The Steady-State Behaviour of a Stochastic Clearing System with Bounded Waiting Times

By

Hisashi MINE and Viliam MAKIŠ

(Received March 6, 1984)

Abstract

The stochastic clearing system considered in this paper is characterized by an uncontrollable Poisson input process and bounded customers' wating times. We assume that all the quantity currently present in the system is instantaneously removed whenever there are at least M items in the queue, or every t time units since the first arrival after the last clearing, whichever occurs first. The objective is to study the steady-state behaviour of this system. Knowledge of this steady-state behaviour can be used for the evaluation of the system performance as a function of the system's parameters. We present explicit expressions for the queue length and waiting time distribution, the average queue length, and the average waiting time under steady-state conditions. This work is related to dispatching in transportation systems with stationary Poisson arrivals.

1. Introduction

Stochastic clearing systems are characterized by an uncontrollable stochastic input process and an output mechanism that removes instantaneously all the quantity currently present in the system whenever it exceeds a level Q. This class of stochastic processes, which has many applications in the study of dispatching problems, queues, dams and inventories, has been investigated by Ross,[5] Stidham,[8,9] Serfozo and Stidham,[6] Whitt,[12] and others. Ross,[5] considered a truck dispatching problem with Poisson arrivals and batch services and proved that the average cost optimal policy is a control limit policy, i.e., a service begins if and only if the queue length is at least as large as some control limit. This result was later extended by Deb and Serfozo[3] to an M/G/1 queue with batch services and a finite and an infinite service capacity.

In all these studies, no restriction was placed on the customers' waiting times. However, it is necessary to take this into account in many real situations such as transportation of perishable items, transhipment of mail and military supplies, and the processing of computer programs. On the other hand, in the serving of people by shuttles and other mass transportation systems, the control limit policy without any guarantee about passenger waiting times is not a very attractive policy if the input intensities are small.[11] Tapiero and Zuckerman[10] considered a vehicle dispatching problem for Poisson arrival processes and the following three dispatching policies:

(i) a capacity policy C, i.e., a vehicle is dispatched whenever it is filled to capacity C (a control limit policy),

(ii) a T-policy consisting in sending a vehicle every T periods,

(iii) a (T, C)-policy consisting in sending a vehicle every T periods or whenever it is filled to capacity C, whichever occurs first.

T and (T, C)-policies were considered to allow the dispatching of vehicles at less than full capacity in cases where the waitng time costs would be too large. The main disadvantage of T and (T, C)-policies is that they allow the dispatching of empty vehicles. Hence, such dispatching rules may be quite unrealistic for one-terminal systems. To overcome this, Makiš[4] considered a controlled M/G/1 batch service queueing system with bounded waiting times, i.e., each time of service is subject to control, and its choice is restricted by the requirement that the customers' waiting times cannot exceed a given constant t. The following cost structure was considered. The waiting cost of i customers in the queue per unit time is $h(i)$, where $h(.)$ is a non-negative real-valued function. The cost of serving i customers is $K+ci$, where $K>0$ and c are any real constant. This cost is charged at the beginning of a service. Viewing the system as a semi-Markov decision process with unbounded costs, it was proved that a policy which minimizes the expected average cost per unit time over an infinite time horizon is of the following simple form:

When the system is in state (i, s), $s<t$ (i is the number of customers waiting in the queue and s is the length of time since the first arrival in a given cycle), a service begins if and only if the server is free and i is at least as large as some control limit $i^*(t)$, and whenever the system reaches state (i, t) a service commences immediately for any positive integer i. This type of operating policy was termed a bounded control limit policy.

The objective of this paper is to derive a steady-state queue length and waiting time distribution for the stochastic clearing system with a Poisson input operated under a bounded control limit policy. Knowledge of this steady-state behaviour can be used for the evaluation of a system performance as a function of the system's parameters (the input intensity, control limit and constant t). The steady-state distributions are obtained using some asymptotic results for regenerative processes. (See e.g. Çinlar[1] and Stidham[7].) In the next paper, we examine the behaviour of the expected average cost per unit time in the class of T-polcies and in the class of bounded control limit policies. We find the optimal operating policies in each class and present some computational results.

2. The Steady-State Queue Length Distribution

Suppose customers arrive according to a Poisson process with intensity $\lambda>0$ and are served in batches under a bounded control limit policy. That is, if η_1 is the first arrival time in a given cycle, the service begins in the time interval $[\eta_1, \eta_1+t)$ if and only if the number of customers waiting in the queue is at least M where M is a given positive integer. Otherwise, the service begins at the time η_1+t. The cycle is completed each time a service begins. Suppose the system is in equilibrium after operating under a bounded control limit policy with a control limit M (henceforth a bounded M-policy) for a sufficient length of time. We find an explicit expression for the queue length distribution under steady-state conditions. The following notation will be used throughout the paper

$$q_k(t)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad R_k(t)=\int_0^t \lambda q_{k-1}(s)\,ds \tag{1}$$

for any non-negative integer k, $R_0(t)=1$ and let $\bar{\eta}_i=\eta_{i+1}-\eta_1$ where η_k is the k-th arrival time in a given cycle. Put $Z_s=(X_s, U_s)$ for $s\geqslant 0$, where X_s is the number of customers waiting in the queue at time s, and U_s is the length of time since the first arrival after the last clearing. We write $Z_s=0$, if there are no customers waiting in the system at time s. Then, $\{Z_s, s\geqslant 0\}$ is a regenerative process and we have from Propositions (7.2) and (7.6) in Çinlar[1] (cf. also Stidham[7])

$$P(0)\equiv\lim_{s\to+\infty}P(Z_s=0)=\frac{1}{\mu}\int_0^{+\infty}P(Z_y=0,\ \tau_1>y\,|\,Z_0=0)\,dy \tag{2}$$

and for $j\geqslant 1$, $u\geqslant 0$

$$P(j, u)\equiv\lim_{s\to+\infty}P(X_s=j,\ U_s\leqslant u)=\frac{1}{\mu}\int_0^{+\infty}P(X_y=j,\ U_y\leqslant u,\ \tau_1>y\,|\,Z_0=0)\,dy \tag{3}$$

where τ_1 is the length of the first cycle and $\mu=E(\tau_1)$ (assuming that $X_0=0$). We proceed to evaluate the expected length of a cycle. Obviously, under a bounded M-policy

$$\tau_1=\min\{\eta_1+t,\ \eta_M\} \tag{4}$$

where η_k is the k-th arrival time in a given cycle. Thus, conditioning on the number of arrivals Y_t in the time interval $(\eta_1, \eta_1+t]$ yields

$$\mathrm{E}\tau_1=\frac{1}{\lambda}+\sum_{k=0}^{M-2}tq_k(t)+\sum_{k=M-1}^{+\infty}\mathrm{E}(\bar{\eta}_{M-1}\,|\,Y_t=k)\,q_k(t). \tag{5}$$

Next, using Theorem 2.3[5] and formula (2.1.6),[2] we have for $k \geqslant M-1$

$$\mathrm{E}(\bar{\eta}_{M-2}|Y_t=k) = \frac{t^{-k}}{B(M-1,\, k-M+2)} \int_0^t s^{M-1}(t-s)^{k-M+1} ds$$
$$= \frac{(M-1)t}{k+1} \tag{6}$$

where

$$B(m,\, n) = \int_0^1 s^{m-1}(1-s)^{n-1} ds$$
$$= \frac{(m-1)!\,(n-1)!}{(m+n-1)!} \tag{7}$$

is a beta function. Thus, (5) and (6) yield after some algebra

$$\mu = \frac{1}{\lambda} \sum_{n=0}^{M-1} R_n(t) \tag{8}$$

where $R_n(t)$ is defined by (1). Next, we have

$$P(Z_y=0,\ \tau_1>y\,|\,Z_0=0) = e^{-\lambda y} \tag{9}$$

and for $1 \leqslant j \leqslant M-1$, $y \leqslant t$ conditioning on the first arrival time yields

$$P(X_y=j,\ U_y \leqslant u,\ \tau_1>y\,|\,Z_0=0) = \int_{y-u}^{y} \lambda e^{-\lambda v} q_{j-1}(y-v)\, dv$$
$$= e^{-\lambda(y-u)} q_j(u) \tag{10}$$

for $u \leqslant y \leqslant t$, and

$$P(X_y=j,\ U_y \leqslant u,\ \tau_1>y\,|\,Z_0=0) = q_j(y) \quad \text{for } y<u \tag{11}$$

where $q_j(y)$ is defined by (1). Similarly, for $y>t$, we get

$$P(X_y=j,\ U_y \leqslant u,\ \tau_1>y\,|\,Z_0=0) = e^{-\lambda(y-u)} q_j(u) \quad \text{for } u \leqslant t<y$$
$$= \int_{y-t}^{y} \lambda e^{-\lambda v} q_{j-1}(y-v)\, dv$$
$$= e^{-\lambda(y-t)} q_j(t) \quad \text{for } t<u. \tag{12}$$

From (2)−(12), we get

$$P(0) = 1/(\lambda\mu)$$

and for $1 \leqslant j \leqslant M-1$

$$P(j,\ u) = \frac{1}{\mu}\left(\int_0^u q_j(y)\, dy + \int_u^{+\infty} e^{-\lambda(y-u)} q_j(u)\, dy\right)$$
$$= \frac{1}{\lambda\mu}\left(\sum_{n=j+1}^{+\infty} q_n(u) + q_j(u)\right)$$
$$= \frac{1}{\mu} \int_0^u q_{j-1}(y)\, dy \quad \text{for } u \leqslant t \tag{13}$$

and

$$P(j, u) = \frac{1}{\mu}\left(\int_0^t q_j(y)\,dy + \int_t^{+\infty} e^{-\lambda(y-t)} q_j(t)\,dy\right)$$
$$= \frac{1}{\lambda\mu} R_j(t) \quad \text{for } u>t. \tag{14}$$

From (8), (13) and (14), we have the following theorem.

Theorem 1. Let $P(j)$, $j \geqslant 0$ be the steady-state probability that there are j customers waiting in the system. Then, for any bounded M-policy

$$P(j) = R_j(t) / \sum_{n=0}^{M-1} R_n(t) \quad \text{for } 0 \leqslant j \leqslant M-1$$
$$=0 \quad \text{otherwise} \tag{15}$$

where $R_n(t)$ is defined by (1).

Corollary 1. Under steady-state conditions, the average queue length is given by

$$L = \left(\sum_{j=0}^{M-1} jR_j(t)\right) / \sum_{n=0}^{M-1} R_n(t). \tag{16}$$

Obviously, (15) and (16) yield for $t \to +\infty$

$$P(j) = 1/M \quad \text{for } 0 \leqslant j \leqslant M-1$$
$$=0 \quad \text{otherwise}$$

and $L = (M-1)/2$, i.e., the steady-state queue length distribution under a capacity policy M is a uniform distribution. In the next section, we derive the waiting time distribution.

3. The Steady-State Waiting Time Distribution

In this section, the steady-state results will be obtained by using (13) and (14). We have the following theorem.

Theorem 2. Let $\overline{W}$ be the waiting time of an arriving customer assuming steady-state conditions. Then, for any bounded M-policy

$$P(\overline{W} \leqslant w) = \frac{1}{\lambda\mu}\left(\sum_{j=1}^{M-1} R_j(t) + R_{M-1}(w) - \sum_{j=1}^{M-2} R_j(t-w)(1 - R_{M-j-1}(w))\right) \quad \text{for } w<t$$
$$=1 \quad \text{for } w \geqslant t \tag{17}$$

where $R_n(s)$ and μ are given by (1) and (8), respectively.

Proof. Let U be the length of time since the first arrival after the last clearing, and X be the number of customers waiting in the queue immediately before the appropriate arrival assuming steady-state conditions. Then (13) and (14) yield for any $w<t$

$$P(\overline{W}\leqslant w)=P(0)\sum_{n=M-1}^{+\infty}q_n(w)+\sum_{j=1}^{M-1}\int_0^t P(\overline{W}\leqslant w\,|\,X=j,\ U=u)\,dP(j,\ u)$$
$$=\frac{1}{\lambda\mu}R_{M-1}(w)+\frac{1}{\mu}\sum_{j=1}^{M-2}(\int_{t-w}^{t}q_{j-1}(u)\,du+\sum_{n=M-j-1}^{+\infty}q_n(w)$$
$$\int_0^{t-w}q_{j-1}(u)\,du)+\frac{1}{\mu}\int_0^t q_{M-2}(u)\,du$$
$$=\frac{1}{\lambda\mu}(R_{M-1}(w)+R_{M-1}(t)+\sum_{j=1}^{M-2}(R_j(t)-R_j(t-w)\sum_{k=0}^{M-j-2}q_k(w))).$$

Since the waiting time of any customer never exceeds t, we have $P(\overline{W}\leqslant w)=1$ for $w\geqslant t$ and this completes the proof.

Corollary 2. The average waiting time of an arriving customer under steady-state conditions is of the form

$$W=(\sum_{j=0}^{M-1}jR_j(t))/(\lambda\sum_{n=0}^{M-1}R_n(t)). \tag{18}$$

Proof. From (17), using the obvious equality

$$\frac{dR_j(w)}{dw}=\lambda q_{j-1}(w)\quad \text{for } j\geqslant 1,$$

we get

$$W=\int_0^{+\infty}w\,dP(\overline{W}\leqslant w)=\frac{1}{\mu}(\int_0^t q_{M-2}(w)\,w\,dw+\sum_{j=1}^{M-2}\sum_{k=0}^{M-j-2}\int_0^t q_{j-1}(t-w)\,q_k(w)\,w\,dw$$
$$+\sum_{j=1}^{M-2}\int_0^t wR_j(t-w)\,q_{M-j-2}(w)\,dw)+t(1-\lim_{s\uparrow t}P(\overline{W}\leqslant s))$$
$$=\frac{1}{\lambda^2\mu}((M-1)R_M(t)+\sum_{j=1}^{M-2}\sum_{k=0}^{M-j-2}\frac{e^{-\lambda t}(\lambda t)^{k+j+1}}{(j-1)!(k)!}B(k+2,\ j)$$
$$+\sum_{j=1}^{M-2}\sum_{k=j}^{+\infty}\frac{e^{-\lambda t}(\lambda t)^{k+M-j}}{k!(M-j-2)!}B(M-j,\ k+1))+\frac{t}{\lambda\mu}\sum_{k=0}^{M-2}q_k(t) \tag{19}$$

where $B(.,\ .)$ is a beta function. From (7) and (19), we have

$$W=\frac{1}{\lambda^2\mu}((M-1)R_M(t)+\sum_{j=1}^{M-1}\frac{j(j+1)}{2}q_j(t)+\frac{(M-1)(M-2)}{2}R_M(t)) \tag{20}$$

and (18) follows from (20) and from

$$\sum_{j=1}^{M-1}\frac{j(j+1)}{2}q_j(t)+\frac{M(M-1)}{2}R_M(t)=\sum_{j=0}^{M-1}jR_j(t).$$

This completes the proof. Observe that formula (18) can be obtained also from (16), using the well-known formula $L=\lambda W$. For $w=0$, (17) yields

$$P(\overline{W}=0)=R_{M-1}(t)/\sum_{n=0}^{M-1}R_n(t)$$

and for $w=t$, we get

$$P(\overline{W}=t)=\sum_{k=0}^{M-2}q_k(t)/\sum_{n=0}^{M-1}R_n(t).$$

Now, we examine the case $M\to+\infty$. Obviously, this yields the following service policy : a service begins every t time units since the first arrival in each cycle. For $M\to+\infty$, we get from (15) and (16)

$$P(j)=R_j(t)/(1+\lambda t) \quad \text{for } j\geqslant 0,$$
$$L=\sum_{j=0}^{+\infty}jR_j(t)/(1+\lambda t)=\lambda t\left(1+\frac{\lambda t}{2}\right)/(1+\lambda t)$$

and from (17), the steady-state waiting time distribution is

$$P(\overline{W}\leqslant w)=\lambda w/(1+\lambda t) \quad \text{for } w<t$$
$$=1 \quad \text{for } w\geqslant t.$$

References

1) Çinlar, E., "Markov Renewal Theory: A Survey," Management Science, Vol. 21, 1975, pp. 727-752.
2) David, H. A., "Order Statistics," John Wiley and Sons, New York, 1970.
3) Deb, R. K., and Serfozo, R. F., "Optimal Control of Batch Service Queues," Advances Appl. Prob., Vol. 5, 1973, pp. 340-361.
4) Makiš, V., "Optimal Control of a Batch Service Queueing System with Bounded Waiting Time," (submitted to Journal of Appl. Prob.).
5) Ross, S. M., "Applied Probability Models with Optimization Applications," Holden Day, San Francisco, Calif., 1970.
6) Serfozo, R. F., and Stidham, S., Jr., "Semistationary Clearing Processes," Stoch. Processes Appl., Vol. 6, 1978, pp. 165-178.
7) Stidham, S., Jr., "Regenerative Processes in the Theory of Queues with Applications to the Alternating-Priority Queue," Advances Appl. Prob., Vol. 4, 1972, pp. 542-577.
8) Stidham, S., Jr., "Stochastic Clearing Systems," Stoch. Processes Appl., Vol. 2, 1974, pp. 85-113.
9) Stidham, S., Jr., "Cost Models for Stochastic Clearing Systems," Opns. Res., Vol. 25, 1977, pp. 100-127.
10) Tapiero, C. S., and Zuckerman, D., "Vehicle Dispatching with Competition," Transpn. Res., Vol. 13*B*, 1979, pp. 207-216.
11) Weiss, H. J., "Further Results on an Infinite Capacity Shuttle with Control at a Single Terminal," Opns. Res., Vol. 29, 1981, pp. 1212-1217.
12) Whitt, W., "The Stationary Distribution of a Stochastic Clearing Process," Opns. Res, Vol. 29, 1981, pp. 294-308.