

## A Control Policy for a Priority Queue with Removable Server

Henk Tijms

*Mathematisch Centrum, Amsterdam, The Netherlands*

(Received February 12, 1973)

This paper studies a control policy for an  $M/G/1$  nonpreemptive-priority queuing system with removable server and two priority classes. This policy turns off the server when the system is empty and turns him on when a given linear combination of the numbers of class 1 and class 2 customers in the system exceeds a certain value. Expressions for the long-run average numbers of class 1 and class 2 customers in the system are derived.

CONSIDER A SERVICE station with a single server at which customers of classes 1 and 2 arrive in accordance with independent Poisson processes with rates  $\lambda_1$  and  $\lambda_2$ , respectively. Customers of class 1 have nonpreemptive priority over customers of class 2. The order in which customers of a given priority class are served is immaterial in our considerations, assuming that this order is independent of the service times. A customer of class  $i$  will be called an  $i$ -customer,  $i=1, 2$ . Let the service times of different customers be independent random variables with finite first moment  $\mu_i$  and finite second moment  $\mu_i^{(2)}$  for  $i$ -customers. Let  $\rho_i = \lambda_i \mu_i$ . It is assumed that  $\rho < 1$ , where  $\rho = \rho_1 + \rho_2$ . The policy for controlling the system is to turn off the server only when the system is empty and to turn him on at the first time when the sum of  $\alpha_1$  times the number of 1-customers present and  $\alpha_2$  times the number of 2-customers present exceeds  $\beta$ , where  $\alpha_1, \alpha_2$ , and  $\beta$  are nonnegative constants with  $\alpha_1 + \alpha_2 > 0$ . We call this policy an  $(\alpha_1, \alpha_2, \beta)$ -policy.

For a single class of customers, such a policy has been studied by BALACHANDRAN,<sup>[1]</sup> BELL,<sup>[2]</sup> HEYMAN,<sup>[4]</sup> YADIN AND NAOR,<sup>[10]</sup> and others. The  $(\alpha_1, \alpha_2, \beta)$ -policy for the priority model was studied by Bell;<sup>[3]</sup> assuming that the service-time distributions do not differ for the two classes of customers and that there is a linear cost structure, he proved that an average-cost optimal policy exists that is of the  $(\alpha_1, \alpha_2, \beta)$ -type.

The purpose of this paper is to derive an expression for the average number of  $i$ -customers in the system,  $i=1, 2$ . Our derivation is based on the theory of regenerative processes (cf. STIDHAM<sup>[7]</sup>). Finally, we superimpose a linear cost structure and determine the best  $(1, 1, \beta)$ -,  $(1, 0, \beta)$ -, and  $(0, 1, \beta)$ -policies with respect to the average-cost criterion.

The results in this paper can be extended without difficulty to cover set-up and close-down times, as in Yadin and Naor.<sup>[10]</sup>

### 1. NOTATION AND PRELIMINARY RESULTS

FOR CONVENIENCE, WE ASSUME throughout that the server is turned off at epoch 0, so no customers are present at epoch 0. Let  $X$  be the next epoch at which the server

is turned off. For any  $t \geq 0$  and  $i=1, 2$ , let  $L_i(t)$  be the number of  $i$ -customers in the system at time  $t$  (including the  $i$ -customer being served at time  $t$ , if any). Observe that  $\int_0^t L_i(s) ds$  represents the total time spent by  $i$ -customers in the system during  $(0, t]$ . By the memoryless property of the Poisson process, any epoch at which the server is turned off is a regeneration epoch for the process  $\{L_i(t), t \geq 0\}$ ,  $i=1, 2$ . Let a *cycle* be the time interval between two successive epochs at which the server is turned off. We shall see in Section 3 that both  $X$  and  $\int_0^X L_i(s) ds$  have a finite expectation. Hence, by the theory of regenerative processes (p. 99 in Ross<sup>[6]</sup> and Theorem 1 of Stidham,<sup>[7]</sup>)

$$L^{(i)} = \lim_{t \rightarrow \infty} (1/t) E \left\{ \int_0^t L_i(s) ds \right\}$$

exists and is finite for  $i=1, 2$ . Also, the random variable  $(1/t) \int_0^t L_i(s) ds$  converges with probability 1 to  $L^{(i)}$  as  $t \rightarrow \infty$ . Furthermore, for  $i=1, 2$ ,

$$L^{(i)} = E \left\{ \int_0^X L_i(s) ds \right\} / EX, \quad (1)$$

i.e., the long-run average number of  $i$ -customers in the system is equal to the quotient of the expected total time spent by  $i$ -customers in the system during one cycle and the expected length of one cycle.

*Remark.* Let the average wait of an  $i$ -customer in the system be defined by  $W^{(i)} = \lim_{n \rightarrow \infty} (1/n) E \left\{ \sum_{k=1}^{k=n} W_{ki} \right\}$ , where  $W_{ki}$  denotes the time spent by the  $k$ th  $i$ -customer in the system (including his service time). Since the expected number of  $i$ -customers served during one cycle is finite, we have by Theorem 2 of Stidham<sup>[7]</sup> that  $W^{(i)}$  is well defined and finite for  $i=1, 2$ . Similarly, we may define  $L_q^{(i)}$  and  $W_q^{(i)}$ , where  $L_q^{(i)}$  represents the average number of  $i$ -customers in the queue and  $W_q^{(i)}$  represents the average wait of an  $i$ -customer in the queue. We have  $L^{(i)} = \lambda_i W^{(i)}$  and  $L_q^{(i)} = \lambda_i W_q^{(i)}$  for  $i=1, 2$  (see JEWELL<sup>[5]</sup> and Stidham.<sup>[8]</sup>) Since  $W_q^{(i)} = W^{(i)} - \mu_i$ , we have  $L_q^{(i)} = L^{(i)} - \rho_i$ ,  $i=1, 2$ .

Let  $\lambda = \lambda_1 + \lambda_2$ . Observe that, if we lump the two separate arrival processes together, the superimposed process is a Poisson process with rate  $\lambda$ . For  $i=1, 2$ , let  $p_i = \lambda_i / \lambda$ , so that  $p_i$  is the probability that an arbitrary customer is an  $i$ -customer. Let the probability distribution function  $F(x)$  be defined as follows: in case  $\alpha_1 \neq \alpha_2$ , the points  $\alpha_1$  and  $\alpha_2$  are points of increase of  $F$  with weights  $p_1$  and  $p_2$ ; otherwise, the point  $\alpha_1 (= \alpha_2)$  is a point of increase of  $F$  with weight 1. Denote by  $F^n(x)$  the  $n$ -fold convolution of  $F$  with itself, and let  $M(x) = \sum_{n=1}^{\infty} F^n(x)$ ,  $x \geq 0$ . The renewal function  $M(x)$  is the unique solution that is bounded on finite intervals to

$$M(x) = F(x) + \int_0^x F(x-y) dM(y), \quad (x \geq 0) \quad (2)$$

(p. 35 in Ross<sup>[6]</sup>). For any  $t \geq 0$ , let  $N_i(t)$  be the number of  $i$ -customers arriving in  $(0, t]$ . For any  $x \geq 0$ , let

$$T(x) = \inf \left\{ t \mid \sum_{i=1}^2 \alpha_i N_i(t) > x \right\}, \quad \nu_i(x) = N_i[T(x)],$$

$$W_i(x) = \int_0^{T(x)} L_i(s) ds. \quad (i=1, 2)$$

Given that an  $(\alpha_1, \alpha_2, x)$ -policy is used,  $\nu_i(x)$  represents the number of  $i$ -customers in the system at the first epoch at which the server is turned on, and  $W_i(x)$  repre-

sents the total time spent by  $i$ -customers in the system up to that epoch. Let  $\nu(x) = \nu_1(x) + \nu_2(x)$ ,  $x \geq 0$ . For any  $x \geq 0$ , let

$$\begin{aligned} a_i(x) &= E\nu_i(x), & b_i(x) &= E\{\nu_i(x)[\nu_i(x) - 1]\}, & w_i(x) &= EW_i(x), \\ a(x) &= E\nu(x), & b(x) &= E\{\nu(x)[\nu(x) - 1]\}, & c(x) &= E\{\nu_1(x)\nu_2(x)\}. \end{aligned} \quad (i=1, 2)$$

**THEOREM 1.** For  $i=1, 2$ , let  $a_i(u) = 0$  for  $u < 0$ , and let  $g_i(x) = p_i + 2p_i a_i(x - \alpha_i)$  for  $x \geq 0$ . Then, for any  $x \geq 0$ ,

$$a_i(x) = p_i\{1 + M(x)\}, \quad (i=1, 2) \quad (3)$$

$$b_i(x) = g_i(x) + \int_0^x g_i(x-y) dM(y) - a_i(x), \quad (i=1, 2) \quad (4)$$

$$a(x) = 1 + M(x), \quad b(x) = 2M(x) + 2 \int_0^x M(x-y) dM(y), \quad (5)$$

$$c(x) = (\nu_2) \{b(x) - b_1(x) - b_2(x)\}. \quad (6)$$

Let  $a(u) = 0$  for  $u < 0$ , and let  $h_i(x) = (1/\lambda)p_i a(x - \alpha_i)$  for  $x \geq 0$  and  $i=1, 2$ . Then,

$$w_i(x) = h_i(x) + \int_0^x h_i(x-y) dM(y) \quad \text{for } x \geq 0 \quad \text{and } i=1, 2. \quad (7)$$

*Proof.* Let  $T$  be the arrival epoch of the first customer and fix  $i$ . We can write  $\nu_i(x) = M + N$ , where  $M = 1$  if the first customer is an  $i$ -customer and  $M = 0$  otherwise, and  $N$  denotes the number of  $i$ -customers arriving in  $(T, T(x)]$ . Clearly, under the condition that the first customer is a  $j$ -customer, the random variable  $N$  has the same distribution as  $\nu_i(x - \alpha_j)$ , where  $\nu_i(u) = 0$  for  $u < 0$ . Now,  $a_i(x) = p_i + p_1 a_i(x - \alpha_1) + p_2 a_i(x - \alpha_2)$ , so  $a_i(x) = p_i + \int_0^x a_i(x-y) dF(y)$ ,  $x \geq 0$ . This is a renewal equation whose unique solution is given by (3). Let  $d_i(x) = E[\nu_i(x)]^2$ . Using  $[\nu_i(x)]^2 = M^2 + 2MN + N^2$ , we obtain  $d_i(x) = g_i(x) + \int_0^x d_i(x-y) dF(y)$ ,  $x \geq 0$ . The unique solution of this renewal equation is given by the sum of the first two terms from the right-side of (4). In the same way the relation (5) can be derived, while (6) follows from  $[\nu(x)]^2 = [\nu_1(x)]^2 + 2\nu_1(x)\nu_2(x) + [\nu_2(x)]^2$ .

By  $ET(y) = (1/\lambda)E\nu(y)$ , we have  $ET(y) = (1/\lambda)a(y)$  for  $y \geq 0$ . Using this and considering the waiting time of the first customer and that of the next customers separately, it follows as above that  $w_i(x)$  satisfies the renewal equation  $w_i(x) = h_i(x) + \int_0^x w_i(x-y) dF(y)$  for  $x \geq 0$ , so  $w_i(x)$  is given by (7). This ends the proof.

We note that, by (2) and (5),  $h_1(x) + h_2(x) = M(x)/\lambda$  for  $x \geq 0$ .

## 2. BASIC RESULT

THIS SECTION GIVES a result that will be basic in our derivation of an expression for  $L^{(i)}$ ,  $i=1, 2$ .

Denote by  $S(n_1, n_2)$  the time elapsed from the start of a service when  $n_1$  1-customers and  $n_2$  2-customers are in the system until the next epoch at which the system is empty. Let  $s(n_1, n_2) = ES(n_1, n_2)$ , and let  $u_i(n_1, n_2)$  be the expected total time spent by  $i$ -customers in the system during the time  $S(n_1, n_2)$ ,  $i=1, 2$  and  $n_1, n_2 = 0, 1, \dots$ .

It is routine to prove the next theorem (see pp. 6-7, 9-11, in TIJMS<sup>[9]</sup>).

**THEOREM 2.** For  $n_1, n_2 = 0, 1, \dots$ , with  $n_1 + n_2 > 0$ ,

$$\begin{aligned} s(n_1, n_2) &= t_b n_1 + \{n_2 + \lambda_2 t_b n_1\} t_{b2}, \\ u_1(n_1, n_2) &= w_1 n_1 + (\frac{1}{2}) t_b n_1 (n_1 - 1) + \{\lambda_2 t_b n_1 + n_2\} u_1(0, 1), \\ u_2(n_1, n_2) &= t_b (1 + \lambda_2 t_{b2}) n_1 n_2 + \{\lambda_2 t_b n_1 + n_2\} u_2(0, 1) \\ &\quad + (\frac{1}{2}) \{\lambda_2 + \lambda_2^2 t_{b2}\} \{t_b^{(2)} n_1 + t_b^2 n_1 (n_1 - 1)\} + (\frac{1}{2}) t_{b2} n_2 (n_2 - 1), \end{aligned}$$

where

$$\begin{aligned} t_b &= \mu_1 / (1 - \rho_1), & t_b^{(2)} &= \mu_1^{(2)} / (1 - \rho_1)^3, & t_{b2} &= \mu_2 / (1 - \rho), \\ w_1 &= \mu_1 / (1 - \rho_1) + \lambda_1 \mu_1^{(2)} / 2(1 - \rho_1)^2, \\ u_1(0, 1) &= (1 - \rho)^{-1} [\mu_1^{(2)} \lambda_1^2 \mu_2 / 2(1 - \rho_1) + \mu_2^{(2)} \lambda_1 / 2 + \rho_1 \mu_2], \\ u_2(0, 1) &= (1 - \rho)^{-1} [\mu_1^{(2)} \lambda_1 \rho_2 / 2(1 - \rho_1)(1 - \rho) + \mu_2^{(2)} \lambda_2 / 2(1 - \rho) + (1 - \rho_1) \mu_2]. \end{aligned}$$

### 3. THE AVERAGE NUMBER OF $i$ -CUSTOMERS IN THE SYSTEM

**THEOREM 3.** For  $i = 1, 2$ , let  $C_i = \rho_i + \lambda_i [2(1 - \rho_1)(1 - \delta_i)]^{-1} [\lambda_1 \mu_1^{(2)} + \lambda_2 \mu_2^{(2)}]$ , where  $\delta_1 = 0$  and  $\delta_2 = \rho$ . Then, for any  $(\alpha_1, \alpha_2, \beta)$ -policy,

$$L^{(1)} = C_1 + \lambda(1 - \rho)[1 + M(\beta)]^{-1} [w_1(\beta) + \mu_1 b_1(\beta) / 2(1 - \rho_1)], \quad (8)$$

$$\begin{aligned} L^{(2)} &= C_2 + \lambda(1 - \rho)[1 + M(\beta)]^{-1} [w_2(\beta) + \mu_1 c(\beta) / (1 - \rho) \\ &\quad + \mu_2 b_2(\beta) / 2(1 - \rho) + \lambda_2 \mu_1^2 b_1(\beta) / 2(1 - \rho_1)(1 - \rho)]. \end{aligned} \quad (9)$$

*Proof.* Using the Theorems 1 and 2, it is readily verified that the expected length of one cycle equals

$$(1/\lambda)[1 + M(\beta)] + E s[\nu_1(\beta), \nu_2(\beta)] = [\lambda(1 - \rho)]^{-1} [1 + M(\beta)].$$

The expected total time spent by  $i$ -customers in the system during one cycle equals  $w_i(\beta) + E u_i[\nu_1(\beta), \nu_2(\beta)]$ ,  $i = 1, 2$ . Now, by (1),

$$L^{(i)} = \lambda(1 - \rho)[1 + M(\beta)]^{-1} \{w_i(\beta) + E u_i[\nu_1(\beta), \nu_2(\beta)]\}. \quad (i = 1, 2)$$

Using Theorems 1 and 2, we obtain (8) and (9) after some algebra.

### 4. SPECIAL CASES OF THE $(\alpha_1, \alpha_2, \beta)$ -POLICY

WE CONSIDER THE following three cases.

*Case 1.*  $\alpha_1 = \alpha_2 = 1$  and  $\beta$  is a nonnegative integer. Then  $\nu_1(\beta)$  has a binomial distribution with parameters  $\beta + 1, p_1$ . Using this, we find

$$M(\beta) = \beta, \quad b_i(\beta) = p_i^2 \beta(\beta + 1), \quad c(\beta) = p_1 p_2 \beta(\beta + 1), \quad w_i(\beta) = (1/2\lambda) p_i \beta(\beta + 1).$$

From (8) and (9) we obtain, after some algebra,

$$L^{(1)} = C_1 + [2\lambda(1 - \rho_1)]^{-1} \lambda_1 (1 - \rho) \beta, \quad L^{(2)} = C_2 + [2\lambda(1 - \rho_1)]^{-1} \lambda_2 \beta.$$

*Case 2.*  $\alpha_1 = 1, \alpha_2 = 0$ , and  $\beta$  is a nonnegative integer. Then  $\nu(\beta)$  has a negative binomial distribution with parameters  $\beta + 1, p_1$ . Using this, we find

$$\begin{aligned}
1+M(\beta) &= (\beta+1)/p_1, & b_1(\beta) &= \beta(\beta+1), \\
b_2(\beta) &= (\lambda_2^2/\lambda_1^2)(\beta+1)(\beta+2), & c(\beta) &= (\lambda_2/\lambda_1)(\beta+1)^2, \\
w_1(\beta) &= (1/2\lambda_1)\beta(\beta+1), & w_2(\beta) &= (\lambda_2/2\lambda_1^2)(\beta+1)(\beta+2).
\end{aligned}$$

Next we find, after some algebra,

$$L^{(1)} = C_1 + [2(1-\rho_1)]^{-1}(1-\rho)\beta, \quad L^{(2)} = C_2 + [2\lambda_1(1-\rho_1)]^{-1}[\lambda_2\beta + 2\lambda_2(1-\rho_1)].$$

Case 3.  $\alpha_1=0$ ,  $\alpha_2=1$ , and  $\beta$  is a nonnegative integer. Then we find

$$L^{(1)} = C_1 + [2\lambda_2(1-\rho_1)]^{-1}[\lambda_1(1-\rho)(\beta+2)], \quad L^{(2)} = C_2 + [2(1-\rho_1)]^{-1}[\beta+2\rho_1].$$

Consider now the following cost structure. There is a holding cost of  $h_i > 0$  per unit time per  $i$ -customer in the system and a fixed cost of  $K > 0$  per cycle for turning the server off and for turning him on. Then, the long-run average cost per unit time equals  $h_1L^{(1)} + h_2L^{(2)} + K\lambda(1-\rho)[1+M(\beta)]^{-1}$ . Routine analysis shows that for Case  $j$  the long-run average cost is convex in  $\beta$  and is minimal for  $\beta$  one of the integers  $[\beta_j^*]$  and  $[\beta_j^*]-1$ , where  $\beta_2^* = (\lambda_1/\lambda)\beta_1^*$  and  $\beta_3^* = (\lambda_2/\lambda)\beta_1^*$  with

$$\beta_1^* = [2K\lambda^2(1-\rho_1)(1-\rho) / \{h_1\lambda_1(1-\rho) + h_2\lambda_2\}]^{1/2}.$$

If we put  $\lambda_2=0$  in the expression for  $\beta_1^*$ , we obtain the well known formula (34) in Yadin and Naor.<sup>[10]</sup>

#### ACKNOWLEDGMENTS

I AM INDEBTED to the referees for their careful reading and helpful suggestions, including several references.

#### REFERENCES

1. K. R. BALACHANDRAN, "Control Policies for a Single Server," *Management Sci.* **19**, 1013-1018 (1973).
2. C. E. BELL, "Characterization and Computation of Optimal Policies for Operating an  $M/G/1$  Queuing System with Removable Server," *Opns. Res.* **19**, 208-218 (1971).
3. ———, "Optimal Average Cost Operating Policy for an  $M/G/1$  Queuing System with Removable Server and Several Priority Classes," Technical Report No. 144, Department of Operations Research, Stanford University, 1971.
4. D. HEYMAN, "Optimal Operating Policies for  $M/G/1$  Queuing Systems," *Opns. Res.* **16**, 362-382 (1968).
5. W. S. JEWELL, "A Simple Proof of  $L = \lambda W$ ," *Opns. Res.* **15**, 1109-1116 (1967).
6. S. M. ROSS, *Applied Probability Models with Optimization Applications*, Holden-Day, Inc., San Francisco, California, 1969.
7. S. STIDHAM, JR., "Regenerative Processes in the Theory of Queues, with Applications to the Alternating Priority Queue," *Adv. Appl. Prob.* **4**, 542-577 (1972).
8. ———, " $L = \lambda W$ : A Discounted Analogue and a New Proof," *Opns. Res.* **20**, 1115-1126 (1972).
9. H. C. TIJMS, "A Priority Queuing System with Removable Server," Report BW 17/73, Mathematisch Centrum, Amsterdam, The Netherlands, 1973.
10. M. YADIN AND P. NAOR, "Queuing Systems with Removable Service Station," *Opnal. Res. Quart.* **14**, 393-405 (1963).