Runnenburg: "Priority in waiting-line problems, I and II," *Proc. Koninkl. Nederl. Akad. Wetensch.*, ser. A, 60, pp. 312–336 (1957).
- [16] H. Levy and L. Kleinrock: "A queue with starter and a queue with vacation: Delay analysis by decomposition," *Oper. Res.*, 34, 3, pp. 426–436 (1986).
- [17] H. Levy and M. Sidi: "Polling systems: Application, modeling, and optimization," *IEEE Trans. Commun.*, Com-38, 10, pp. 1750–1760 (1990).
- [18] M. Nomura and K. Tsukamoto: "Traffic analysis on polling systems," *The Trans. of the Institute of Electronics and Communication Engineers of Japan*, J61-B, 7, pp. 600–607 (1978) (in Japanese).
- [19] T. Ozawa: "Alternating service queues with mixed exhaustive and K -limited services," *Performance Evaluation*, 11, 3, pp. 165–175 (1990).
- [20] R. Ramaswamy and L. D. Servi: "The busy period of the $M/G/1$ vacation model with a Bernoulli schedule," *Commun. Statist.-Stochastic Models*, 4, 3, pp. 507–521 (1988).
- [21] L. D. Servi: "Average delay approximation of $M/G/1$ cyclic service queues with Bernoulli schedules," *IEEE J. on Selected Areas in Communications*, SAC-4, 6, pp. 813–822 (1986).
- [22] L. Takács: "*Introduction to the Theory of Queues*," Oxford University Press, New York (1962).
- [23] L. Takács: "Two queues attended by a single server," *Oper. Res.*, 16, 3, pp. 639–650 (1968).
- [24] H. Takagi: "*Analysis of Polling Systems*," MIT Press, Cambridge (1986).
- [25] H. Takagi: "Priority queues with setup times," *Oper. Res.*, 38, 4, pp. 667–677 (1990).
- [26] H. Takagi: "*Queueing Analysis: A Foundation of Performance Evaluation Vol. 1, Vacation and Priority Systems, Part 1*," Elsevier Science Publisher B. V., North-Holland (1991).
- [27] Y. Takahashi: "Batch input queueing models with multi-class customers and their applications to communication systems," Doctoral dissertation, Tokyo Inst. Tech. (1990).
- [28] Y. Tedijanto: "Exact results for the cyclic-service queue with a Bernoulli schedule," *Performance Evaluation*, 11, 1, pp. 107–115 (1990).
- [29] V. P. Vinnitskii and G. L. Brodetskii: "Alternating priorities and "Rest periods" in a single-server queueing system," *Cybernetics*, 9, 2, pp. 321–327 (1973).
- [30] P. D. Welch: "*Some contribution to the theory of priority queues*," Ph. D. Thesis, Columbia University, New York (1963).
- [31] P. D. Welch: "On a generalized $M/G/1$ queueing process in which the first customer of each busy period receives exceptional service," *Oper. Res.*, 12, 5, pp. 736–752 (1964).
- [32] U. Yechiali: "Optimal dynamic control of polling systems," *Queueing, Performance and Control in ATM (ITC-13)*, Edited by J. W. Cohen and C. D. Pack, pp. 205–217, North-Holland (1991).

Appendix. Validity of Equation (13.b)

If we restrict ourselves to the real space \mathbf{R} , all the functions appearing below (e. g., f and g) are real-valued.

(1) Convergence of $\{\varphi(y_i)\}$

From the expressions of $\delta(y)$ and $\sigma(x)$ [e. g. (14.d)], we have for the second derivatives of δ and σ

$$\delta''(y) > 0, \sigma''(x) > 0 \quad (y, x \in [0, 1])$$

and for the images of δ and σ

$$\begin{aligned} \delta([0, 1]) &\subseteq [0, 1] \quad (0 < \delta(0) < 1, \delta(1) = 1), \\ \sigma([0, 1]) &\subseteq [0, 1] \quad (0 < \sigma(0) < 1, 0 < \sigma(1) \leq 1). \end{aligned}$$

The graphs of $\delta(y)$ and $\sigma(x)$ are then illustrated as in Fig. 1. There exists an intersection point $P := (\delta(\xi), \xi)$ of $\{(x, y) \in [0, 1]^2 | x = \delta(y)\}$ and $\{(x, y) \in [0, 1]^2 | y = \sigma(x)\}$.

Recall that the sequence $\{y_i\}$ is defined as $y_i = \sigma\{\delta(y_{i-1})\}$. If we start with $y_0 \in [0, 1]$, y_i , then tends to ξ as $i \rightarrow \infty$. Since φ is continuous, $\varphi(y_i) \rightarrow \varphi(\xi)$ as $i \rightarrow \infty$.

(2) Convergence of the infinite summation in (13.b)

We define $s_n := \sum_{i=0}^n g\{y_i | y_0 = y\}$ for brevity. It is sufficient to prove that $\{s_n\}$ forms a Cauchy sequence, since \mathbf{R} is complete. Namely, we show that for any $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon) \geq 1$ such that

$$n, m \geq n_0 \text{ implies } |s_n - s_m| < \varepsilon. \tag{A.1}$$

We note that the idle probability π_0 is assumed to be positive ($\pi_0 > 0$). From the fact that $\varphi(y_i)$ converges to $\varphi(\xi)$, there exists $i_0 \geq 1$ such that

$$i \geq i_0 \text{ implies } |\varphi(y_i) - \varphi(\xi)| < \pi_0 \varepsilon / 2. \tag{A.2}$$

From equations (13.a) and (A.2), we have for all $n, m \geq i_0$

$$\begin{aligned} |s_n - s_m| &= \left| \sum_{i=m+1}^n g\{y_i | y_0 = y\} \right| \\ &= \frac{1}{\pi_0} |\varphi(y_{n+1}) - \varphi(y_{m+1})| \\ &\leq \frac{1}{\pi_0} \{|\varphi(y_{n+1}) - \varphi(\xi)| + |\varphi(\xi) - \varphi(y_{m+1})|\} \\ &< \varepsilon. \end{aligned}$$

Here, we assume that $n \geq m$. In the other case ($n < m$), the same equalities and inequalities are valid with $|\sum_{i=n+1}^m g\{y_i | y_0 = y\}|$ instead of $|\sum_{i=m+1}^n g\{y_i | y_0 = y\}|$ in the right-hand side of the first equation. The statement (A.1) thus follows if we select i_0 as n_0 . This completes the proof.

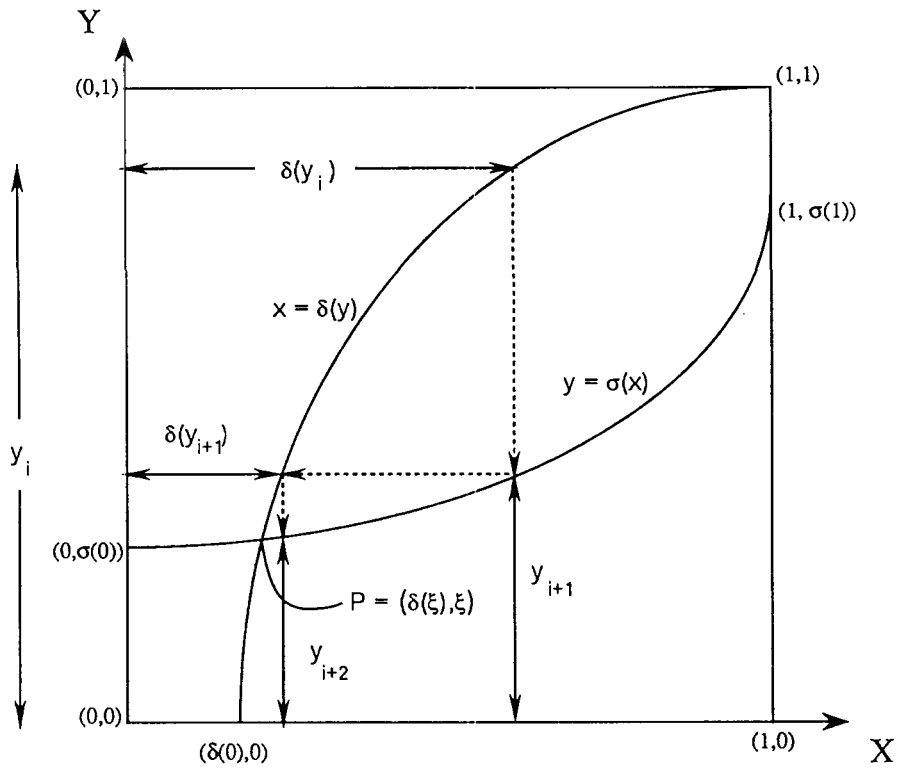


Fig.1- Iterative procedure for determining a sequence of $\{y_i\}$

Tsuyoshi Katayama
 Department of Electronics & Informatics
 Toyama Prefectural University
 Toyama 939-03, Japan