

## A QUEUEING MODEL FOR CUSTOMERS REQUESTING SERVICE POSITIONS AT A COUNTER

Ken'ichi Katou  
*Kanagawa University*

Yukio Takahashi<sup>1</sup>  
*Tokyo Institute of Technology*

(Received October 9, 2014; Revised September 18, 2015)

*Abstract* A new queueing model for customers requesting service positions at a counter is proposed. Each arriving customer requests a service at a particular position, and multiple servers provide services for these customers at their requested positions. The servers can attend different counter positions, but they may not change their order. This model was devised to evaluate the performance of a movable compact shelving system in a recently built university library, and it was used to plan that system.

In this paper, we analyze two simple cases, and then some variations of the model are discussed in connection with different service disciplines. The simulation results show some basic properties of the model. We present in some detail the application of our model to a movable compact shelving system.

**Keywords:** Queue, service counter, service position, movable compact shelving system

### 1. Introduction

We propose a new queueing model for customers requesting services at particular positions at a counter. In the model, each arriving customer requests a particular position at the counter (Figure 1). There are multiple servers at the counter, and at the start of each transaction, one of the servers who is assigned to the requested service moves to the requested position. We assume that the servers are not allowed to change their order at the counter (*the position order constraint*). This model was devised to evaluate the performance of a movable compact shelving system in a recently built university library, and it was used to plan the composition of that system (see Section 5). The model will be also applied to a gantry crane system at a wharf<sup>†</sup> and to some other storage and service systems.

In the model, we consider the following service station situation. The station has a waiting area and a counter. In the waiting area, customers form a queue in the order of arrival. At the counter, multiple servers provide services to customers at the positions requested by the customers. The servers can move along the counter, but their order cannot change. That is, suppose that there are  $c$  servers and that the  $i$ th server is at position  $x_i(t)$  at time  $t$ , where  $0 < x_i(t) < l$ ,  $i = 1, 2, \dots, c$ , and  $l$  is the length of the counter. The servers must satisfy the constraint  $0 < x_1(t) < x_2(t) < \dots < x_c(t) < l$  for any  $t$ . As mentioned above, we will refer to this as “the position order constraint”. The times required for the customers and servers to walk to new locations are disregarded in this model.

Consider the example depicted in Figure 1, in which  $l = 1$ . Server 1 is serving Customer 2 at position  $x_1 = 0.1$ , and Server 3 is serving Customer 1 at  $x_3 = 0.6$ . Server 2 is idle at  $x_2 = 0.4$ , but this server cannot serve Customer 3, who is waiting in the queue, because

---

\*Professor Emeritus

<sup>†</sup>The authors are grateful to Dr. Yutaka Takahashi, Professor, Kyoto University, for suggesting this application.

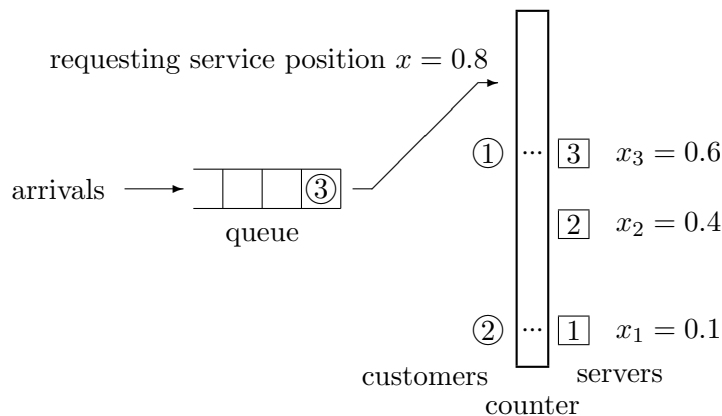


Figure 1: A queueing model with service-position requests

Customer 3 is requesting a service at position  $x = 0.8$ , and Server 2 is not allowed to pass Server 3. Thus, Customer 3 must wait in the queue until Server 3 has finished serving Customer 1.

We will refer to a model of this type<sup>‡</sup> as a *queueing model with service-position requests*. In order to specify a queueing model, it is necessary to clarify the input process for arriving customers, the service-time distribution, the service-position distribution, the number of servers, the service discipline, and other factors. We will use a variant of Kendall's notation:  $A/B,C/c(\text{ordered})/D$  for a  $c$ -server model with service-position requests, where  $A$  denotes the input process,  $B$  denotes the service-time distribution, and  $D$  denotes the capacity of the queueing system or the service discipline. The new component  $C$  denotes the service-position distribution. In some models, the service discipline can be complex and must be stated in words instead of by a symbol.

We note that if  $C$  is a continuous distribution on the interval  $(0, l)$  with distribution function  $F(x)$ , the behavior of the model  $A/B,C/c(\text{ordered})/D$  is substantially equivalent to that of the model  $A/B,U/c(\text{ordered})/D$ , where  $U$  represents the uniform distribution on  $(0, 1)$ . This is because the order constraint is a condition on the relative positions of the servers, and there is a one-to-one correspondence between the service positions of the two models:  $x \in (0, l) \leftrightarrow F(x) \in (0, 1)$ . Hence, most of the important characteristic quantities are the same in both models, and we may consider the model  $A/B,U/c(\text{ordered})/D$  to be a standard queueing model with service-position requests when the service-position distribution is continuous. On the other hand, when  $C$  is a discrete distribution, the characteristic quantities may depend on the individual probabilities of the atoms of the distribution, and it is not easy to find a standard model.

The purpose of this paper is to propose queueing models with service-position requests and to investigate their basic properties through analyses and simulations. The remainder of this paper is constructed as follows. We first analyze two simple cases, in Section 2, a loss model  $M/M,U/2(\text{ordered})/\text{loss}$ , and in Section 3, a saturated model  $*/M,U/2(\text{ordered})$  with a first-come first-served (FCFS) service rule. Proofs of theorems and corollaries are presented in the appendices. In Section 4, we discuss variations of the model brought by different service disciplines, and some simulation results are investigated to see typical behaviors of the model. Section 5 shows how the model was used in the application to the

<sup>‡</sup>We may include the single server case with  $c = 1$ , though the position order constraint becomes meaningless. A single server model is essentially same as the corresponding ordinary single server queue except that service positions differ for every customer.

movable shelving system. It is a typical application of this model, and it provides some clues for future studies. Readers who are mainly interested in the application are recommended to start reading from Section 5, and then to go back to Section 4.

## 2. Loss Model M/M,U/2(ordered)/loss

As one of the simplest queueing models with service-position requests, we first consider an M/M-type loss model with two servers.

### Model

We consider an M/M,U/2(ordered)/loss model. There are two servers at the counter. Customers arrive at the service station via a Poisson process with rate  $\lambda$ . Each customer requests a position at the counter, where they will be served by one of the servers. Service times are subject to a common exponential distribution with mean  $1/\mu$ , and service positions are subject to the uniform distribution on  $(0, l)$ , where  $l$  is the length of the counter. We will use  $l$  (rather than setting  $l = 1$ ) in order to clarify equations and results. We assume that service times and service positions are mutually independent, and they are also independent from the input process.

We will call the two servers the left server and the right server. They must satisfy the position order constraint stated in the previous section. Namely, the left server cannot move to the right of the right server, and the right server cannot move to the left of the left server.

### Service discipline

Services for customers are executed according to the following service discipline. To save space, alternative cases are listed in brackets.

1. Upon arrival: Suppose that a customer arrives at the service station and requests service position  $x \in (0, l)$ .
  - (a) If both servers are busy, then this customer is lost (i.e., the customer is denied service and leaves the station).
  - (b) If the left [right] server is idle and the right [left] server is serving another customer at position  $y \in (0, l)$ , then (i) if  $y > x$  [if  $y < x$ ], the left [right] server moves to position  $x$  and serves the new customer, or (ii) if  $y \leq x$  [if  $y \geq x$ ], this customer is lost.
  - (c) If both servers are idle and if  $x \leq \frac{1}{2}l$  [if  $x > \frac{1}{2}l$ ], then the left [right] server moves to position  $x$  and serves the customer.
2. Upon completion:
  - (a) When the service of the left [right] server is completed, the server becomes idle and moves to position 0 [to position  $l$ ].

**Remark 2.1.** *The service discipline described above is just one of many possible disciplines. For example, another service discipline, which also seems natural, is derived by changing rules 2. (a) and 1. (c) above as follows (see Section 4).*

2. (a') *When a service is completed, the server becomes idle and remains where the previous service was performed.*
1. (c') *Upon the arrival of a customer requesting a service at position  $x$ , if both servers are idle, the nearest server moves to position  $x$  and serves the customer.*

*The authors think that this change makes the analysis of the model more difficult in this case.*

## Balance equations

In order to analyze the model with a Markov process, we define the following states of the system:

$$\begin{aligned}
 S_0 &= \{\text{there are no customers at the station}\}, \\
 S_1(x) &= \{\text{the left server is serving a customer at position } x, \\
 &\quad \text{and the right server is idle}\}, \quad x \in (0, l), \\
 S_2(y) &= \{\text{the right server is serving a customer at position } y, \\
 &\quad \text{and the left server is idle}\}, \quad y \in (0, l), \quad \text{and} \\
 S_3(x, y) &= \{\text{the left server is serving a customer at position } x, \\
 &\quad \text{and the right server is serving a customer at position } y\}, \quad 0 < x < y < l.
 \end{aligned}$$

Then, the stochastic behavior of the system can be represented by a continuous-time Markov process  $\{X(t)\}$  on the state space

$$S = S_0 \cup \left( \bigcup_{x \in (0, l)} S_1(x) \right) \cup \left( \bigcup_{y \in (0, l)} S_2(y) \right) \cup \left( \bigcup_{0 < x < y < l} S_3(x, y) \right). \quad (2.1)$$

The infinitesimal generator of the transition function of  $X(t)$  is described in Appendix A. We note that the transition function is time homogeneous, the transition rates are uniformly bounded, and the sample path of the process is a step function with probability one [1], and thus its basic properties can be easily generalized from those of Markov processes with a finite number of states (e.g., see Theorem 9.10, p.244 [2]).

As seen in Appendix A, the infinitesimal generator is rather complicated, and it is not easily analyzed. Instead, we shall approach the analysis through the balance equations. We assume that the Markov process  $\{X(t)\}$  has a steady state, and denote the steady state probability and steady state probability densities as follows:

$$\begin{aligned}
 f_0 &: \text{the probability that the state is in } S_0, \\
 f_1(x) &: \text{the probability density that the state is in } S_1(x), \quad x \in (0, l), \\
 f_2(y) &: \text{the probability density that the state is in } S_2(y), \quad y \in (0, l), \quad \text{and} \\
 f_3(x, y) &: \text{the joint probability density that the state is in } S_3(x, y), \quad 0 < x < y < l.
 \end{aligned}$$

Then, (2.2) through (2.6) are the balance equations for the steady state, where  $h_{1/2}(x)$  takes the value of 1 for  $x \in (0, \frac{1}{2}l)$ , and 0 otherwise. The meaning of each equation is explained below. Note that, for example,  $l^{-1}\lambda dt dx$  is the probability that a customer arrives at the counter in time interval  $(t, t + dt)$  and requests a service position in the interval  $(x, x + dx)$

for some suitable  $t$  and  $x$  and for small quantities  $dt > 0$  and  $dx > 0$ .

$$\lambda f_0 = \mu \int_0^l f_1(x) dx + \mu \int_0^l f_2(y) dy, \tag{2.2}$$

$$\{\mu + l^{-1}\lambda(l-x)\} f_1(x) = l^{-1}\lambda h_{l/2}(x) f_0 + \mu \int_x^l f_3(x,y) dy, \quad 0 < x < l, \tag{2.3}$$

$$\{\mu + l^{-1}\lambda y\} f_2(y) = l^{-1}\lambda\{1 - h_{l/2}(y)\} f_0 + \mu \int_0^y f_3(x,y) dx, \quad 0 < y < l, \tag{2.4}$$

$$2\mu f_3(x,y) = l^{-1}\lambda f_1(x) + l^{-1}\lambda f_2(y), \quad 0 < x < y < l, \quad \text{and} \tag{2.5}$$

$$1 = f_0 + \int_0^l f_1(x) dx + \int_0^l f_2(y) dy + \iint_{0 < x < y < l} f_3(x,y) dx dy. \tag{2.6}$$

- The left-hand side of (2.2) is the rate at which the system moves from state  $S_0$  due to the arrival of a customer, and the right-hand side is the rate at which the system moves into state  $S_0$  due to the completion of the service provided by the left server in state  $S_1(x)$  or due to the completion of the service provided by the right server in state  $S_2(y)$ .
- The left-hand side of (2.3) is the rate at which the system moves from state  $S_1(x)$  due to the completion of the service by the left server or due to the arrival of a customer who can be served by the right server. The right-hand side of (2.3) is the rate at which the system moves into state  $S_1(x)$  due to the arrival of a customer who can be served by the left server or due to the completion of the service by the right server when both servers are busy.
- Equation (2.4) is derived in a way similar to that of (2.3) by exchanging the roles of the servers on the left and on the right.
- The left-hand side of (2.5) is the rate at which the system moves from state  $S_3(x,y)$  due to the completion of a service, and the right-hand side of (2.5) is the rate at which the system moves into state  $S_3(x,y)$  due to the arrival of a customer when one of the servers is busy and the other is idle.
- Equation (2.6) is a constraint on the total probability.

### Solution to the balance equations

By exploiting the special relation (2.5), we can solve the balance equations. The solution is presented in the theorem below. A formal proof of the theorem is given in Appendix A. To save space here, a rough idea of the derivation process is shown in a footnote<sup>§</sup>. In order to simplify the equations, we let  $\rho = \frac{\lambda}{2\mu}$ .

---

<sup>§</sup>From the model structure, we can easily guess the symmetric relation (2.10). By using (2.5) and (2.10), we can reduce the balance equations (2.2) to (2.5) into an integral equation for  $f_1(x)$  with  $f_0$  as a parameter. Taking derivatives and using the method of separation of variables, we obtain the solution (2.8) for  $f_1(x)$ , which includes  $f_0$  as a multiplicative constant. The value of  $f_0$  can be determined by using (2.6).

**Theorem 2.1.** *The balance equations (2.2) to (2.6) have the following solution:*

$$f_0 = \left[ (1 + \rho)^2 + (2 + \rho) \log \frac{(1 + \frac{1}{2}\rho)^2}{1 + \rho} \right]^{-1}, \quad (2.7)$$

$$f_1(x) = \begin{cases} l^{-1}C_1 \left[ \frac{(2 + \rho)l}{l + \rho(l - x)} + \log \frac{l + \rho x}{l + \rho(l - x)} + C_2 \right], & 0 < x \leq \frac{1}{2}l, \\ -l^{-1}C_1 \left[ \frac{(2 + \rho)l}{l + \rho(l - x)} + \log \frac{l + \rho x}{l + \rho(l - x)} - C_2 \right], & \frac{1}{2}l < x < l, \end{cases} \quad (2.8)$$

$$\text{where } C_1 = \frac{\rho}{2 + \rho} f_0 \quad \text{and} \quad C_2 = 2 + \rho + \log\{1 + \rho\}, \quad (2.9)$$

$$f_2(y) = f_1(l - y), \quad 0 < y < l, \quad \text{and} \quad (2.10)$$

$$f_3(x, y) = l^{-1}\rho \{f_1(x) + f_2(y)\}, \quad 0 < x < y < l. \quad (2.11)$$

**Remark 2.2.** (a) *The above solution of the balance equations (2.2) to (2.6) is unique (we omit the proof) and it gives a stationary distribution of the Markov process  $\{X(t)\}$ .*

(b) *We see that the system has a stationary distribution for any choice of positive  $\lambda$  and  $\mu$ , since the value of  $f_0$  in (2.7) is always finite. This follows from the loss model assumption.*

(c) *The density function  $f_1(x)$  is not continuous at  $x = \frac{1}{2}l$ . This is because the equation (2.3) includes a discontinuous function  $h_{l/2}(x)$ .*

### Loss probability

The loss probability, i.e., the probability in the steady state that an arriving customer is lost, can be calculated as

$$P_* = l^{-1} \int_0^l x f_1(x) dx + l^{-1} \int_0^l (l - y) f_2(y) dy + \iint_{0 < x < y < l} f_3(x, y) dy. \quad (2.12)$$

The (conditional) loss probability  $P(a)$  for a customer requesting service position  $a \in (0, l)$  can be calculated by

$$P(a) = \int_a^l f_1(x) dx + \int_0^a f_2(y) dy + \iint_{0 < x < y < l} f_3(x, y) dy. \quad (2.13)$$

By performing the integrations in these equations, we obtain the following corollary (the proof is given in Appendix A).

**Corollary 2.1.** *The loss probabilities are given by*

$$P_* = 1 - \frac{1}{\rho} + \left(1 + \frac{1}{\rho}\right) f_0 = 1 - \left\{ 1 + \rho + \frac{2 + \rho}{\rho} \log \frac{\{1 + \frac{1}{2}\rho\}^2}{1 + \rho} \right\} f_0, \quad \text{and} \quad (2.14)$$

$$P(a) = \begin{cases} 1 - \left\{ 1 + \rho + \log \frac{\{1 + \rho\}\{l + \rho a\}}{l + \rho(l - a)} \right\} f_0, & 0 < a \leq \frac{1}{2}l, \\ 1 - \left\{ 1 + \rho + \log \frac{\{1 + \rho\}\{l + \rho(l - a)\}}{l + \rho a} \right\} f_0, & \frac{1}{2}l < a < l. \end{cases} \quad (2.15)$$

Note that  $P(a)$  is decreasing on  $a$  in the interval  $(0, \frac{1}{2}l)$ , and it is increasing in  $(\frac{1}{2}l, l)$ . This means that  $P(a)$  depends heavily on  $a$ , and a customer who requests a service position near the middle of the counter is less likely to be lost than a customer who requests a service position near either of the ends (see Figure 3).

When the traffic intensity  $\rho$  is very small, the asymptotic loss probabilities are those shown in (2.16) and (2.17) (the proof is given in Appendix A).

**Corollary 2.2.** *As  $\rho$  tends to zero, we have*

$$P_* = \frac{1}{2}\rho + o(\rho), \quad \text{and} \tag{2.16}$$

$$P(a) = \frac{|2a - l|}{l}\rho + o(\rho), \quad 0 < a < l. \tag{2.17}$$

If we remove the position order constraint from our model assumptions, we can analyze the model as if it were the ordinary M/M/2/2 model. On the other hand, if we add to the model the constraints that the left server can only serve customers requesting service positions in the interval  $(0, \frac{1}{2}l)$ , and the right server can only serve customers requesting service positions in  $(\frac{1}{2}l, l)$ , then the model can be regarded as a composite of two ordinary M/M/1/1 queues. Let  $P_*^{M/M/1/1}$  be the loss probability of the ordinary M/M/1/1 model with traffic intensity  $\rho$ , and let  $P_*^{M/M/2/2}$  be that of the ordinary M/M/2/2 model with the same traffic intensity. Then we can easily prove the following inequalities (the proofs are given in Appendix A).

**Corollary 2.3.** *For any  $\rho > 0$ ,*

$$P_*^{M/M/2/2} = \frac{2\rho^2}{1 + 2\rho + 2\rho^2} < P_* < \frac{\rho}{1 + \rho} = P_*^{M/M/1/1}. \tag{2.18}$$

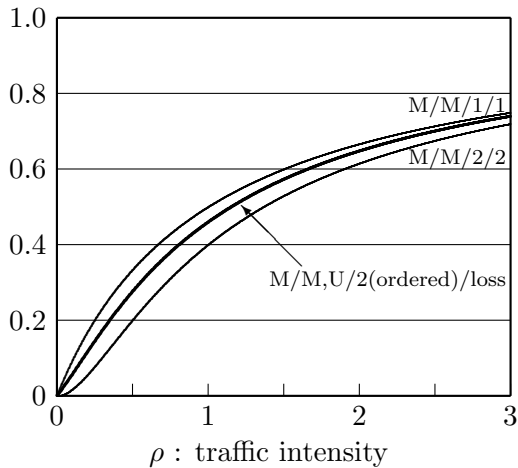
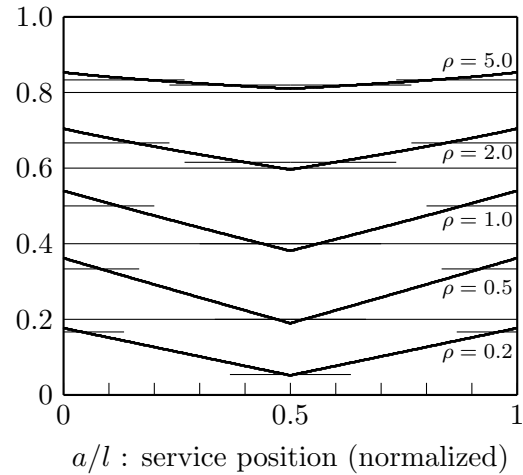
When the traffic is light, we have  $P_* = \frac{1}{2}\rho + o(\rho)$  from (2.16), which is about half the size of  $P_*^{M/M/1/1} = \rho + o(\rho)$  but far larger than  $P_*^{M/M/2/2} = 2\rho^2 + o(\rho^2)$ .

**Some numerical results**

We shall now show some numerical results for the loss probabilities (2.14) and (2.15). Figure 2 shows a graph of  $P_*$  vs.  $\rho$ , together with corresponding graphs of  $P_*^{M/M/1/1}$  and  $P_*^{M/M/2/2}$ . As stated in Corollary 2.3,  $P_*$  is between  $P_*^{M/M/1/1}$  and  $P_*^{M/M/2/2}$ .

Figure 3 shows graphs of  $P(a)$  for various values of  $\rho$ . These are V-shaped, and we see that customers requesting service positions near an end of the counter have larger loss probabilities. In the figure, the value of  $P_*^{M/M/1/1}$  is indicated by thin horizontal lines near  $a/l = 0$  and 1, and the value of  $P_*^{M/M/2/2}$  is indicated by a thin horizontal line near  $a/l = 0.5$ . These graphs show that  $P(0)$  is a little bit larger than  $P_*^{M/M/1/1}$ , and  $P(\frac{1}{2}l)$  is a little bit less than  $P_*^{M/M/2/2}$ . (These properties can be proved analytically from (2.15), but we omit the proof.) The difference between  $P(0)$  and  $P_*^{M/M/1/1}$  is less than 9%, and the difference between  $P(\frac{1}{2}l)$  and  $P_*^{M/M/2/2}$  is less than 6%.

Intuitively, these properties might be roughly interpreted as follows (though this is not a strict discussion). Near the ends of the counter, customers tend to receive services from only a specific server, as in the M/M/1/1 model, but near the center of the counter, customers can receive services from either of two servers, as in the M/M/2/2 model.

Figure 2: Loss probability  $P_*$ Figure 3: Loss probability  $P(a)$ 

### 3. Saturated Model $*/M,U/2(\text{ordered})/FCFS$

As a second example of a simple queueing model with service-position requests, we consider an M/M-type saturated model with two servers, in which there are always customers in the queue, and they are served on an FCFS basis.

#### Model

We consider a saturated  $*/M,U/2(\text{ordered})$  model. To represent the saturated situation, we assume there to exist infinitely many customers waiting in the queue. Service times are subject to a common exponential distribution with mean  $1/\mu$ , and service positions are subject to the uniform distribution on  $(0, l)$ , where  $l$  is the length of the counter. We assume service times and service positions to be mutually independent.

There are two servers at the counter. Again, we will call them the left server and the right server. They satisfy the position order constraint. Services for customers are provided strictly in an FCFS way. We will refer to this model as  $*/M,U/2(\text{ordered})/FCFS$ .

Services for customers are executed according to the following service discipline.

#### Service discipline

A new service is started immediately after the previous one is completed. We will examine the moment that a service is completed. Suppose that, just prior to that, the left server is at position  $x$ , the right server is at position  $y$ , the customer at the head of the queue (Customer 1) requests service position  $z$ , and the next customer in the queue (Customer 2) requests service position  $z'$ , where  $0 < x < y < l$  and  $0 < z, z' < l$ . New services are started according to the rules (a) and (b) below. To save space, alternatives are in brackets. We note that when Customer 1 is served, every other customer in the queue moves one place ahead, and when Customer 2 is served, the customers again move one place ahead. The new top two customers are now called Customer 1 and Customer 2.

- (a) If the left [right] server completes a service while the right [left] server is still serving, then (i) if  $z < y$  [if  $z > x$ ], Customer 1 is served at position  $z$  by the left [right] server, or (ii) if  $z \geq y$  [if  $z \leq x$ ], no new service is begun and the left [right] server becomes idle.
- (b) If the left [right] server completes a service while the right [left] server is idle (this occurs only when  $z \leq x < y$  [only when  $x < y \leq z$ ]), then Customer 1 is served at position  $z$  by the left [right] server. Further, (i) if  $z' > z$  [if  $z' < z$ ], then Customer 2 will then be served by the right [left] server, or (ii) if  $z' \leq z$  [if  $z' \geq z$ ], the right [left] server will



continue to be idle.

**The balance equations**

In order to analyze the model using a Markov process, we consider the following states of the system.

$$S_1(x, z) = \{ \text{the left server is serving a customer at position } x, \text{ the right server is idle, and Customer 1 requests service position } z \}, \quad 0 < z < x < l,$$

$$S_2(y, z) = \{ \text{the right server is serving a customer at position } y, \text{ the left server is idle, and Customer 1 requests service position } z \}, \quad 0 < y < z < l,$$

$$S_3(x, y) = \{ \text{the left server is serving a customer at position } x, \text{ and the right server is serving a customer at position } y \}, \quad 0 < x < y < l.$$

The stochastic behavior of the system can be represented by a continuous-time Markov process  $\{X(t)\}$  on the state space

$$S = \left( \bigcup_{0 < z < x < l} S_1(x, z) \right) \cup \left( \bigcup_{0 < y < z < l} S_2(y, z) \right) \cup \left( \bigcup_{0 < x < y < l} S_3(x, y) \right). \tag{3.1}$$

Here, we do not show the infinitesimal generator, but the Markov process has stationary transition functions, its transition rates are uniformly bounded, and its sample path is a step function with probability 1.

**Remark 3.1.** *The state space above is somewhat redundant. For the purpose of deriving a Markov process, the auxiliary variable  $z$  can be omitted. However, it is easier to write the balance equations with auxiliary variable  $z$  for the state representation, and so we use the states described above.*

Assuming the existence of the steady state of the Markov process  $\{X(t)\}$ , we denote the state probability densities in the steady state as follows:

$f_1(x, z)$  : the joint probability density that the state is in  $S_1(x, z)$ ,  $0 < z < x < l$ ,

$f_2(y, z)$  : the joint probability density that the state is in  $S_2(y, z)$ ,  $0 < y < z < l$ , and

$f_3(x, y)$  : the joint probability density that the state is in  $S_3(x, y)$ ,  $0 < x < y < l$ .

$$\tag{3.2}$$

We then have the following balance equations. The meaning of each equation is explained below.

$$\mu f_1(x, z) = \mu l^{-1} \int_x^l f_1(x', x) dx' + \mu l^{-1} \int_x^l f_3(x, y') dy', \quad 0 < z < x < l, \tag{3.3}$$

$$\mu f_2(y, z) = \mu l^{-1} \int_0^y f_2(y', y) dy' + \mu l^{-1} \int_0^y f_3(x', y) dx', \quad 0 < y < z < l, \tag{3.4}$$

$$2\mu f_3(x, y) = \mu l^{-1} \int_x^l f_1(x', x) dx' + \mu l^{-1} \int_0^y f_2(y', y) dy' + \mu l^{-1} \int_0^y f_3(x', y) dx' + \mu l^{-1} \int_x^l f_3(x, y') dy', \quad 0 < x < y < l, \tag{3.5}$$

$$1 = \iint_{0 < z < x < l} f_1(x, z) dx dz + \iint_{0 < y < z < l} f_2(y, z) dy dz + \iint_{0 < x < y < l} f_3(x, y) dx dy \tag{3.6}$$

- The left-hand side of (3.3) is the rate at which the system moves from state  $S_1(x, z)$  due to the left server completing a service. The right-hand side of (3.3) is the rate at which the system moves into state  $S_1(x, z)$ . It consists of two terms:
  - (1) The first term comes from the event that a service is completed by the left server at position  $x'$  in state  $S_1(x', x)$  (with rate  $\mu$ ) and a new service for Customer 1 is begun by the same server at position  $x$  while Customer 2 requests service position  $z$  (with density  $l^{-1}$ ), where  $z < x$ ; this is case (b) (ii) in the service rule above.
  - (2) The second term comes from the event that a service is completed by the right server at position  $y'$  in state  $S_3(x, y')$  (with rate  $\mu$ ) and the server then becomes idle since Customer 1 requests service position  $z$  (with density  $l^{-1}$ ), where  $z < x$ ; this is alternate case (a) (ii) (in brackets) in the service rule above.
- The leading term in equation (3.4) is found by exchanging the roles of the left server and the right server in (3.3). Note that the right-hand side of (3.4) is derived from two cases in the service rule above: alternate (b) (ii) (in brackets) and (a) (ii).
- The left-hand side of (3.5) is the rate at which the system moves from state  $S_3(x, y)$  due to the completion of a service. The right-hand side of (3.5) is the rate at which the system moves into state  $S_3(x, y)$ . It consists of four terms:
  - (1) the first term comes from the event that a service is completed by the left server at position  $x'$  in state  $S_1(x', x)$  (with rate  $\mu$ ), the same server begins a new service for Customer 1 at position  $x$ , and the right server begins a new service for Customer 2 at position  $y$  (with density  $l^{-1}$ ), where  $y > x$ ; this is case (b) (i) in the service rule above.
  - (2) The second term comes from the event corresponding to the one in (1) above but where the roles of the left and right servers are exchanged; this is alternate case (b) (i) (in brackets).
  - (3) The third term comes from the event that a service is completed by the left server at position  $x'$  in state  $S_3(x', y)$  (with rate  $\mu$ ), the same server begins a new service for Customer 1 at position  $x$  (with density  $l^{-1}$ ), where  $x < y$ , and the right server continues to deliver a service at position  $y$ ; this is case (a) (i).
  - (4) The fourth term comes from the corresponding event in (3) above but where the roles of the left and right servers are exchanged; this is alternate case (a) (i) (in brackets).
- Equation (3.6) is a constraint on the total probability.

### Solution to the balance equations

Using the special structures in (3.3) to (3.5), we can solve the balance equations as in Theorem 3.1 below. A formal proof of the theorem is given in Appendix B. We show a rough idea of the derivation process in a footnote<sup>¶</sup>.

**Theorem 3.1.** *We let*

$$g(x) = \frac{54l^2}{3 + 4 \log 2} \cdot \frac{l - x}{(2l - x)^2(l + x)^3}. \quad (3.7)$$

*Then the balance equations (3.3) to (3.6) have a solution given by*

$$f_1(x, z) = g(x), \quad 0 < z < x < l, \quad (3.8)$$

$$f_2(y, z) = f_1(l - y, l - z) = g(l - y), \quad 0 < y < z < l, \quad \text{and} \quad (3.9)$$

$$f_3(x, y) = \frac{1}{2} \{g(x) + g(l - y)\}, \quad 0 < x \leq y < l. \quad (3.10)$$

**Remark 3.2.** *The above solution is unique (we omit the proof), and it gives a stationary distribution of the Markov process  $\{X(t)\}$ .*

**Throughput**

From the solution given in Theorem 3.1, we can calculate the throughput of our model. It is given in the following corollary (the proof is given in Appendix B).

**Corollary 3.1.** *The throughput  $\Lambda$  of the saturated model  $*/M,U/2(\text{ordered})/FCFS$  is given by*

$$\Lambda = \frac{9}{3 + 4 \log 2} \mu \quad (= 1.559\mu). \tag{3.11}$$

**Remark 3.3.** *The throughput is less than  $2\mu$ . This is because in the model, the servers have some idle periods even in the saturated situation.*

**4. Models  $M/M,U/c(\text{ordered})/FCFS$  and  $M/M,U/c(\text{ordered})/PSP$**

In Sections 2 and 3, we considered very simple models. These were the only models that the authors could analyze. In this section, using simulations, we will study some more-complex  $M/M$ -type queueing models with service-position requests.

There are a variety of queueing models with service-position requests. We may choose any arrival process, any service-time distribution, and any service-position distribution. The number of servers is also an important parameter. Further, diverse service disciplines exist. Compared with ordinary queueing models, the service disciplines in these models are complex. This is because they give not only the rules by which customers are served, but also the rules by which servers are selected to serve the customers.

Here we introduce two examples of service disciplines. For a customer who is requesting service position  $x$ , a server will be said to be *available* if that server is idle and is able to move to  $x$  under the position order constraint by moving, if necessary, other idle servers.

**Discipline 1: the first-come first-served (FCFS) discipline**

We begin with the FCFS discipline, which is the one used in the saturated model  $*/M,U/2(\text{ordered})/FCFS$  in Section 3. In this discipline, customers are assigned to servers in the order of their arrival, as follows.

1. Upon arrival: A customer arrives at the service station.
  - (a) If there are no customers in the queue and if there is at least one idle server available, then the customer is immediately served by the *nearest* available server.
  - (b) Otherwise (that is, there is at least one customer in the queue or no server is available), the customer goes to the end of the queue.
2. Upon completion of a service: The service being performed by Server  $i$  is completed.

---

¶From the right-hand sides of (3.3) and (3.4), we can easily see that  $f_1(x, z)$  and  $f_2(y, z)$  are independent of  $z$ , that is, constants as functions of  $z$ . Further, by comparing the right-hand sides of (3.3) to (3.5), we see that  $f_3(x, y) = \frac{1}{2}\{f_1(x, z) + f_2(y, z)\}$ . From the structure of the model, we can assume the symmetry relation  $f_2(y, z) = f_1(l - y, l - z)$ . Thus, we see that the functions  $f_1(x, y)$ ,  $f_2(y, z)$ , and  $f_3(x, y)$  are given as functions of  $g(x)$  by (3.8) to (3.10), respectively, where  $g(x)$  satisfies the integral equation

$$(l + x)g(x) = \int_0^{l-x} g(z) dz + 2 \int_x^l g(z) dz.$$

To derive equation (3.7) from the above integral equation and (3.6), we used a series expansion technique.

- (a) If there are no customers in the queue or if Server  $i$  is not available for the customer at the head of the queue, then Server  $i$  becomes idle and remains in the same position.
- (b) If there is at least one customer waiting in the queue and Server  $i$  is available for the customer at the head of the queue, then that customer is immediately served by Server  $i$  and every other customer (if any) in the queue moves one place ahead. In this case, we also check (c), below.
- (c) If there is at least one customer waiting in the queue, and there is at least one server available for the customer at the head of the queue, then that customer is served by the nearest available server, and every other customer (if any) in the queue moves one place ahead. This step (c) is repeated until the queue is emptied or there is no available server for the customer at the head of the queue.

### Discipline 2: the performable service priority (PSP) discipline

The second service discipline is a more practical one, which we will call the *performable service priority* (PSP) discipline. In this discipline, a customer in the queue for whom at least one server is available has priority over customers for whom there is no server available. The assignment rule is further specified below. Suppose that there are  $c$  servers at the counter and that they are numbered from left to right as Server 1, Server 2,  $\dots$ , Server  $c$ . We also introduce two imaginary servers, Server 0 and Server  $c + 1$ . Server 0 is considered to be always busy at position 0, and Server  $c + 1$  is always busy at position  $l$ .

1. Upon arrival: A customer arrives at the service station.
  - (a) If there is at least one available server, the customer is served immediately by the nearest available server.
  - (b) Otherwise (that is, if there is no available server), the customer goes to the end of the queue.
2. Upon completion of a service: Server  $i$  completes a service.
  - (a) If there are no customers in the queue for whom Server  $i$  is available, then Server  $i$  becomes idle and remains in position.
  - (b) If there is at least one customer in the queue for whom Server  $i$  is available, new services are begun, as follows. Note that in this case one of Servers  $i-1$  and  $i+1$  or both are busy.
    - (i) When Servers  $i-1$  and  $i+1$  are both busy, the first customer in the queue for whom Server  $i$  is available proceeds to the counter and is served by Server  $i$ . All customers who were behind that customer move one place ahead.
    - (ii) When Server  $i-1$  is busy and Server  $i+1$  is idle, then there exists some  $j$  ( $> i$ ) such that Server  $j$  is busy but Servers  $i, i+1, \dots, j-1$  are idle. These  $j-i$  idle servers can provide services for at most  $j-i$  customers in the queue for whom Server  $i$  is available, and so it is necessary to determine a rule for assigning servers to customers; we use the following rule.
      - (1) When there are greater than or equal to  $j-i$  customers in the queue for whom Server  $i$  is available, the first  $j-i$  of these customers are served. We sort these  $j-i$  customers according to the service positions requested and assign Servers  $i, i+1 \dots j-1$  in the order of their positions. The remaining customers in the queue reform the queue in the order in which they arrived.
      - (2) When there are  $k$  ( $< j-i$ ) customers in the queue for whom Server  $i$  is available, we sort these  $k$  customers according to the service positions requested and assign Servers  $i, i+1, \dots, i+k-1$ . The remaining customers

reform the queue in the order in which they arrived.

- (iii) When Server  $i - 1$  is idle and Server  $i + 1$  is busy, then for some  $j (< i)$ , Server  $j$  is busy but Servers  $j + 1, j + 2, \dots, i$  are idle. These  $i - j$  idle servers can provide services for at most  $i - j$  customers for whom Server  $i$  is available. The rule for assigning servers is similar to the one used in (ii) (1) and (2) above, but the ordering of the service positions is reversed.

**M/M,U/c(ordered)/FCFS and M/M,U/c(ordered)/PSP models**

We now consider M/M-type,  $c$ -server models with Discipline 1 or Discipline 2 as the service discipline. We will refer to the model as M/M,U/c(ordered)/FCFS when Discipline 1 is used, and as M/M,U/c(ordered)/PSP when Discipline 2 is used.

We let  $\lambda$  be the arrival rate and  $\mu$  be the service rate of each server. The traffic intensity is defined by  $\rho = \frac{\lambda}{c\mu}$ . We denote by  $W_{q^*}$  the mean waiting time and by  $W_q(a)$  the (conditional) mean waiting time for a customer requesting service position  $a, 0 < a < l$ . We denote by  $\rho_{\max}$  the stability threshold; that is, the model is stable for  $\rho < \rho_{\max}$  and is unstable for  $\rho > \rho_{\max}$ . For the model M/M,U/c(ordered)/FCFS, we attach the superscript FCFS, such as  $W_{q^*}^{\text{FCFS}}$ , and for the model M/M,U/c(ordered)/PSP, we attach the superscript PSP. Similarly, we denote by  $W_{q^*}^{\text{M/M}/c}$  the mean waiting time in the ordinary M/M/c model.

**Remark 4.1.** Models M/M,U/c(ordered)/FCFS and M/M,U/c(ordered)/PSP are related to the ordinary M/M/c model in the following manner.

- (a) If  $c = 1$ , we may regard each of them as the ordinary M/M/1 model, though service positions differ for every customer.
- (b) If we remove the position order constraint from the model assumptions, we may regard each of them as the ordinary M/M/c model.
- (c) In M/M,U/c(ordered)/PSP, if we add the constraint that Server  $j$  can serve only customers requesting service positions in the interval  $((j - 1)c^{-1}l, jc^{-1}l), j = 1, 2, \dots, c$ , then the model can be considered to be a composite of  $c$  ordinary M/M/1 models.

**Remark 4.2.** (a) In M/M,U/c(ordered)/PSP with  $c \geq 2$ , from (b) and (c) in Remark 4.1 above, we may expect that the following inequalities hold:

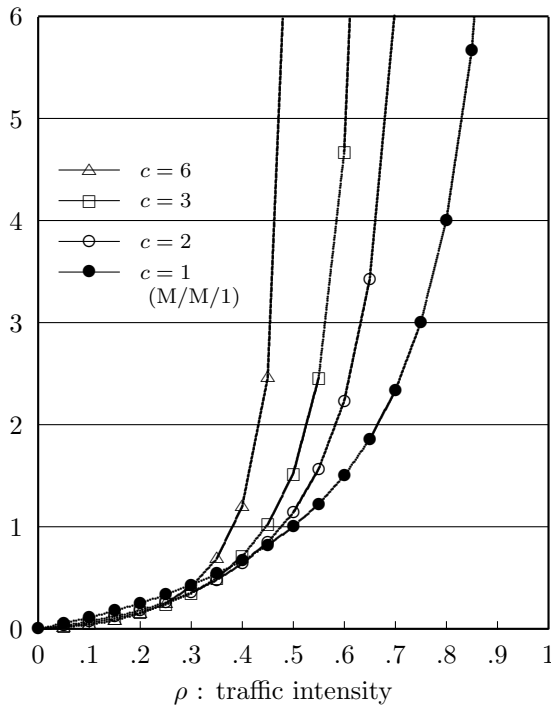
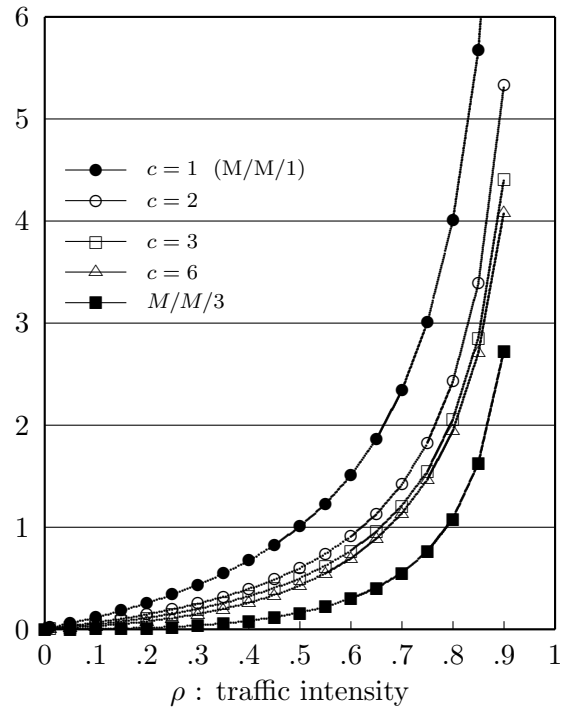
$$W_{q^*}^{\text{M/M}/c} < W_{q^*}^{\text{PSP}} < W_{q^*}^{\text{M/M}/1}. \tag{4.1}$$

- (b) In M/M,U/c(ordered)/FCFS with  $c \geq 2$ , some servers may be idle even if there are one or more customers in the queue, and so  $\rho_{\max}^{\text{FCFS}} < 1$ . The corresponding saturated model  $*/M,U/c(ordered)/FCFS$  should have throughput  $c\mu\rho_{\max}^{\text{FCFS}}$ . So, when  $c = 2$ , we know from (3.11) that  $\rho_{\max}^{\text{FCFS}} = 0.7795$ .

**Some simulation results**

The authors executed extensive simulation experiments for M/M,U/c(ordered)/FCFS and M/M,U/c(ordered)/PSP. Each experiment simulated more than 30 million customers. Confidence intervals are shown in Tables 1 and 2, but for simplicity, we omit the confidence intervals in the graphs. Hereafter in this section, we always set  $\mu = 1$ .

**The mean waiting time in M/M,U/c(ordered)/FCFS** The estimated value of  $W_{q^*}^{\text{FCFS}}$  is depicted in Figure 4 for  $c = 1, 2, 3$ , and 6. We see that, for  $\rho < 0.3$ ,  $W_{q^*}^{\text{FCFS}}$  for  $c \geq 2$  is smaller than  $W_{q^*}^{\text{M/M}/1}$  (which is equal to  $W_{q^*}^{\text{FCFS}}$  for  $c = 1$ ). However, for  $\rho > 0.45$ , it becomes larger than  $W_{q^*}^{\text{M/M}/1}$ . Further, the graph of  $W_{q^*}^{\text{FCFS}}$  grows more rapidly as  $c$  increases.

Figure 4:  $W_{q^*}^{\text{FCFS}}$  for  $\mu = 1$ Figure 5:  $W_{q^*}^{\text{PSP}}$  for  $\mu = 1$ 

**The stability threshold in M/M,U/c(ordered)/FCFS** The mean waiting time  $W_{q^*}^{\text{FCFS}}$  becomes infinity for  $\rho > \rho_{\max}^{\text{FCFS}}$ . The stability thresholds  $\rho_{\max}^{\text{FCFS}}$  for  $c \geq 3$  were estimated from simulation experiments on a saturated model \*/M,U/c(ordered)/FCFS, like the one analyzed in Section 3. Table 1 lists the estimated value of  $\rho_{\max}^{\text{FCFS}}$  and the standard errors ( $\pm$  half the width of the 95% confidence interval). We see that  $\rho_{\max}^{\text{FCFS}}$  decreases as  $c$  increases. It is likely that this is because as  $c$  increases, interactions among servers caused by the position order constraint become larger.

**The mean waiting time in M/M,U/c(ordered)/PSP** The estimated  $W_{q^*}^{\text{PSP}}$  is depicted in Figure 5 and its value and standard error are presented in Table 2 for some selected cases. We see that the model is stable for  $\rho < 1$  and that  $W_{q^*}^{\text{PSP}}$  decreases as  $c$  increases. For  $c = 2$ ,  $W_{q^*}^{\text{PSP}}$  is about a half  $W_{q^*}^{\text{M/M/1}}$ , and as we predicted in (4.1),  $W_{q^*}^{\text{PSP}}$  is always between  $W_{q^*}^{\text{M/M/1}}$  and  $W_{q^*}^{\text{M/M/c}}$ , though it is shown only for  $c = 3$  in Figure 5. For  $c \geq 3$ , the dependency of  $W_{q^*}^{\text{PSP}}$  on  $c$  looks small for large  $\rho$ .

**The mean waiting time conditioned on the service position requested** Figures 6 and 7 show the graphs of  $W_q^{\text{FCFS}}(a)$  and  $W_q^{\text{PSP}}(a)$ , respectively, estimated from simulations. Here,  $W_q^{\text{FCFS}}(a)$  and  $W_q^{\text{PSP}}(a)$  are estimated for  $a = \frac{(j-0.5)l}{100}$ ,  $j = 1, 2, \dots, 100$ , from simulation data of customers requesting service positions in the interval  $(\frac{j-1}{100}l, \frac{j}{100}l)$ . The number of customers used to estimate a single  $W_q^{\text{FCFS}}(a)$  or  $W_q^{\text{PSP}}(a)$  was about 300,000 (= 30 million/100), and the estimated values may not be accurate.

We see that  $W_q^{\text{PSP}}(a)$  is decreasing on  $c$ , irrespective of  $a$ , but the dependency of  $W_q^{\text{FCFS}}(a)$  on  $c$  is not so simple. All the graphs, except for the case  $c = 1$ , are U-shaped<sup>||</sup>, and  $W_q^{\text{FCFS}}(0)$  and  $W_q^{\text{FCFS}}(l)$  [ $W_q^{\text{PSP}}(0)$  and  $W_q^{\text{PSP}}(l)$ ] are larger than  $W_q^{\text{FCFS}}(\frac{1}{2}l)$  [ $W_q^{\text{PSP}}(\frac{1}{2}l)$ ]. This indicates that in a queueing model with service-position requests, the waiting times may not be

<sup>||</sup>Although it is not presented here, the graph of  $W_q^{\text{FCFS}}(a)$  for  $\rho$  near  $\rho_{\max}^{\text{FCFS}}$  is rather flat, but still U-shaped.

Table 1: Stability threshold  $\rho_{\max}^{\text{FCFS}}$  (estimated)

$c = 1$	$c = 2$	$c = 3$	$c = 6$
1.0000	0.7795	$0.6734 \pm 0.0003$	$0.5259 \pm 0.0003$

Table 2: Mean waiting time  $W_{q^*}^{\text{PSP}}$  (estimated,  $\mu = 1$ )

$\rho$	$c = 1$	$c = 2$	$c = 3$	$c = 6$
0.005	0.00503	$0.00269 \pm 0.00003$	$0.00179 \pm 0.00002$	$0.00092 \pm 0.00002$
0.1	0.1111	$0.06074 \pm 0.00015$	$0.04401 \pm 0.00012$	$0.02662 \pm 0.00010$
0.3	0.4286	$0.24503 \pm 0.00043$	$0.19520 \pm 0.00039$	$0.14501 \pm 0.00034$
0.5	1.0000	$0.5939 \pm 0.0012$	$0.4959 \pm 0.0010$	$0.4244 \pm 0.0010$
0.7	2.3333	$1.4117 \pm 0.0038$	$1.2055 \pm 0.0032$	$1.1311 \pm 0.0035$
0.9	9.0000	$5.3205 \pm 0.0367$	$4.4061 \pm 0.0289$	$4.0816 \pm 0.0469$

uniform over the positions requested. The nonuniformity occurs even in an FCFS model, where customers receive services in the order of their arrivals. However, the grade of nonuniformity is much more intense in PSP models. In this sense, the M/M,U/ $c$ (ordered)/PSP model with  $c \geq 2$  might disadvantage customers requesting service positions near the ends of the counter. However, we note that  $W_q^{\text{PSP}}(a)$  is always smaller than  $W_q^{\text{M/M/1}}$ .

One possible explanation for the U shape might be a similar one to the intuitive explanation for the V shape of the loss probability in the M/M,U/2(ordered)/loss given in the last paragraph of Section 2; however, the authors have not yet succeeded in giving a more persuasive explanation. Another interesting property is that the graphs of  $W_q^{\text{PSP}}(a)$  for  $c \geq 2$  start from and end at almost the same point. The authors do not have any proper explanation for it, too.

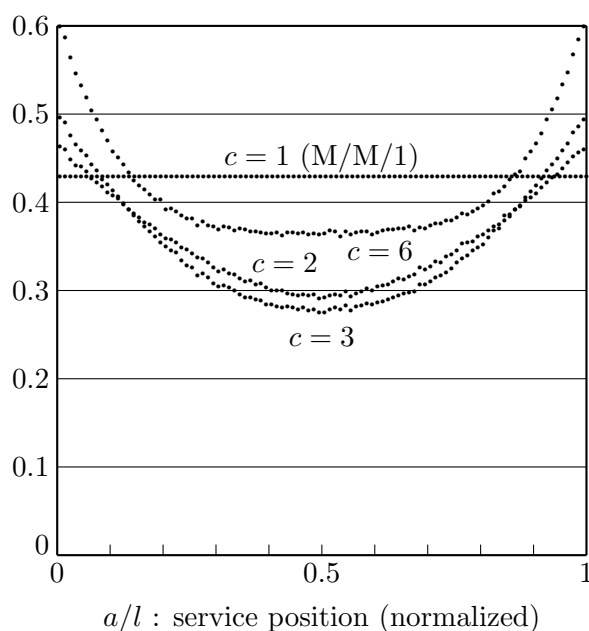
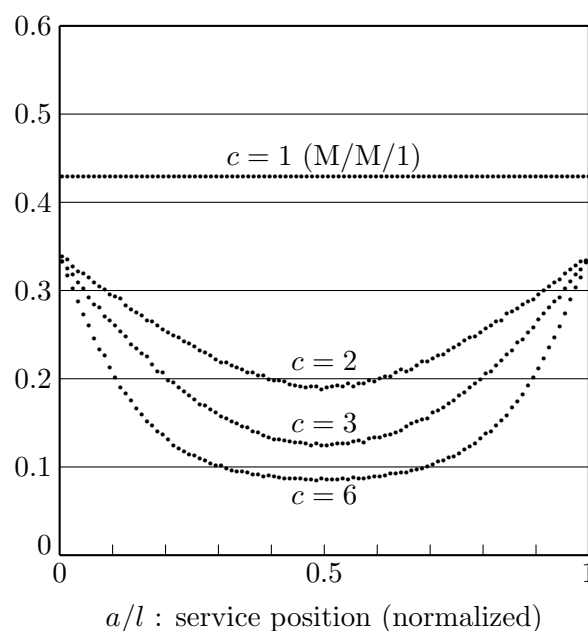
## 5. Application to a Movable Compact Shelving System in a University Library

The Ookayama Campus Library of the Tokyo Institute of Technology was built in February 2011. Several years before that, it was found that the old library building was not sufficiently earthquake resistant, and in 2007, the university decided to build a new building near the old one. The total floor space was set to be approximately 8,600 m<sup>2</sup>. The queueing model with service-position requests was devised during the planning of the storage system for the new library.

The library's collections included more than 640,000 volumes, and a large portion of them were bound periodicals in the field of science and technology. It was planned to store most of the collections in open-shelf stacks in the first- and second-level basements of the new building. To use the space efficiently, an electrical movable compact shelving system was considered.

### Movable compact shelving systems

In [3], p.559, movable compact shelving systems are explained as follows.

Figure 6:  $W_q^{\text{FCFS}}(a)$  for  $\rho = 0.3$  and  $\mu = 1$ Figure 7:  $W_q^{\text{PSP}}(a)$  for  $\rho = 0.3$  and  $\mu = 1$ 

“Mobile storage systems feature storage units — shelving, file cabinets, map cases, and other types — mounted on wheeled carriages that run on tracks in the floor. Rather than having to provide space for an aisle between each range (row) of storage units, a mobile system makes it possible to provide only one aisle for a number of ranges. The system consists of a number of modules (also referred to as blocks and pods). Each module consists of a fixed carriage at each end with movable carriages between the fixed ones. One may create an accessible aisle anywhere within a module by one of three methods: manually by pushing on a fixed handle on the end of the mobile range, mechanical-assist using a crank handle, or electrically, using push-button controls.”

The new library intended to store most of its vast collections of bound periodicals in a movable compact shelving system. It is sometimes thought that movable compact shelving systems reduce the ease of access to collections; it is true that only one face of the multiple ranges in a module is accessible at any one time, and a user must open the aisle before accessing the desired face of the range. However, typically, bound periodicals are retrieved without browsing, and the desired aisle for a user can be automatically created by placing a request at a terminal of the controlling computer. In this sense, bound periodicals are a good fit for movable compact shelving systems.

### Some plans

Figures 8 to 10 show three of the plans that were examined for the movable compact shelving system in a room in the second-level basement of the new library. The room is about  $46.4 \text{ m} \times 9.6 \text{ m}$ . Each plan has a number of double-faced *movable* ranges (carriages), several single-faced *fixed (stationary)* ranges, and several double-faced *fixed* ranges. We will consider a double-faced fixed range to be a union of two single-faced fixed ranges that belong to different modules or are facing a passage.

Plan 1 in Figure 8 consists of 11 modules\*\*. Each module has a single-faced range at

\*\*Strictly speaking, Plan 1 consists of 11 modules *and* four single-faced fixed ranges next to Modules 5, 6, 9, and 10, each of which is one side of a double-faced fixed range and faces a passage. In evaluation of plans, we consider these four single-faced fixed ranges can be reached by users without waiting.



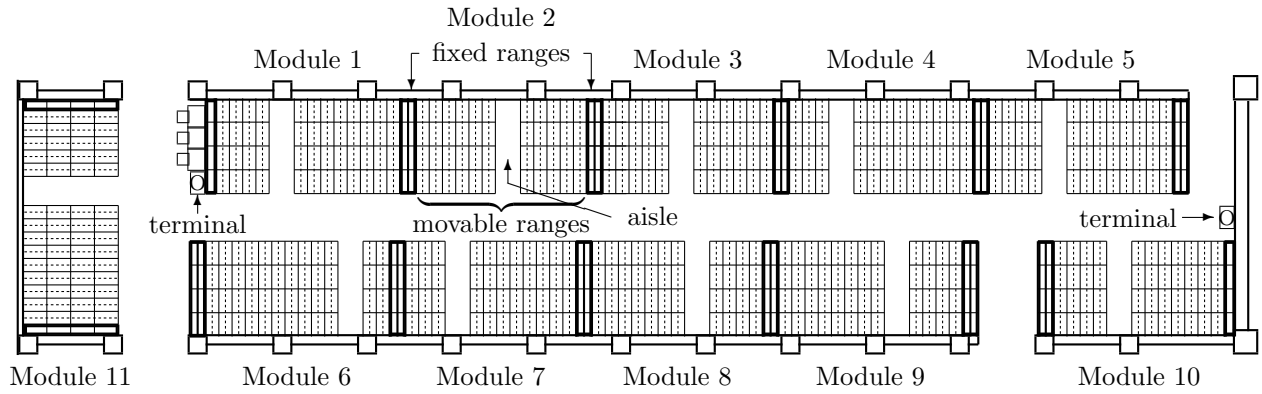
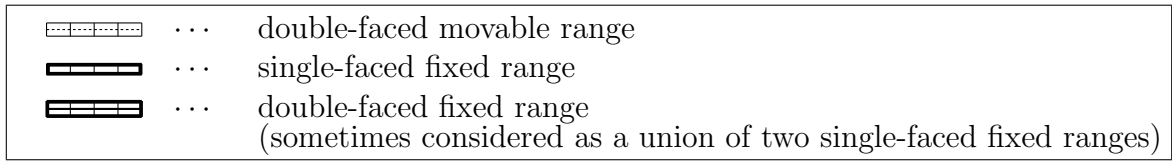


Figure 8: Plan 1: Each module has a single aisle

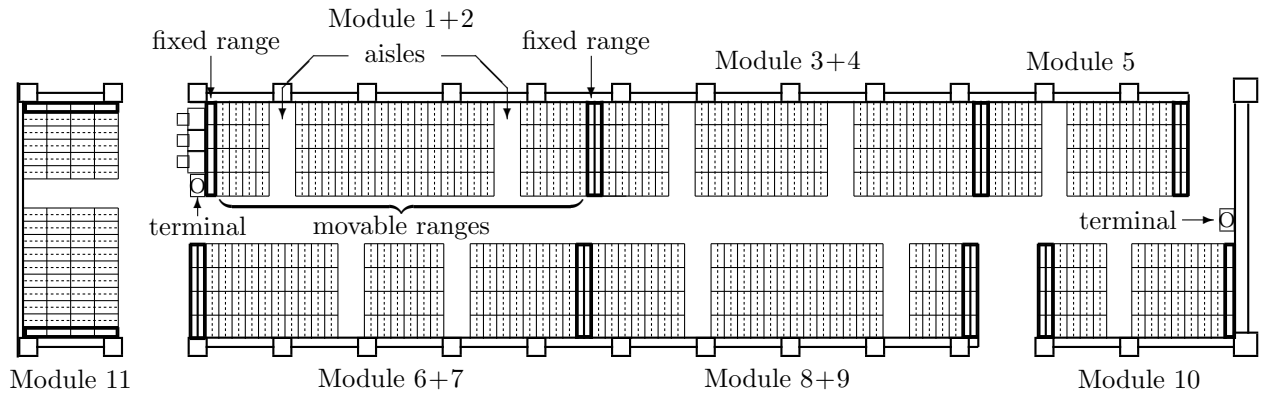


Figure 9: Plan 2: Some modules have two aisles

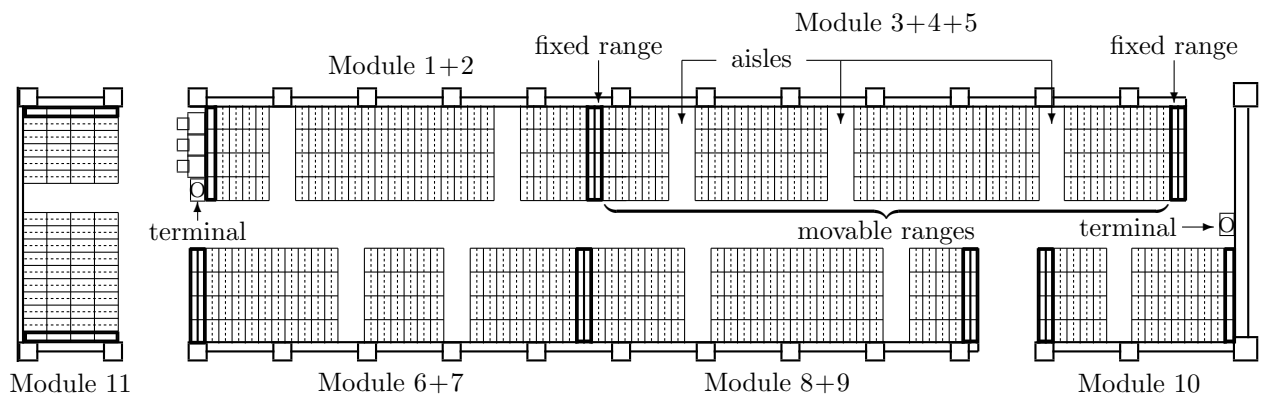


Figure 10: Plan 3: One module has three aisles and others have one or two aisles

each end and some double-faced movable ranges between them. Each module also has a space for an aisle (90 cm). The number of positions where an aisle can be created is equal to the number of movable ranges plus one. For example, Module 2 has 11 movable ranges and 12 positions where an aisle can be created. Each face of a range consists of four sections, each of which is 90 cm wide and has six shelves. Hence, Module 2, for example, can store about 17,000 ( $\doteq 6 \times 4 \times 2 \times 12 \times 30$ ) volumes. The room can store a total of about 200,000 volumes in Plan 1.

Usually in movable compact shelving systems [3], as described above, each module has a single aisle, as in Plan 1. However, from the viewpoint of queueing theory, we may consider plans having modules with two or more aisles, such as in Plans 2 and 3 in Figures 9 and 10, respectively. Plan 2 has four modules with two aisles, and Plan 3 has one module with three aisles and three modules with two aisles. In these plans, a module with two [three] aisles is derived by merging two [three] modules, each with a single aisle. Hence, the accommodation capacity and the ratio of the number of aisles to the number of ranges are the same in all three plans.

### Simulation model

In order to evaluate these plans, the authors did extensive simulation experiments using a queueing model with service-position requests. In the library application, a user of a movable compact shelving system wants a book in a module and thus requests that an aisle be created in the range where the book is stored. If we regard the module as a service station and the requested aisle position as a service position, the stochastic behavior of customers in the module can be modeled as a queueing model with service-position requests, although in this case, the service-position distribution is discrete.

For the simulation, we chose the  $M/M, U_d/c(\text{ordered})/PSP$  model, where  $U_d$  stands for a discrete uniform distribution that takes  $d$  different values with equal probability. A user who wants a book stored in a module can create an aisle in one of two ways: (i) give an instruction to the computer from a terminal, or (ii) push a button on the side panel of the range. Operation (ii) is reflected in the model as the priority of a customer who finds an available server over customers who have no available servers; we note that this is the PSP discipline, whereas the FCFS discipline cannot handle this operation.

In the model  $M/M, U_d/c(\text{ordered})/PSP$ , the service-position distribution is discrete, so it is necessary to add a rule for the case in which multiple customers request the same position. We adopted a simple rule specific to this application:

- Any single server can serve any number of customers simultaneously if these customers are requesting the same position.

Table 3: Modules in Plans 1 to 3

$d$ : the number of possible positions where aisles can be created  
 $c$ : the number of aisles created

Plan 1			Plan 2			Plan 3		
Modules	$d$	$c$	Modules	$d$	$c$	Modules	$d$	$c$
2, 3, 7, 8, 10	12	1	10	12	1	10	12	1
1, 4, 5, 6, 9	13	1	5	13	1	11	15	1
11	15	1	11	15	1	1+2, 6+7, 8+9	25	2
			1+2, 3+4, 6+7, 8+9	25	2	3+4+5	38	3

Table 4: The probability of wait  $Q(d,c)$  in M/M, $U_d/c$ (ordered)/PSP (estimated)

$\lambda_{\text{room}}$	$c = 1$						$c = 2$		$c = 3$	
	$d = 12$		$d = 13$		$d = 15$		$d = 25$		$d = 38$	
	$\lambda$	$Q(12,1)$	$\lambda$	$Q(13,1)$	$\lambda$	$Q(15,1)$	$\lambda$	$Q(25,2)$	$\lambda$	$Q(38,3)$
0.1	0.0085	0.0085	0.0092	0.0091	0.011	0.011	0.018	0.0044	0.027	0.0030
1	0.085	0.084	0.092	0.91	0.11	0.11	0.18	0.050	0.27	0.038
2	0.17	0.17	0.18	0.18	0.21	0.21	0.35	0.11	0.54	0.091
4	0.34	0.34	0.37	0.36	0.42	0.42	0.70	0.27	1.1	0.24
10	0.85	0.78	0.92	0.83	1.1	0.92	1.8	0.80	2.7	0.81

Table 5: The mean waiting time  $W_q(d,c)$  ( $\mu = 1$ ) in M/M, $U_d/c$ (ordered)/PSP (estimated)

$\lambda_{\text{room}}$	$c = 1$						$c = 2$		$c = 3$	
	$d = 12$		$d = 13$		$d = 15$		$d = 25$		$d = 38$	
	$\lambda$	$W_q(12,1)$	$\lambda$	$W_q(13,1)$	$\lambda$	$W_q(15,1)$	$\lambda$	$W_q(25,2)$	$\lambda$	$W_q(38,3)$
0.1	0.0085	0.0085	0.0092	0.0092	0.011	0.011	0.018	0.0044	0.027	0.0030
1	0.085	0.092	0.092	0.10	0.11	0.12	0.18	0.050	0.27	0.036
2	0.17	0.20	0.18	0.22	0.21	0.26	0.35	0.11	0.54	0.086
4	0.34	0.49	0.37	0.55	0.42	0.69	0.70	0.30	1.1	0.25
10	0.85	2.8	0.92	3.6	1.1	6.0	1.8	2.3	2.7	2.2

As a special case of this rule, if an arriving customer finds that the position they wish to request is already being served, then the arriving customer is immediately served.

We let  $\lambda$  be the arrival rate of the Poisson arrival process and  $\mu$  be the service rate for exponentially distributed service times. We assume that the arrival process, requested service positions, and service times are stochastically independent.

**Simulation results**

Here, we briefly summarize simulation results for Plans 1 to 3. Plan 1 consists of eleven modules, Plan 2 consists of seven modules, and Plan 3 consists of six modules (see footnote 7). These modules are listed in Table 3 together with  $d$ , the number of possible positions where aisles can be created, and  $c$ , the number of aisles created.

We let  $\lambda_{\text{room}}$  be the rate of arrival to the room, and we assume that the arrival rate to a module with  $d$  possible positions for aisles is given by

$$\lambda = \frac{d}{D_{\text{room}}} \lambda_{\text{room}} = \frac{d}{142} \lambda_{\text{room}}. \tag{5.1}$$

Here,  $D_{\text{room}} = 142$  is the total number of ranges in each of Plans 1 to 3, with each double-faced range counted as a unit, and we implicitly assume the arrivals are distributed uniformly over the ranges.

Tables 4 and 5 show some simulation results for the probability of wait  $Q(d,c)$  and the mean waiting time  $W_q(d,c)$  in M/M, $U_d/c$ (ordered)/PSP, respectively, for various values of  $\lambda_{\text{room}}$ . The combinations of  $d$  and  $c$  are chosen so that they cover all the modules listed in Table 3. The estimated values given in these tables are based on experiments that each simulated more than 30 million customers. The results are sufficiently precise for evaluating plans with two significant digits; for simplicity, we omit the confidence intervals.

Table 6: The probability of wait  $Q^{\text{Plan } i}$  in Plans 1 to 3 (estimated)

$\lambda_{\text{room}}$	$Q^{\text{Plan } 1}$	$Q^{\text{Plan } 2}$	$Q^{\text{Plan } 3}$
0.1	0.0089	0.0058	0.0050
1	0.089	0.062	0.055
2	0.18	0.13	0.12
4	0.35	0.29	0.28
10	0.81	0.80	0.80

Table 7: The mean waiting time  $W_q^{\text{Plan } i}$  in Plans 1 to 3 (estimated,  $\mu = 1$ )

$\lambda_{\text{room}}$	$W_q^{\text{Plan } 1}$	$W_q^{\text{Plan } 2}$	$W_q^{\text{Plan } 3}$
0.1	0.0089	0.0058	0.0050
1	0.097	0.064	0.056
2	0.21	0.14	0.13
4	0.53	0.38	0.34
10	3.4	2.8	2.7

Using the results in Tables 4 and 5, we can calculate the aggregated (i.e., weighted averaged) probability of wait and the aggregated mean waiting time for Plans 1 to 3. For Plan  $i$ , let  $Q^{\text{Plan } i}$  be the probability of wait and let  $W^{\text{Plan } i}$  be the mean waiting time. Further, for the  $k$ -th module in Plan  $i$ , we let  $d_k^{\text{Plan } i}$  be the number of possible positions for aisles,  $c_k^{\text{Plan } i}$  be the number of aisles, and  $\lambda_k^{\text{Plan } i}$  be the arrival rate. Note that, on average,  $\lambda_k^{\text{Plan } i} / \lambda_{\text{room}}$  of the customers who arrive at the room go to the  $k$ -th module. Then,  $Q^{\text{Plan } i}$  and  $W^{\text{Plan } i}$  are given by

$$Q^{\text{Plan } i} = \sum_k \frac{\lambda_k^{\text{Plan } i}}{\lambda_{\text{room}}} Q(d_k^{\text{Plan } i}, c_k^{\text{Plan } i}) \quad \text{and} \quad W_q^{\text{Plan } i} = \sum_k \frac{\lambda_k^{\text{Plan } i}}{\lambda_{\text{room}}} W_q(d_k^{\text{Plan } i}, c_k^{\text{Plan } i}). \quad (5.2)$$

From Tables 3 to 5, we can easily estimate the values of  $Q^{\text{Plan } i}$  and  $W_q^{\text{Plan } i}$ . The results are listed in Tables 6 and 7. We can see that, when the traffic is light,  $W_q^{\text{Plan } 2}$  is about 70% of  $W_q^{\text{Plan } 1}$ , and  $W_q^{\text{Plan } 3}$  is about 60% of  $W_q^{\text{Plan } 1}$ .

## Discussion

In the actual design process, after many discussions, it was decided to adopt Plan 2 for the movable compact shelving system, and a system based on this plan was installed. We list here some of the points considered in these discussions.

1. From the point of view of the mean waiting time, it is clearly effective to use larger modules with multiple aisles. The performance of Plan 2 was considerably better than that of Plan 1, but Plan 3 was only a little better than Plan 2.
2. It is difficult to quantitatively estimate the traffic intensity  $\lambda_{\text{room}}$ , but it was anticipated that  $\lambda_{\text{room}}$  would be so small that waiting times would not cause serious problems. This is because the collections stored in the system are mainly bound periodicals.
3. As in the case of M/M,U/c(ordered)/PSP, which was discussed in the preceding section, in M/M,U<sub>d</sub>/c(ordered)/PSP, the mean waiting time conditioned on the service position requested (the one corresponding to  $W_q^{\text{PSP}}(a)$ ) is U-shaped, and customers requesting positions near the ends of the counter are relatively disadvantaged. However, the conditional mean waiting time is always less than  $W_q^{\text{M/M/1}}$ , the mean waiting time in the single-server case, and thus it is considered to be acceptable.
4. If a module has two aisles, there is an additional advantage. By using the space for two aisles, we can create a wider (180 cm) aisle in the module. This is very convenient for the library staff, especially when they shift sections of books. However, in order to use this function, each carriage (range) must have longer overhead wiring ducts, which store the electric power and signal cables and connect adjacent carriages. If the widest

possible aisle is increased, the length of the duct must be increased, and too wide of an aisle (e.g., 270 cm, combining three aisles) is not practical.

5. In addition to rails, frames, and shelves, each module consists of many parts, including motors, gears, wheels, electric control circuits, microprocessors, and sensors. For example, even in the smallest module in Plan 1, there are more than 600 sensors. If one of these parts fails, the entire module will be inaccessible to users until the failed part can be fixed or replaced. So, from the point of view of reliability, smaller modules are preferable. For this reason, a three-aisle module was considered too large.

## 6. Concluding Remarks

In this paper, we proposed a new type of queueing model for customers requesting service positions at a counter. The model was first considered in the process of a performance evaluation for a movable compact shelving system in a university library, as discussed in Section 5. The characteristic of the model is in the position order constraint on servers. Two very simple models are analyzed in Sections 2 and 3. The analyses show some interesting characteristics of the models, including the V-shaped form of the loss probability  $P(a)$  conditioned on the service position  $a$  in the M/M,U/2(ordered)/loss model. However, until now, we have only been able to analyze these two models.

To understand the behaviors of more sophisticated models, simulation experiments were performed for the models M/M,U/ $c$ (ordered)/FCFS and M/M,U/ $c$ (ordered)/PSP, and the results are reported in Section 4. Through the simulations, we discovered that  $\rho_{\max}^{\text{FCFS}}$ , the stability threshold in the former model, is less than 1, and  $\rho_{\max}^{\text{PSP}}$ , the stability threshold in the latter model, is equal to 1. The mean waiting time  $W_{q^*}^{\text{PSP}}$  in the latter model is expected to be between those of the ordinary M/M/1 and M/M/ $c$  models. The mean waiting times  $W_q^{\text{FCFS}}(a)$  and  $W_q^{\text{PSP}}(a)$  conditioned on the service position  $a$  in these models are U-shaped. Further, as a function of  $c$ , both  $\rho_{\max}^{\text{FCFS}}$  and  $W_q^{\text{PSP}}(a)$  are decreasing.

These properties are features of the queueing models with service-position requests, but they have not yet been fully explained. In order to better understand their properties, we need further study of these models as well as other queueing models with service-position requests.

## Appendix A. Generator for the Model M/M,U/2(ordered)/loss and Proofs of Theorem 2.1 and Corollaries 2.1 to 2.3

### The infinitesimal generator of the Markov process $X(t)$ describing the stochastic behavior of M/M,U/2(ordered)/loss

For the loss model M/M,U/2(ordered)/loss, the state space  $S$  is defined in (2.1). It consists of a singleton, two intervals in  $\mathbb{R}^1$  and a Borel set in  $\mathbb{R}^2$ . So, we can introduce a natural topology on  $S$ , and let  $B(S)$  denote the family of Borel subsets of  $S$ . The infinitesimal generator of the Markov process  $\{X(t)\}$  on  $S$  is defined by

$$G(\xi, A) = \lim_{t \downarrow 0} P(X(t) \in A | X(0) = \xi), \quad \xi \in S, A \in B(S), \quad (\text{A.1})$$

if it exists. To give a concrete expression, we introduce some notations. For  $\zeta \in S$ , we let  $\{\zeta\}$  denote a singleton in  $B(S)$ . For  $\xi, \xi_1, \xi_2, \eta, \eta_1, \eta_2 \in (0, l)$  such that  $\xi_1 < \xi_2$  and  $\eta_1 < \eta_2$ ,

we let

$$\bar{S}_1(\xi_1, \xi_2) = \bigcup_{\xi_1 < x < \xi_2} S_1(x), \quad \bar{S}_2(\eta_1, \eta_2) = \bigcup_{\eta_1 < y < \eta_2} S_2(y), \quad (\text{A.2})$$

$$\bar{S}_{31}(\xi; \eta_1, \eta_2) = \bigcup_{\eta_1 < y < \eta_2} S_3(\xi, y), \quad \bar{S}_{32}(\xi_1, \xi_2; \eta) = \bigcup_{\xi_1 < x < \xi_2} S_3(x, \eta), \quad (\text{A.3})$$

$$\text{and } \bar{S}_3(0, l) = \bigcup_{0 < x < y < l} S_3(x, y). \quad (\text{A.4})$$

Then, the generator  $G(\xi, A)$  for a typical pair  $(\xi, A)$  is written as (A.5) to (A.20) below (see Service discipline in Section 2 and explanations just after the balance equations (2.2) to (2.6)). Here  $x$  and  $y$  are positive numbers such that  $0 < x < y < l$ .

$$G(S_0; \{S_0\}) = -\lambda \quad (\text{A.5})$$

$$G(S_0; \bar{S}_1(\xi_1, \xi_2)) = l^{-1}\lambda(\xi_2 - \xi_1), \quad 0 < \xi_1 < \xi_2 < \frac{1}{2}l \quad (\text{A.6})$$

$$G(S_0; \bar{S}_2(\eta_1, \eta_2)) = l^{-1}\lambda(\eta_2 - \eta_1), \quad \frac{1}{2}l < \eta_1 < \eta_2 < l \quad (\text{A.7})$$

$$G(S_0, A) = 0, \quad A \subset \bar{S}_1(\frac{1}{2}l, l) \cup \bar{S}_2(0, \frac{1}{2}l) \cup \bar{S}_3(0, l) \quad (\text{A.8})$$

$$G(S_1(x); \{S_0\}) = \mu \quad (\text{A.9})$$

$$G(S_1(x); \{S_1(x)\}) = -l^{-1}\lambda(l - x) - \mu \quad (\text{A.10})$$

$$G(S_1(x); \bar{S}_{31}(x; \eta_1, \eta_2)) = l^{-1}\lambda(\eta_2 - \eta_1), \quad x < \eta_1 < \eta_2 < l \quad (\text{A.11})$$

$$G(S_1(x); A) = 0, \quad A \subset (\bar{S}_1(0, l) - \{S_1(x)\}) \cup \bar{S}_2(0, l) \cup (\bar{S}_3(0, l) - \bar{S}_{31}(x; x, l)) \quad (\text{A.12})$$

$$G(S_2(y); \{S_0\}) = \mu \quad (\text{A.13})$$

$$G(S_2(y); \{\bar{S}_2(y)\}) = -l^{-1}\lambda y - \mu \quad (\text{A.14})$$

$$G(S_2(y); \bar{S}_{32}(\xi_1, \xi_2; y)) = l^{-1}\lambda(\xi_2 - \xi_1), \quad 0 < \xi_1 < \xi_2 < y \quad (\text{A.15})$$

$$G(S_2(y); A) = 0, \quad A \subset \bar{S}_1(0, l) \cup (\bar{S}_2(0, l) - \{S_2(y)\}) \cup (\bar{S}_3(0, l) - \bar{S}_{32}(y; 0, y)) \quad (\text{A.16})$$

$$G(S_3(x, y); \{S_3(x, y)\}) = -2\mu \quad (\text{A.17})$$

$$G(S_3(x, y); \{S_1(x)\}) = \mu \quad (\text{A.18})$$

$$G(S_3(x, y); \{S_2(y)\}) = \mu \quad (\text{A.19})$$

$$G(S_3(x, y); A) = 0, \quad A \subset S - \{S_3(x, y)\} - \{S_1(x)\} - \{S_2(y)\}. \quad (\text{A.20})$$

For a general pair  $(\xi, A)$ ,  $\xi \in S$ ,  $A \in B(S)$ , the expression for  $G(\xi, A)$  is easily induced from the above cases.

### Proof of Theorem 2.1

A direct integration of  $f_1(x)$  in (2.8) leads us to

$$\begin{aligned} \int_0^l f_1(x) dx &= \int_0^{l/2} f_1(x) dx + \int_{l/2}^l f_1(x) dx \\ &= \frac{1}{\rho} C_1 \left[ \log\{1 + \rho\} + \frac{1}{2} C_2 \rho \right] - \frac{1}{\rho} C_1 \left[ \{1 + \rho\} \log\{1 + \rho\} - \frac{1}{2} C_2 \rho \right] \\ &= C_1 \left[ C_2 - \log\{1 + \rho\} \right] = \rho f_0. \end{aligned} \quad (\text{A.21})$$

Here we use the expressions for  $C_1$  and  $C_2$  in (2.9) to get the last expression. From (2.10),  $\int_0^l f_2(y) dy$  takes the same value as (A.21). Then we have

$$\int_0^l f_1(x) dy + \int_0^l f_2(y) dy = 2\rho f_0, \tag{A.22}$$

and this proves that the solution (2.7) through (2.10) satisfies the equation (2.2).

To show the solution satisfies the equation (2.3), we notice that from (2.11) and (2.10)

$$\begin{aligned} \int_x^l f_3(x, y) dy &= l^{-1}\rho \int_x^l f_1(x) dy + l^{-1}\rho \int_x^l f_2(y) dy \\ &= l^{-1}\rho(l-x)f_1(x) + l^{-1}\rho \int_0^{l-x} f_1(y) dy. \end{aligned} \tag{A.23}$$

For the calculation of the last integral, first we consider the case  $\frac{1}{2}l < x < l$ . A direct integration of  $f_1(x)$  in (2.8) leads us to

$$\begin{aligned} l^{-1}\rho \int_0^{l-x} f_1(y) dy &= l^{-2}C_1 \left[ -\{l + \rho(l-x)\} \log \frac{l + \rho x}{l + \rho(l-x)} + l \log \{1 + \rho\} + C_2 \rho(l-x) \right] \\ &= -l^{-2}C_1 \{l + \rho(l-x)\} \left[ \frac{(2 + \rho)l}{l + \rho(l-x)} + \log \frac{l + \rho x}{l + \rho(l-x)} - C_2 \right] \\ &\qquad\qquad\qquad + l^{-1}C_1 [2 + \rho + \log\{1 + \rho\} - C_2] \\ &= l^{-1}\{l + \rho(l-x)\}f_1(x) \qquad\qquad\qquad \text{(from (2.9)).} \end{aligned} \tag{A.24}$$

Hence (A.23) is written as

$$\int_x^l f_3(x, y) dy = l^{-1}\{l + 2\rho(l-x)\}f_1(x). \tag{A.25}$$

Noticing that  $h_{l/2}(x) = 0$ , this proves that the solution satisfies the equation (2.3) when  $\frac{1}{2}l < x < l$ . Next we consider the case  $0 < x < \frac{1}{2}l$ . The last term of (A.23) is calculated as

$$\begin{aligned} l^{-1}\rho \int_0^{l-x} f_1(y) dy &= l^{-1}\rho \int_0^{l/2} f_1(y) dy + l^{-1}\rho \int_{l/2}^{l-x} f_1(y) dy \\ &= l^{-1}C_1 \log\{1 + \rho\} + \frac{1}{2}l^{-1}C_1 C_2 \rho \\ &\qquad\qquad\qquad + l^{-2}C_1 \{l + \rho(l-x)\} \log \frac{l + \rho x}{l + \rho(l-x)} + l^{-2}C_1 C_2 \rho (\frac{1}{2}l - x) \\ &= l^{-2}C_1 \{l + \rho(l-x)\} \left[ \frac{(2 + \rho)l}{l + \rho(l-x)} + \log \frac{l + \rho x}{l + \rho(l-x)} + C_2 \right] \\ &\qquad\qquad\qquad - l^{-1}C_1 [2 + \rho - \log\{1 + \rho\} + C_2] \\ &= l^{-1}\{l + \rho(l-x)\}f_1(x) - 2l^{-1}\rho f_0 \qquad\qquad\qquad \text{(from (2.9)).} \end{aligned} \tag{A.26}$$

Substituting this to (A.23), we easily see that the solution satisfies (2.3) when  $0 < x < \frac{1}{2}l$ , since  $h_{l/2}(x) = 1$  in this case.

The equation (2.4) is easily derived from (2.3) by using (2.10), and hence it is satisfied by the solution. The equation (2.5) is essentially the same as (2.11), and it is satisfied by the solution.

Finally, in order to check the solution satisfies the equation (2.6), we calculate the integral of  $f_3(x, y)$  as follows. From (2.11) and (2.10) we have

$$\begin{aligned} \iint_{0 < x < y < l} f_3(x, y) dx dy &= \rho l^{-1} \int_0^l \int_x^l f_1(x) dy dx + \rho l^{-1} \int_0^l \int_0^y f_2(y) dx dy \\ &= \rho l^{-1} \int_0^l (l-x) f_1(x) dx + \rho l^{-1} \int_0^l y f_2(y) dy = 2\rho l^{-1} \int_0^l (l-x) f_1(x) dx. \end{aligned} \quad (\text{A.27})$$

Further,

$$\begin{aligned} 2\rho l^{-1} \int_0^l (l-x) f_1(x) dx &= 2\rho l^{-1} \int_0^{l/2} (l-x) f_1(x) dx + 2\rho l^{-1} \int_{l/2}^l (l-x) f_1(x) dx \\ &= \rho^{-1} l^{-2} C_1 \left[ \{3l^2 + 2\rho l^2 + \rho^2(l-x)^2\} \log\{l + \rho(l-x)\} \right. \\ &\quad \left. + \{(1+\rho)^2 l^2 - \rho^2(l-x)^2\} \log\{l + \rho x\} + \rho\{2+\rho\} l x - C_2 \rho^2(l-x)^2 \right]_{l/2}^{l/2} \\ &\quad - \rho^{-1} l^{-2} C_1 \left[ \{3l^2 + 2\rho l^2 + \rho^2(l-x)^2\} \log\{l + \rho(l-x)\} \right. \\ &\quad \left. + \{(1+\rho)^2 l^2 - \rho^2(l-x)^2\} \log\{l + \rho x\} + \rho\{2+\rho\} l x + C_2 \rho^2(l-x)^2 \right]_{l/2}^l \\ &= \{2+\rho\} \left\{ 2 \log \left\{ 1 + \frac{1}{2} \rho \right\} - \log\{1+\rho\} \right\} f_0 + \rho^2 f_0 \quad (\text{from (2.9)}). \end{aligned} \quad (\text{A.28})$$

Substituting (A.22) and (A.28) into (2.6) and using the expression (2.7) for  $f_0$ , we see that the solution satisfies the equation (2.6).  $\square$

### Proof of Corollary 2.1

For  $P_*$ , we calculate (2.12). From (A.27), we easily see that

$$l^{-1} \int_0^l x f_1(x) dx = \int_0^l f_1(x) dx - \frac{1}{2\rho} \iint_{0 < x < y < l} f_3(x, y) dx dy. \quad (\text{A.29})$$

The integral  $l^{-1} \int_0^l (l-y) f_2(y) dy$  has the same expression as above. From (A.21) we know that  $\int_0^l f_1(x) dx = \rho f_0$ , and from (2.6) we know that  $\iint_{0 < x < y < l} f_3(x, y) dx dy = 1 - f_0 - 2\rho f_0$ . Hence we have

$$\begin{aligned} P_* &= l^{-1} \int_0^l x f_1(x) dx + l^{-1} \int_0^l (l-y) f_2(y) dy + \iint_{0 < x < y < l} f_3(x, y) dy \\ &= 2 \int_0^1 f_1(x) dx + \left( 1 - 2 \cdot \frac{1}{2\rho} \right) \iint_{0 < x < y < 1} f_3(x, y) dx dy \\ &= 2\rho f_0 + \left( 1 - \frac{1}{\rho} \right) (1 - f_0 - 2\rho f_0) = 1 - \frac{1}{\rho} + \left( 1 + \frac{1}{\rho} \right) f_0. \end{aligned} \quad (\text{A.30})$$

This proves the first expression of (2.14). The equivalence of the second expression to the first is easily confirmed by using (2.7).



For  $P(a)$ , we calculate (2.13). Assume that  $0 < a \leq \frac{1}{2}l$ . A direct integration leads us to

$$\int_a^l f_1(x) dx = l^{-1} \left[ \rho(l-a) - \frac{l+\rho a}{2+\rho} \log \frac{\{l+\rho a\}\{1+\rho\}}{l+\rho(l-a)} \right] f_0, \quad \text{and} \quad (\text{A.31})$$

$$\int_0^a f_2(y) dy = l^{-1} \left[ \rho a - \frac{l+\rho(l-a)}{2+\rho} \log \frac{\{l+\rho a\}\{1+\rho\}}{l+\rho(l-a)} \right] f_0. \quad (\text{A.32})$$

Then, using  $\iint_{0 < x < y < l} f_3(x, y) dx dy = 1 - f_0 - 2\rho f_0$ , from (2.13) we have (2.15) for  $0 < a \leq \frac{1}{2}l$ . The expression for  $\frac{1}{2}l < a < l$  can be obtained in a similar manner.  $\square$

**Proof of Corollary 2.2**

Since  $f_0 = 1 - 2\rho + \frac{5}{2}\rho^2 + o(\rho^2)$ , (2.16) and (2.17) are easily derived from (2.14) and (2.15), respectively.  $\square$

**Proof of Corollary 2.3**

From (2.14), we have

$$P_*^{M/M/1/1} - P_* = \frac{(2+\rho)f_0}{\rho(1+\rho)} \log \frac{(1+\frac{1}{2}\rho)^2}{1+\rho} > 0. \quad (\text{A.33})$$

Using the inequality  $\log(1+x) \leq x$  for  $x > -1$ , we also have

$$\begin{aligned} P_* - P_*^{M/M/2/2} &= \frac{(1+\rho)f_0}{\rho(1+2\rho+2\rho^2)} \left\{ \rho^2 - (2+\rho) \log \frac{(1+\frac{1}{2}\rho)^2}{1+\rho} \right\} \\ &\geq \frac{(1+\rho)f_0}{\rho(1+2\rho+2\rho^2)} \cdot \frac{\rho^2(2+3\rho)}{4(1+\rho)} > 0. \end{aligned} \quad (\text{A.34})$$

These inequalities prove (2.18).  $\square$

**Appendix B. Proofs of Theorem 3.1 and Corollary 3.1**

**Proof of Theorem 3.1**

For brevity of expressions, we put  $G = \frac{54}{3+4\log 2}$ . First we decompose the fraction in (3.7) into partial fractions.

$$G^{-1}g(x) = \frac{l^2(l-x)}{(2l-x)^2(l+x)^3} = \frac{2}{9} \frac{l}{(l+x)^3} + \frac{1}{27} \frac{1}{(l+x)^2} - \frac{1}{27} \frac{1}{(2l-x)^2}. \quad (\text{B.1})$$

Then the indefinite integral of  $G^{-1}g(x)$  is given by

$$\int G^{-1}g(x) dx = -\frac{1}{9} \frac{l}{(l+x)^2} - \frac{1}{27} \frac{1}{l+x} - \frac{1}{27} \frac{1}{2l-x} + C. \quad (\text{B.2})$$

Using (3.8) and (3.10), the right-hand side (rhs) of (3.3) is calculated as follows:

$$\begin{aligned}
\text{rhs of (3.3)} &= \mu l^{-1} \int_x^l g(x') dx' + \frac{1}{2} \mu l^{-1} \int_x^l \{g(x) + g(l-y)\} dy \\
&= \mu l^{-1} \int_x^l \{g(y) + \frac{1}{2} g(l-y)\} dy + \frac{1}{2} \mu l^{-1} (l-x) g(x) \\
&= G \mu l^{-1} \left[ -\frac{1}{9} \frac{l}{(l+y)^2} - \frac{1}{27} \frac{1}{l+y} - \frac{1}{27} \frac{1}{2l-y} \right. \\
&\quad \left. + \frac{1}{18} \frac{l}{(2l-y)^2} + \frac{1}{54} \frac{1}{2l-y} + \frac{1}{54} \frac{1}{l+y} \right]_x^l + \frac{1}{2} \mu l^{-1} (l-x) g(x) \\
&= G \mu l^{-1} \left\{ \frac{1}{9} \frac{l}{(l+x)^2} + \frac{1}{54} \frac{1}{l+x} + \frac{1}{54} \frac{1}{2l-x} - \frac{1}{18} \frac{l}{(2l-x)^2} \right\} + \frac{1}{2} \mu l^{-1} (l-x) g(x) \\
&= G \mu l^{-1} \left\{ \frac{1}{9} \frac{l}{(l+x)^2} + \frac{1}{54} \frac{1}{l+x} - \frac{1}{54} \frac{l+x}{(2l-x)^2} \right\} + \frac{1}{2} \mu l^{-1} (l-x) g(x) \\
&= \frac{1}{2} \mu l^{-1} (l+x) g(x) + \frac{1}{2} \mu l^{-1} (l-x) g(x) = \mu g(x) = \mu f_1(x, z). \tag{B.3}
\end{aligned}$$

This proves that the solution (3.8) through (3.10) satisfies the equation (3.3). Further it is easily seen that the solution also satisfies the equations (3.4) and (3.5), as stated in footnote 5.

To see the solution satisfies the equation (3.6), we calculate the integrals as follows.

$$\begin{aligned}
&\iint_{0 < z < x < l} f_1(x, z) dz dx + \iint_{0 < y < z < l} f_2(y, z) dz dy + \iint_{0 < x < y < l} f_3(x, y) dy dx \\
&= \int_0^l \int_0^x g(x) dz dx + \int_0^l \int_y^l g(l-y) dz dy + \int_0^l \int_x^l \frac{1}{2} \{g(x) + g(l-y)\} dy dx \\
&= 2 \int_0^l x g(x) dx + \int_0^l (l-x) g(x) dx = \int_0^l (l+x) g(x) dx \\
&= G \int_0^l (l+x) \left\{ \frac{2}{9} \frac{l}{(l+x)^3} + \frac{1}{27} \frac{1}{(l+x)^2} - \frac{1}{27} \frac{1}{(2l-x)^2} \right\} dx \\
&= G \left[ -\frac{2}{9} \frac{l}{l+x} + \frac{1}{27} \log(l+x) - \frac{1}{9} \frac{l}{2l-x} - \frac{1}{27} \log(2l-x) \right]_0^l \\
&= G \left\{ \frac{1}{18} + \frac{2}{27} \log 2 \right\} = G \cdot \frac{3 + 4 \log 2}{54}. \tag{B.4}
\end{aligned}$$

Hence if we put  $G = \frac{54}{3 + 4 \log 2}$  as we have done at the beginning of the proof, the solution satisfies the equation (3.6).  $\square$

### Proof of Corollary 3.1

The throughput of the model is calculated by

$$\Lambda = \mu \iint_{0 < z < x < l} f_1(x, z) dz dx + \mu \iint_{0 < y < z < l} f_2(y, z) dz dy + 2\mu \iint_{0 < x < y < l} f_3(x, y) dy dx. \tag{B.5}$$

By using the function  $g(x)$ , it is expressed as  $\Lambda = 2\mu \int_0^l z g(z) dz + 2\mu \int_0^l (l-z) g(z) dz = 2\mu l \int_0^l g(z) dz$ . Then a direct integration leads us to  $\Lambda = \frac{1}{6} G \mu$ , and this proves (3.11).  $\square$

### **Acknowledgement**

The authors express their sincere thanks to the editor and two anonymous referees for their valuable comments, which have contributed very much to improve the paper.

### **References**

- [1] J.L. Doob: *Stochastic Processes* (Wiley, New York, 1953).
- [2] S.N. Ethier and T.G. Kurtz: *Markov Processes — Characterization and Convergence* (Wiley, New Jersey, 1986, 2005).
- [3] G. Novak: Movable Compact Shelving Systems: Selection and Specifications. *Library Technology Reports*, **35** (1999), 557–708.

Ken'ichi Katou  
Kanagawa University  
2946 Tuchiya, Hiratuka  
Kanagawa 259-1293, Japan  
E-mail: kkatou@kanagawa-u.ac.jp