*Probability in the Engineering and Informational Sciences*, **17**, 2003, 527–543. Printed in the U.S.A.

# ON THE WORKLOAD PROCESS IN A FLUID QUEUE WITH A RESPONSIVE BURSTY INPUT AND SELECTIVE DISCARDING

Parijat Dube
*IBM T.J. Watson Research Center*
*Yorktown Heights, New York 10598*
*E-mail: pdube@us.ibm.com*

Eitan Altman
*INRIA, Sophia Antipolis Cedex, France*
*E-mail: altman@sophia.inria.fr*

We analyze a feedback system consisting of a finite buffer fluid queue and a responsive source. The source alternates between silence periods and active periods. At random epochs of times, the source becomes ready to send a burst of fluid. The length of the bursts (length of the active periods) are independent and identically distributed with some general distribution. The queue employs a threshold discarding policy in the sense that only those bursts at whose commencement epoch (the instant at which the source is ready to send) the workload (i.e., the amount of fluid in the buffer) is less than some preset threshold are accepted. If the burst is rejected then the source backs off from sending. We work within the framework of Poisson counter-driven stochastic differential equations and obtain the moment generating function and hence the probability density function of the stationary workload process. We then comment on the stability of this fluid queue. Our explicit characterizations will further provide useful insights and "engineering" guidelines for better network designing.

A preliminary version of this paper appeared in the proceedings of Seventeenth International Teletraffic Congress, ITC 17, Salvador de Bahia, Brazil, December 2–7, 2001.

## 1. INTRODUCTION

Selective message discarding policies have been proposed [9] and are implemented in routers (e.g., in Cisco BP 8600 series) to prevent network congestion. This is particularly the case with the router supporting UBR (unspecified bit rate) service class of ATM, where message (i.e., a frame) discarding is employed to achieve the twin goals of reduced network congestion and increased goodput [10]. In the ATM context, message discarding is based on the idea that loss of a single packet results in the corruption of the entire message (to which it belongs) and hence it is advantageous to discard the entire remaining message. Two discarding mechanisms are frequently used: the *partial message discarding* (PMD), in which packets that belong to an already corrupted message are discarded, and the *early message discarding* (EMD), in which in addition to partial discarding, an admission control is applied to reject an entire message if upon arrival of its first packet, the queue exceeds some threshold value $K$ (threshold discarding) [10]. We have focused both on the discrete as well as on the fluid analysis (back-to-back message arrival with exponentially distributed message lengths) of the first mechanism in [4] and of the second in [5,6] with the goal of obtaining explicit expressions for performance metrics like the stationary distribution of the workload process and the goodput ratio. In both the packet model in [4,6] and the fluid model in [4,5], the source was nonresponsive to message discarding at the network element and thus the system was effectively open-loop. The discarding policies worked independently without any cooperation from the source (the source continues sending even if its data are being rejected by the network node). Also in the fluid model in [4,5], we have back-to-back messages with exponential distribution of length and thus the fluid arrival rate was deterministic.

In this article, we study responsive sources (in the spirit of [8,12]). We propose a model for the feedback system consisting of the network node with selective burst (the burst can be seen as a message) discarding and the source which responds to congestion signals (in our case, the congestion signal being positive if the queue length at the network node is higher than some preset threshold and is negative if the queue length is less than some preset threshold). This will help us in understanding the improvement in the performance achievable with combining selective burst discarding with congestion feedback to sources and responsiveness of sources in backing off from sending.

Our model consists of a finite buffer fluid queue fed by a source that alternates between off periods and active periods. An off period corresponds to a time interval when the source is not sending fluid and an active period corresponds to an interval when the source is actually transmitting fluid. At random epochs of times, the source becomes ready to send a burst of fluid. The lengths of the bursts (lengths of the active periods) are independent and identically distributed with an exponential distribution. The queue employs a threshold discarding policy in the sense that only those bursts at whose arrival epochs[1] the workload (i.e., the amount of fluid in the buffer) is less than some preset threshold are accepted. If the burst is rejected, then

[1]By arrival epoch of a burst we mean the time instant when the source is ready to transmit a “potential” burst.

the source backs off instantaneously from sending and goes into a silence (back-off) period. The off-period distribution is characterized as follows. After an exponentially distributed silence time, a new batch arrives; if it is accepted, then the off period ends. Otherwise, a new exponentially distributed silence period commences and so on. The off period is then the sum of the consecutive silence periods.

Our analysis employs *Poisson counter-driven stochastic differential equations* [2,3] for describing the workload dynamics. We obtain closed-form expressions for the distribution of the stationary workload process by first finding the Laplace–Stieltjes transform (LST) of the stationary workload process and then inverting it.

In Section 2, we formally define our model and state our main results on the LST and the density of the stationary workload process. The infinite buffer case along with its stability analysis is presented in Section 3. In Section 4, we present an approach for analyzing a policy which has partial discarding of bursts in addition to threshold burst discarding.[2]

## 2. MODEL: FORMAL DEFINITION

The fluid arrival rate is $h$ in the active period and zero in the off period (which models either the thinking time of the source or forced back-off by the source due to positive congestion feedback); the server has a constant capacity $c$. Let the buffer size be $B$ (maximum amount of fluid) and the threshold be $K$, $K < B$. Let the silence and back-off periods be exponentially distributed, both with parameter $\lambda_1$, and let the burst sizes also have an exponential distribution with parameter $\lambda_2$. Let the distribution of the off period between messages be as described in Section 1.

We have analyzed the same model but with infinite buffer in [7]. We extend here the analysis to the finite buffer case and obtain expressions for the stationary distribution of the workload process. The results for the infinite buffer can be obtained by taking $B \to \infty$ in the expressions for the finite buffer case and we present them as a special case of our model.

The discarding policy is such that if at the commencement epoch of a message the workload process $v(t)$ is less than $K$, the message is admitted, otherwise not. We assume[3] that $c < h$. Figure 1 explains the model (the source behavior and the workload process in the queue).

We write the dynamics of the system in terms of Poisson counter-driven stochastic differential equations [3]. Let $N_1$ and $N_2$ be Poisson counters with parameters $\lambda_1$ and $\lambda_2$, respectively. We define a new variable $x \in \{0,1\}$ as the indicator of the *actual arrival process* to the buffer. $x(t)$ captures the behavior of the discarding policy. The dynamics of $x(t)$ and of the workload $v(t)$ are

$$dx(t) = (-x(t) + 1)\, dN_1 \mathbf{I}(v(t) < K) - x(t)\, dN_2, \tag{1}$$

$$dv(t) = -c\mathbf{I}(v(t) > 0)\, dt + hx(t)\mathbf{I}(v(t) < B)\, dt + cx(t)\mathbf{I}(v(t) = B)\, dt. \tag{2}$$

[2]Recall that under partial discarding, once the buffer starts overflowing, the source backs off instantaneously and goes into a silence period.

[3]For the case $c \geq h$, the workload will always be zero w.p. 1.

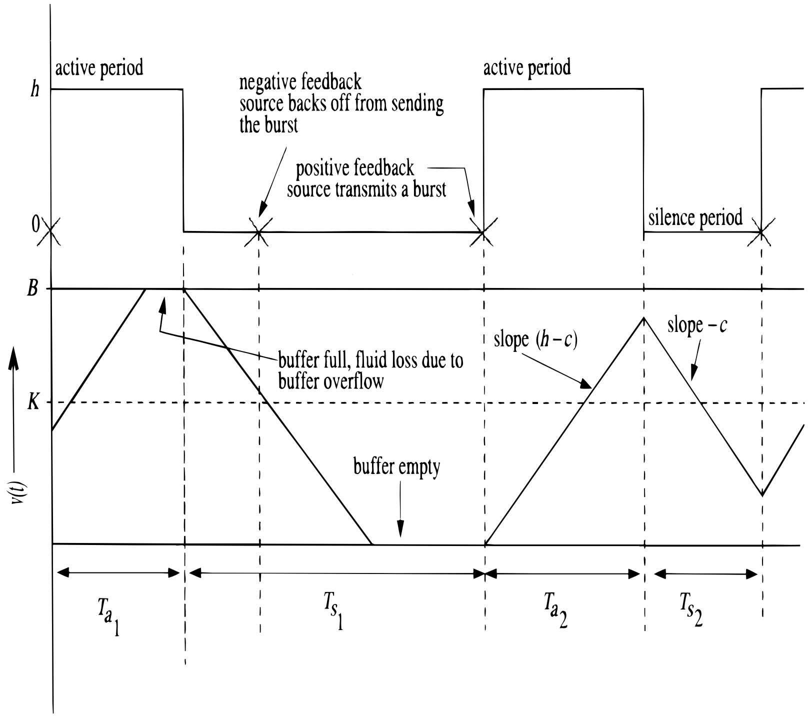


**FIGURE 1.** The dynamics of the arrival process and the workload in our model. $T_{a_n}$ and $T_{s_n}$ denote the $n$th active and silence periods, respectively.

*Remark 1:* Observe that in [3] a model for a finite buffer fluid queue without discarding has been studied. However, the expression for $dv(t)$ in Eq. (28) in [3, Sect. 4] does not have the boundary term $\mathbf{I}(v = B)$. This will result in a "chattering" of the fluid level at the boundary.

We next present our main results: The LST $V(s)$ $(= E[e^{-sv}])$ of $v$ is obtained from which we obtain the stationary probability density function, $\rho(v)$ of $v$. The LST will be obtained by deriving a recursive formula for the moments $E[v^n]$ for all positive integers $n$.

PROPOSITION 1: *$V(s)$ is given by*

$$V(s) = \frac{(p_2 - p_1)g_6}{1 + g_5 s}\left[\frac{(h-c)}{\lambda_2 + (h-c)s}\left(e^{-Ks} - e^{-Bs}e^{-\lambda_2(B-K)/(h-c)}\right) + \frac{(h-c)}{\lambda_2}\left(e^{-\lambda_2(B-K)/(h-c)} - 1\right)\right] + g_3\frac{(1-p_2)(e^{-Bs}-1)}{1+g_5 s} + 1 + \frac{s}{1+g_5 s}\left(\frac{h(h-c)g_0(1-p_2)}{c(\lambda_1+\lambda_2)} - \frac{hg_5}{c}E[x]\right) + \frac{g_5(1-p_2)s(e^{-Bs}-1)}{1+g_5 s}, \quad \textbf{(3)}$$

*where* $p_1 = \text{Prob}(v < K)$, $p_2 = \text{Prob}(v < B)$, *and*

$$g_0 = \left(1 - \frac{h\lambda_1}{c(\lambda_1 + \lambda_2)}\right)^{-1}, \qquad g_3 = 1 - \left(\frac{h\lambda_2}{c(\lambda_1 + \lambda_2)}\right) g_0,$$

$$g_5 = \frac{(h-c)}{\lambda_1 + \lambda_2}\, g_0, \qquad g_6 = \left[1 - \frac{\lambda_2}{\lambda_1 + \lambda_2}\, g_0\right] \frac{1}{(h-c)E[Y_1]},$$

$$E[x] = \frac{\lambda_1}{\lambda_1 + \lambda_2}\left(1 - (p_2 - p_1)\left(1 - \frac{c}{h}\right)\right), \qquad E[Y_1] = \frac{1}{\lambda_2}\,(1 - e^{-[(B-K)/(h-c)]\lambda_2}). \tag{4}$$

Corollary 1: *The stationary probability density function* $\rho(\cdot)$ *of* $v(t)$ *is given by*

$$\rho(m) = \begin{cases} \rho_0 & \textit{for } m = 0 \\ \rho_1(m) & \textit{for } 0 < m < K \\ \rho_2(m) & \textit{for } K \geq m < B \\ 1 - p_2 & \textit{for } m = B, \end{cases}$$

*where*

$$\rho_0 = \left[1 + \frac{h}{c}\left(\frac{(h-c)g_0(1-p_2)}{g_5(\lambda_1 + \lambda_2)} - E[x]\right) - (1 - p_2)\right], \tag{5}$$

$$\rho_1(m) = \frac{(p_2 - p_1)(h-c)g_6}{g_5\lambda_2}\,(e^{-\lambda_2(B-K)/(h-c)} - 1)e^{-g_5^{-1}m}$$
$$+ \left((1-p_2)(1-g_3) - \frac{h}{c}\left(\frac{(h-c)g_0(1-p_2)}{g_5(\lambda_1+\lambda_2)} - E[x]\right)\right) g_5^{-1} e^{-g_5^{-1}m}, \tag{6}$$

$$\rho_2(m) = \rho_1(m) + \frac{(p_2 - p_1)g_6}{\left(1 - \frac{\lambda_2 g_5}{(h-c)}\right)}\,(e^{-\lambda_2(m-K)/(h-c)} - e^{-g_5^{-1}(m-K)}) \tag{7}$$

*and where* $\delta(\cdot)$ *is the Dirac delta function and* $p_1$ *and* $p_2$ *are given by*

$$\begin{bmatrix} 1 + a_1 - a_2(1 + e^{-g_5^{-1}K}) & -1 - a_1 + (a_2 + a_6)(1 + e^{-g_5^{-1}K}) + (g_3 - 1)e^{-g_5^{-1}K} \\ -1 + a_3(a_2 + a_4) - a_5 & 1 + a_3(-a_2 + (g_3 - 1) - a_4 + a_6) + a_5 \end{bmatrix}$$
$$\times \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} a_6 - a_7 + e^{-g_5^{-1}K}((g_3 - 1) + a_6 - a_7) \\ a_3((g_3 - 1) + a_6 - a_7) \end{bmatrix} \tag{8}$$

*with*

$$a_1 = \frac{g_6(h-c)}{\lambda_2}\,(e^{-\lambda_2(B-K)/(h-c)} - 1)(1 - e^{-g_5^{-1}K}),$$

$$a_2 = \frac{\lambda_1}{\lambda_1+\lambda_2}\left(\frac{h}{c} - 1\right),$$

$$a_3 = e^{-g_5^{-1}K} - e^{-g_5^{-1}B},$$

$$a_4 = \frac{g_6(h-c)}{\lambda_2}\,(e^{-\lambda_2(B-K)/(h-c)} - 1),$$

$$a_5 = \frac{g_6(h-c)}{\lambda_2}\,\frac{(e^{-\lambda_2(B-K)/(h-c)} - e^{-g_5^{-1}(B-K)})}{1 - \dfrac{g_5\lambda_2}{(h-c)}},$$

$$a_6 = \frac{h}{c}\,\frac{(h-c)g_0}{g_5(\lambda_1+\lambda_2)},$$

$$a_7 = \frac{h}{c}\,\frac{\lambda_1}{\lambda_1+\lambda_2}.$$

### 2.1. Model Analysis and Proofs of the Main Results

Observe that the fluid level $v$ never remains steady at $K$ (visiting it only at isolated points in time) and hence $v$ does not have any probability mass at $K$. Thus, the distribution function is continuous at $K$. However, there is a probability mass at $v = B$ due to the buffer size being finite. From stochastic calculus,[4] we can write from (1) and (2)[5]

$$dv^{n+1} = (n+1)v^n\big(-c\mathbf{I}(v>0)\,dt + hx\mathbf{I}(v<B)\,dt + cx\mathbf{I}(v=B)\,dt\big) \tag{9}$$

$$\begin{aligned} dv^n x &= xnv^{n-1}\,dv + v^n\,dx \\ &= nv^{n-1}x\big(-c\mathbf{I}(v>0)\,dt + hx\mathbf{I}(v<B)\,dt + cx\mathbf{I}(v=B)\,dt\big) \\ &\quad + v^n[(-x+1)\,dN_1\mathbf{I}(v<K) - x\,dN_2]. \end{aligned} \tag{10}$$

Note that $P(v=B) = 1 - p_2$. From (9) and (10), we get

$$\begin{aligned} dE[v^{n+1}] &= (n+1)\big(-cE[v^n]\,dt + hE[v^n x \mid v<B]p_2\,dt + cE[v^n x\mathbf{I}(v=B)]\,dt\big), \\ dE[v^n x] &= \lambda_1\,dt\,E[v^n(-x+1)\mathbf{I}(v<K)] - \lambda_2\,dt\,E[v^n x] - cnE[v^{n-1}x]\,dt \\ &\quad + hnE[v^{n-1}x^2 \mid v<B]p_2\,dt + cnE[v^{n-1}x^2\mathbf{I}(v=B)]\,dt. \end{aligned} \tag{11}$$

[4]See the Appendix and [3].

[5]For notational convenience, henceforth we will not show $t$ in parentheses.

Observe that, $E[v^{n-1}x^2|v < B]p_2 = E[v^{n-1}x^2] - E[v^{n-1}x^2|v = B](1 - p_2) = E[v^{n-1}x] - B^{n-1}(1 - p_2)$. This is because $v = B$ only when $x = 1$ and also $E[v^{n-1}x^2] = E[v^{n-1}x]$ for any $n \geq 1$. Thus, we get

$$\begin{aligned} dE[v^n x] = (h-c)nE[v^{n-1}x]\,dt - hnB^{n-1}(1-p_2)\,dt + E[v^n|v<K]p_1\lambda_1\,dt \\ - E[v^n x|v<K]\lambda_1 p_1\,dt - E[v^n x|v<B]\lambda_2 p_2\,dt \\ - B^n\lambda_2(1-p_2)\,dt + cnB^{n-1}(1-p_2)\,dt. \end{aligned} \qquad \textbf{(12)}$$

By the total probability argument, we can write

$$E[v^n|v<B]p_2 = E[v^n|v<K]p_1 + E[v^n|K<v<B](p_2-p_1),$$
$$E[v^n x|v<B]p_2 = E[v^n x|v<K]p_1 + E[v^n x|K<v<B](p_2-p_1).$$

Thus, we can write (11) and (12) as

$$\begin{aligned} dE[v^{n+1}] &= (n+1)\big(-cE[v^n|v<K]p_1 - cE[v^n|K<v<B](p_2-p_1) \\ &\quad + hE[v^n x|v<K]p_1 + hE[v^n x|K<v<B](p_2-p_1)\big)\,dt \\ dE[v^n x] &= -ncE[v^{n-1}x]\,dt + nhE[v^{n-1}x|v<B]p_2\,dt \\ &\quad + ncB^{n-1}(1-p_2)\,dt - E[v^n x|v<K]p_1\lambda_1\,dt \\ &\quad + E[v^n|v<K]p_1\lambda_1\,dt - E[v^n x|v<K]p_1\lambda_2\,dt \\ &\quad - E[v^n x|K<v<B](p_2-p_1)\lambda_2\,dt - B^n(1-p_2)\lambda_2\,dt. \end{aligned}$$

Thus, for the existence of the steady state, the following should vanish:

$$\begin{pmatrix} p_1\lambda_1 & 0 & -(\lambda_1+\lambda_2)p_1 & -\lambda_2(p_2-p_1) \\ -cp_1 & -c(p_2-p_1) & hp_1 & h(p_2-p_1) \end{pmatrix} \begin{pmatrix} E(v^n|v<K) \\ E(v^n|K<v<B) \\ E(v^n x|v<K) \\ E(v^n x|K<v<B) \end{pmatrix}$$
$$+ \begin{pmatrix} n(h-c)E[v^{n-1}x] - (1-p_2)(n(h-c)B^{n-1} + B^n\lambda_2) \\ 0 \end{pmatrix}. \qquad \textbf{(13)}$$

Thus, we have two equations in six unknowns (the four conditional expectations and $p_1$ and $p_2$). However, an important observation to be made here is that for any $n \geq 1$, $E[v^n|K<v<B]$ and $E[v^n x|K<v<B]$ can be calculated alternatively as follows:

When $v \geq K$ and the state of the modulating process changes from zero to one for the first time (and succeeding times), then the incoming fluid is not accepted. This is done until $v < K$. Also, even if the current on period started when $v < K$ (and hence accepted) but $v$ reaches the level $B$ before the on period ends, then the excess fluid corresponding to this on period (which arrives when $v = B$) is lost due to buffer overflow. The queue length can only cross the threshold of $K$ at any time $t$, if $x(t) = x(t^-) = 1$. Because the sojourn time in a state is exponentially distributed, the excess time the Markov chain spends in state 1 after time $t$ is again exponentially distributed given that the chain was in state 1 at time $t$. The fluid level will rise steadily with a rate $(h-c)$ from $t$ until the Poisson process $N_2$ causes $x$ to change from one to zero or until $v = B$ (whichever occurs earlier). Then, either $v$ will stay at $B$ (if the excess time the Markov chain spends in state 1 after time $t$ is greater than $(B-K)/(h-c)$) or $v$ will start decreasing at a steady rate of $c$ until the buffer level is $K$. During this period (when $v > K$), even if a new message arrives (with fluid arriving at rate $h$), the fluid is not accepted (the fact highlighted by the presence of an indicator function $\mathbf{I}(v < K)$ in (1)), and even if the current on period started when $v < K$ (and hence accepted) but $v$ reaches the level $B$ before the on period ends, then the excess fluid corresponding to this on period (which arrives when $v = B$) is lost (the fact highlighted by the presence of an indicator function $\mathbf{I}(v < B)$ in (2)). Let $T_i$, $i = 1, 2, \ldots$ be the random variable denoting the time spent by $v$ after crossing $K$ at the $i$th cross and during which the condition $K < v < B$ is true. Thus,

$$T_i = Y_i + \frac{Y_i(h-c)}{c} = Y_i \frac{h}{c}, \tag{14}$$

where

$$Y_i = \min\left(\frac{B-K}{h-c}, X_i\right)$$

and $X_i$ is the excess sojourn time in state 1. Note that the condition $v(t) > K$ implies that at the time of crossing the level $K$, say at time $t_1 \le t$, $x(t_1) = 1$. Thus, the sequence $\{T_i\}$ is independent and identically distributed (since $X_i \sim \exp \lambda_2$) and

$$E[Y_i] = E\left[X_i \mathbf{I}\left(X_i < \frac{B-K}{h-c}\right) + \left(\frac{B-K}{h-c}\right)\mathbf{I}\left(X_i > \frac{B-K}{h-c}\right)\right],$$

which gives the expression for $E[Y_1]$ in (4). Observe that we can write

$$\begin{aligned} E[v^n | K < v < B] = {} & E\left[v^n | K < v < B, \frac{dv}{dt} = (h-c)\right] P\left(\frac{dv}{dt} = (h-c) | K < v < B\right) \\ & + E\left[v^n | K < v < B, \frac{dv}{dt} = -c\right] P\left(\frac{dv}{dt} = -c | K < v < B\right). \end{aligned} \tag{15}$$

From (14), we have $Y_i = (c/h)T_i$. Thus, we have

$$P\left(\frac{dv}{dt} = (h-c) | K < v < B\right) = \frac{c}{h}$$

and

$$P\left(\frac{dv}{dt} = -c | K < v < B\right) = 1 - P\left(\frac{dv}{dt} = h - c | K < v < B\right) = 1 - \frac{c}{h}.$$

Thus, from (15), we can write

$$E[v^n|K<v<B] = E\left[v^n|K<v<B, \frac{dv}{dt}=(h-c)\right]\frac{c}{h} + E\left[v^n|K<v<B, \frac{dv}{dt}=-c\right]\left(1-\frac{c}{h}\right).$$

Now, observe that when the condition $(dv/dt = h-c, K<v<B)$ is true, $v(t) := v_1(t) = (h-c)(t-t_0)+K$, where $t_0$ is the last time the workload process was at $K$. Thus, we have by the Renewal Reward Theorem [13]

$$\begin{aligned} E\left[v^n|K<v<B, \frac{dv}{dt}=h-c\right] &= \frac{E\left[\int_{t_0}^{t_0+Y_1}(v_1(t))^n\,dt\right]}{E[Y_1]}, \\ &= E\left[\frac{(K+Y_1(h-c))^{n+1}-K^{n+1}}{(n+1)(h-c)}\right](E[Y_1])^{-1} \\ &= E\left[v^n|K<v<B, \frac{dv}{dt}=-c\right] \\ &= E[v^n|K<v<B]. \end{aligned} \tag{16}$$

We proceed to evaluate $E[v^n x|K<v<B]$. By similar arguments, it is given by

$$E\left[v^n x|K<v<B, \frac{dv}{dt}=(h-c)\right]\frac{c}{h} + E\left[v^n x|K<v<B, \frac{dv}{dt}=-c\right]\left(1-\frac{c}{h}\right).$$

Also $x=1$ when $(dv/dt = h-c, K<v<B)$ and $x=0$ when $(dv/dt=-c, K<v<B)$. Thus

$$\begin{aligned} E[v^n x|K<v<B] &= E\left[v^n x|K<v<B, \frac{dv}{dt}=h-c\right]\frac{c}{h} \\ &= \left(E\left[\frac{(K+Y_1(h-c))^{n+1}-K^{n+1}}{(n+1)(h-c)}\right](E[Y_1])^{-1}\right)\frac{c}{h}. \end{aligned} \tag{17}$$

Observe that from (13) for $n=1$, we will require an expression for $E[x]$. Next, we continue the analysis and obtain an expression for $E[x]$ in terms of $p_1$ and $p_2$.

LEMMA 1: *$E[x]$ is given by*

$$E[x] = \frac{\lambda_1}{\lambda_1+\lambda_2}\left(1-(p_2-p_1)\left(1-\frac{c}{h}\right)\right). \tag{18}$$

PROOF: In the steady state from (1), we get

$$\lambda_2 E[x] = p_1\lambda_1 - E[x|v<K]p_1\lambda_1. \tag{19}$$

Also,

$$\begin{aligned} E[x] &= E[x|v<K]p_1 + E[x|K<v<B](p_2-p_1) + E[x|v=B](1-p_2) \\ &= E[x|v<K]p_1 + \Bigg(E\left[x|K<v<B, \frac{dv}{dt}=h-c\right] \\ &\qquad \times P\left(\frac{dv}{dt}=h-c|K<v<B\right) \\ &\qquad + E\left[x|K<v<B, \frac{dv}{dt}=-c\right] \\ &\qquad \times P\left(\frac{dv}{dt}=-c|K<v<B\right)\Bigg) \\ &\quad \times (p_2-p_1) + (1-p_2) \\ &\Rightarrow E[x|v<K]p_1 = E[x] - \frac{c}{h}(p_2-p_1) - (1-p_2). \end{aligned} \tag{20}$$

The equivalence follows as $x=0$ for the case $(K<v<B, dv/dt=-c)$. From (19) and (20), we obtain (18). ■

Proof of Proposition 1: We eliminate $E[v^n x|v<K]$ from the steady-state equations. After some calculations, we obtain

$$\begin{aligned} E[v^n|v<K] = \Bigg[\frac{h}{c}\Big(\big(E[v^n x|K<v<B]\lambda_1(p_2-p_1) + n(h-c)E[v^{n-1}x] \\ - (1-p_2)B^n(n(h-c)B^{-1}+\lambda_2)\big)/(\lambda_1+\lambda_2)\Big) \\ - E[v^n|K<v<B](p_2-p_1)\Bigg]\left(p_1 - \frac{hp_1\lambda_1}{c(\lambda_1+\lambda_2)}\right)^{-1}. \end{aligned} \tag{21}$$

Replacing the left-hand side of the last equation by $p_1^{-1}(E[v^n] - E[v^n|K<v<B](p_2-p_1) - B^n(1-p_2))$ and using $E[v^n x|K<v<B] = E[v^n|K<v<B] \times (c/h)$ from (17), we get, from (21),

$$\begin{aligned} E[v^n] = g_1 nE[v^{n-1}x] + g_2(p_2-p_1)E[v^n|K<v<B] + g_3(1-p_2)B^n \\ + g_4 n(1-p_2)B^{n-1}, \end{aligned} \tag{22}$$

with

$$g_0 = \left(1 - \frac{h\lambda_1}{c(\lambda_1+\lambda_2)}\right)^{-1}, \qquad g_1 = \frac{h(h-c)}{c(\lambda_1+\lambda_2)}g_0, \qquad g_2 = \left[1 - \frac{\lambda_2}{\lambda_1+\lambda_2}g_0\right],$$

$$g_3 = \left[1 - \left(\frac{h\lambda_2}{c(\lambda_1+\lambda_2)}\right)g_0\right], \qquad g_4 = \frac{-h(h-c)}{c(\lambda_1+\lambda_2)}g_0.$$

Also, with $E[v^n x] = E[v^n x\mathbf{I}(v < B)] + B^n(1 - p_2)$, from the steady-state behavior of (9), we can write (for $n \geq 1$)

$$E[v^n x] = \frac{c}{h} E[v^n] + B^n \frac{(h-c)}{h} (1 - p_2). \tag{23}$$

From (16), (23), and (22) we get, for $n \geq 2$,

$$\begin{aligned} E[v^n] = {} & g_5 nE[v^{n-1}] + g_1 \frac{(h-c)}{h} (1-p_2) nB^{n-1} \\ & + g_6 \frac{(p_2 - p_1)}{(n+1)} E[(K + Y_1(h-c))^{n+1} - K^{n+1}] \\ & + g_3(1-p_2)B^n + g_4(1-p_2)nB^{n-1}. \end{aligned} \tag{24}$$

For $n = 1$ from (16) and (22), we have

$$E[v] = g_1 E[x] + g_6 \frac{(p_2 - p_1)}{2} E[(K + Y_1(h-c))^2 - K^2] + g_7(1-p_2)B^n, \tag{25}$$

where

$$g_5 = g_1 \frac{c}{h}, \qquad g_6 = \frac{g_2}{E[Y_1](h-c)}, \qquad g_7 = \left[1 - \left(\frac{h(h-c)B^{-1} + h\lambda_2}{c(\lambda_1 + \lambda_2)}\right) g_0\right],$$

$E[Y_1]$ being obtained earlier in (4).

Thus, we have a recursive relation between $E[v^n]$ and $E[v^{n-1}]$ for $n \geq 1$. With these, we will proceed to find the LST of the stationary workload. Further multiplying both the sides of (24) by $(-s)^n/n!$ and summing from $n = 2$ to $\infty$, we can write (with $g_1 + g_4 = -g_5$)

$$\begin{aligned} \sum_{n\geq 2} \frac{(-s)^n}{n!} E[v^n] \\ = {} & g_5(-s) \sum_{n\geq 2} E[v^{n-1}] \frac{(-s)^{n-1}}{(n-1)!} - g_5(1-p_2)(-s) \sum_{n\geq 2} \frac{(-Bs)^{n-1}}{(n-1)!} \\ & + g_6(p_2 - p_1)E\left[\frac{1}{(-s)} \sum_{n\geq 2} [(K + Y_1(h-c))^{n+1} - K^{n+1}] \frac{(-s)^{n+1}}{(n+1)!}\right] \\ & + g_3(1-p_2) \sum_{n\geq 2} B^n \frac{(-s)^n}{n!}. \end{aligned} \tag{26}$$

Adding to (26) and multiplying (25) by $(-s)$, we get

$$\begin{aligned}
&\sum_{n\geq 1} \frac{(-s)^n}{n!} E[v^n] \\
&\quad = g_5(-s) \sum_{n\geq 2} E[v^{n-1}] \frac{(-s)^{n-1}}{(n-1)!} + g_5(1-p_2)s \sum_{n\geq 2} \frac{(-Bs)^{n-1}}{(n-1)!} \\
&\qquad + g_6(p_2-p_1)E\left[ \frac{1}{(-s)} \sum_{n\geq 1} [(K+Y_1(h-c))^{n+1} - K^{n+1}] \frac{(-s)^{n+1}}{(n+1)!} \right] \\
&\qquad + g_3(1-p_2) \sum_{n\geq 1} B^n \frac{(-s)^n}{n!} - sg_1 E[x] + \frac{h(h-c)g_0}{c(\lambda_1+\lambda_2)} s(1-p_2). \qquad \textbf{(27)}
\end{aligned}$$

Also, observe that

$$\begin{aligned}
&E\left[ \frac{1}{-s} \sum_{n\geq 1} [(K+Y_1(h-c))^{n+1} - K^{n+1}] \frac{(-s)^{n+1}}{(n+1)!} \right] \\
&\quad = E\left[ \frac{e^{-Ks}}{s} (1-e^{-Y_1(h-c)s}) - Y_1(h-c) \right],
\end{aligned}$$

$$\begin{aligned}
&E\left[ \frac{e^{-Ks}}{s} (1-e^{-Y_1(h-c)s}) - Y_1(h-c) \right] \\
&\quad = \frac{(h-c)}{\lambda_2+(h-c)s} (e^{-Ks} - e^{-Bs}e^{-\lambda_2(B-K)/(h-c)}) \\
&\qquad + \frac{(h-c)}{\lambda_2} (e^{-\lambda_2(B-K)/(h-c)} - 1).
\end{aligned}$$

Furthermore, $V(s) = \sum_{n\geq 0}(-s)^n E v^n/n!$; thus, we get from (27)

$$\begin{aligned}
V(s) - 1 = {} & -g_5 s(V(s)-1) + g_6(p_2-p_1) \\
& \times \left[ \frac{(h-c)}{\lambda_2+(h-c)s} (e^{-Ks} - e^{-Bs}e^{-\lambda_2(B-K)/(h-c)}) \right. \\
& \qquad \left. + \frac{(h-c)}{\lambda_2} (e^{-\lambda_2(B-K)/(h-c)} - 1) \right] \\
& + g_3(1-p_2)(e^{-Bs}-1) + s\left( \frac{h(h-c)g_0(1-p_2)}{c(\lambda_1+\lambda_2)} - g_1 E[x] \right) \\
& + g_5(1-p_2)s(e^{-Bs}-1),
\end{aligned}$$

which implies (3). ■

Proof of Corollary 1: Taking the inverse LST of (3), we get, for $0 \le v \le K$, (6). For $K < v \le B$, the inverse LST implies

$$\rho(v) = \rho_2(v) = \rho_1(v) + \mathcal{L}^{-1}\left(\frac{(p_2 - p_1)g_6(h-c)e^{-Ks}}{(1+g_5 s)(\lambda_2 + (h-c)s)}\right), \tag{28}$$

where $\mathcal{L}^{-1}$ denotes the inverse LST.[6] Thus (with $*$ denoting the convolution operator),

$$\mathcal{L}^{-1}\left(\frac{1}{(s+g_5^{-1})\left(s+\dfrac{\lambda_2}{h-c}\right)}\right) = e^{-g_5^{-1}v} * e^{-\lambda_2 v/(h-c)}$$

$$= \frac{1}{\left(g_5^{-1} - \dfrac{\lambda_2}{(h-c)}\right)}\,(e^{-\lambda_2 v/(h-c)} - e^{-g_5^{-1}v}). \tag{29}$$

Thus, from (29) and (28), we get

$$\rho_2(m) = \rho_1(m) + \frac{(p_2-p_1)g_6}{\left(1 - \dfrac{\lambda_2 g_5}{(h-c)}\right)}\,(e^{-\lambda_2(m-K)/(h-c)} - e^{-g_5^{-1}(m-K)}), \tag{30}$$

which is (7). Observe that $\int_0^K \rho_1(v)\,dv = p_1$ and $\int_K^B \rho_2(v)\,dv = p_2 - p_1$. Thus, we get from integrating (5) and (6), and (7) with limits $[0, K)$ and $[K, B]$ respectively, two linear equations in two unknowns $p_1$ and $p_2$ which gives explicit closed form solutions for $p_1$ and $p_2$ in (8). ∎

## 3. INFINITE BUFFER CASE: THE WORKLOAD PROCESS AND THE STABILITY ANALYSIS

Taking $B \to \infty$ in Corollary 1, we have the following.

Corollary 2 ([7, Prop. 2]): *The stationary probability density function $\rho(\cdot)$ of $v(t)$ is*

$$\rho(m) = \begin{cases} \rho_1(m) & \text{for } 0 \le m < K \\[1ex] \dfrac{k_2(1-p)\dfrac{a\lambda_2}{(h-c)}}{\left(a - \dfrac{\lambda_2}{(h-c)}\right)}\,(e^{-[\lambda_2/(h-c)](m-K)} - e^{-a(m-K)}) + \rho_1(m) & \\ \qquad \text{for } m \ge K, & \end{cases}$$

[6]For a random variable $X$ with distribution $F(x)$ and LST $\mathcal{L}(s) (= \int_0^\infty e^{-sx}\,dF(x))$, we mean by $\mathcal{L}^{-1}$, the probability density (if it exists) of $X$ (i.e., $\mathcal{L}^{-1}[\mathcal{L}(s)] = dF(x)/dx$).

*where*

$$\rho_1(m) = \delta(m)\left(1 - \frac{h}{c}E[x]\right) + \left(\frac{h}{ck_1}E[x] - \frac{k_2}{k_1}(1-p)\right)e^{-am},$$

$$k_1 = \left[\frac{(h-c)}{(\lambda_1+\lambda_2)}\left(1 - \frac{h\lambda_1}{c(\lambda_1+\lambda_2)}\right)^{-1}\right],$$

$$k_2 = \left[1 - \frac{\lambda_2}{\lambda_1+\lambda_2}\left(1 - \frac{h\lambda_1}{c(\lambda_1+\lambda_2)}\right)^{-1}\right], \qquad a = k_1^{-1}$$

$$E[x] = \frac{\lambda_1}{\lambda_1+\lambda_2}\left(p\left(1-\frac{c}{h}\right)+\frac{c}{h}\right), \qquad p = \frac{1}{1 + \dfrac{\alpha}{1 - \dfrac{\alpha c}{h}}},$$

$$\alpha = \frac{\lambda_1 e^{-aK}h}{(\lambda_1+\lambda_2)c}(1 - k_2 + k_2 e^{-aK})^{-1}.$$

*Remark 2:* Taking $K$ to infinity in Corollary 2, we get

$$\rho(m) = \left(1 - \frac{\lambda_1 h}{c(\lambda_1+\lambda_2)}\right)\delta(m) + \frac{h\lambda_1}{c(\lambda_1+\lambda_2)}\lambda_2 e^{\lambda_2 m},$$

which is the probability density function for the stationary workload in a single server infinite buffer queue and is of course the same as the results of Anick, Mitra, and Sondhi [1].

Next, we establish the stability of the queue for the infinite buffer case. In particular, we show that the workload process is a renewal process. To that end, we consider the Markov chain $\mathcal{V}_n$, which is the workload as seen by the commencement of the $n$th potential arriving message. We will show that $V_n$ is a Harris recurrent Markov chain and that the empty state is recurrent whenever $c > 0$. This will then imply that the original workload process is a renewal process. To that end, we recall the following sufficient condition for the Harris recurrence, which follows from [11, Thm. 14.0.1, p. 330].

Lemma 2: *Assume that $\mathcal{V}_n$ is $\Psi$-irreducible (for some $\Psi$) and aperiodic. Let there be a function $f$, some $\epsilon > 0$, and some small set $C$ such that*

$$E[f(V_{n+1}) - f(V_n) | V_n = v] < -\epsilon + \mathbf{I}\{v \notin C\}. \qquad \textbf{(31)}$$

*Then, the Markov chain $\mathcal{V}_n$ is positive recurrent, it has a stationary probability $\pi$, and the $n$-step transition probabilities $P^n$ converge to the $\pi$ in total variation as $n \to \infty$.*

In Lemma 2, a set $C$ is *small* if there exists some integer $n$, a constant $g > 0$, and a probability measure $\phi$ over the state space $R$ such that

$$[P^n]_{xA} > g\phi(A) \quad \text{for all } x \in C \text{ and measurable set } A \subset R. \tag{32}$$

THEOREM 1: *Assume that $c > 0$. Then, the process $\mathcal{V}_n$ is Harris recurrent.*

PROOF: Define $C = [0, K]$ and $f(v) = v$. Let $T$ be an exponentially distributed random variable with parameter $\lambda_2$. Then, for all $v \notin C$,

$$E[\mathcal{V}_{n+1} - \mathcal{V}_n | \mathcal{V}_n = v] < -cE\left[\min\left(T, \frac{K}{c}\right)\right] =: -\epsilon.$$

Thus, (31) holds. Next, we check that $C$ is indeed a small set. Let $Z_i$ be an exponentially distributed random variable with parameter $\lambda_i$, $i = 1,2$. Viewing $Z_1$ as the length of the off period and $Z_2$ as the length of the on period, we have for any $v \in C$ and measurable $A \subset R$,

$$[P^2]_{vA} \geq P\left(\frac{K}{h-c} < Z_2 < \frac{2K}{h-c}, Z_1 > \frac{3K}{c}\right) P_{0A}.$$

Thus, (32) holds with $\phi(A) = P_{0A}$ and $g = P(Z)$, where

$$Z = \left\{\frac{K}{h-c} < Z_2 < \frac{2K}{h-c}, Z_1 > \frac{3K}{c}\right\}.$$

Note that on $Z$, a message starts to arrive when $v < K$ and then no more messages are accepted until the system empties.

Observe that $\mathcal{V}_n$ is $\phi$-irreducible [11, pp. 70 and 87] because the probability of eventually reaching any measurable set $A \subset R$ from any state $x$ is greater than $g\phi(A)$. Finally, the aperiodicity follows because the probability to go from state 0 to state 0 is strictly positive. ■

## 4. COMBINING PARTIAL AND THRESHOLD DISCARDING

Observe that in our discarding policy, we continue accepting the fluid of an accepted burst, even if during the arrival period of the burst, the queue hits $B$ and the buffer overflows. Thus, the distribution function has a positive mass at $v = B$. A much more efficient policy is one in which a source also backs off when the queue hits $B$. Thus, there may be partial discarding of an accepted message.

The difference between our model and the one in which we combine both threshold (like in our model) and partial discarding policies is that now when $B$ is reached, immediately discarding begins. Unlike the threshold discarding, this time not a whole message is discarded but only the remaining message (after $B$ is reached). Hence, the probability mass that we had at $B$ (see Fig. 1) disappears.

In both types of discarding (partial at the boundary and complete at thresholds), we assume that when discarding begins, the source changes immediately from the active period to a silence period. Thus, fluid stops being injected to the system once discarding begins and a "thinking time" begins until the arrival of the next batch.

Consider a sample path of $v(t)$ in our previous model. Let $\tau_i$ be the $i$th time $v(t)$ hits $B$ and let $\sigma_i$ be the $i$th time it leaves $B$. Define $S_i, i = 1,2,\ldots$ to be the time interval $(\tau_i, \sigma_i]$. We now construct a new sample $\hat{v}(t)$ that is obtained by *eliminating* the periods $S_i$ from $v(t)$ (by simply "cutting" them out). Then, it is easy to see that $\hat{v}(t)$ has the same distribution as the process in the new model that combines both discarding mechanisms. We conclude that the stationary probability density function of the new model is given by $\rho(\cdot)/p_2$, where $\rho$ and $p_2$ (the probability of not being at $B$) are given in Corollary 1; we thus have also the probability distribution of the model with combined discarding mechanisms.

### References


1. Anick, D., Mitra, D., & Sondhi, M.M. (1982). Stochastic theory of a data-handling system with multiple sources. *The Bell System Technical Journal* 61(8): 1871–1894.
2. Brockett, R.W. (1983). *Stochastic control*. Lecture Notes. Harvard University, Cambridge, MA.
3. Brockett, R.W., Gong, W.B., & Guo, Y. (1999). New analytical methods for queueing systems. *IEEE Conference on Decision and Control*.
4. Dube, P. & Altman, E. (2001). Queueing and fluid analysis of partial message discard policy. In *Proceedings of the 9th IFIP Working Conference on Performance Modelling and Evaluation of ATM & IP Networks (IFIP ATM & IP 2001)*.
5. Dube, P. & Altman, E. (2002). Fluid analysis of early message discard policy under heavy traffic. In *Proceedings of INFOCOM 2002*.
6. Dube, P. & Altman, E. (2002). Queueing analysis of early message discarding policy. In *Proceedings of IEEE ICC 2002*.
7. Dube, P. & Altman, E. (2000). On fluid analysis of queues with selective message discarding policies. In *Proceedings of 38th Annual Allerton Conference on Communication, Control and Computing*.
8. Floyd, S. (1994). TCP and explicit congestion notification. *ACM Computer Communication Review* 24(5): 10–23.
9. Floyd, S. & Romanow, A. (1994). Dynamics of TCP traffic over ATM networks. In *Proceedings of ACM SIGCOMM '94*, pp. 79–88.
10. Lapid, Y., Rom, R., & Sidi, M. (1998). Analysis of discarding policies in high-speed networks. *IEEE Journal on Selected Areas in Communications* 16(5): 764–777.
11. Meyn, S.P. & Tweedie, R.L. (1993). *Markov chain and stochastic stability*. New York: Springer-Verlag.
12. Ramakrishnan, K.K., Floyd, S., & Black, D. (2001). The addition of explicit congestion notification (ECN) to IP. *RFC 3168, Proposed Standard*. Available at ftp://ftp.isi.edu/innotes/rfc3168.txt.
13. Wolff, R. (1989). *Stochastic modeling and theory of queues*. Englewood Cliffs, NJ: Prentice-Hall.


## APPENDIX[7]

Consider a stochastic differential equation driven by $n$ independent Poisson counters $N_1, \ldots, N_n$,

$$dx = f(x)\,dt + \sum_{i=1}^{n} g_i(x)\,dN_i, \qquad x \in R^n.$$

[7]This material summarizes results from [3] which we make use of.

Then, we have the following Itô rule:

If $\phi : R^n \to R$ is a differentiable function, then

$$d\phi(x(t)) = \left\langle \frac{\partial \phi}{\partial x}, f(x) \right\rangle dt + \sum_{i=1}^{n} [\phi(x(t) + g_i(x(t))) - \phi(x(t))]\, dN_i.$$

Also, since $x(t)$ is continuous from the left and the Poisson counter is taken to be continuous from the right, we have

$$\frac{d}{dt}(E[x(t)]) = E[\,f(x(t))] + \sum_{i=1}^{n} (E[g_i(x(t))])\lambda_i,$$

and, similarly, one can interchange the expectation and derivative operators for $\phi(x(t))$ also. In our analysis for any integer $n \geq 1$, we have the following relations:

$$\begin{aligned}
\frac{dv^n(t)}{dt} &= \frac{\partial v^n}{\partial v}\, dv = nv^{n-1}\, dv, \\
\frac{d(v^n(t)x(t))}{dt} &= nv^{n-1}x\, dv + v^n\, dx + [v^n(x - x + 1) - v^n x]\mathbf{I}(v < K)\, dN_1 \\
&\quad + [v^n(x - x) - v^n x]\, dN_2 \\
&= nv^{n-1}x\, dv + v^n\, dx.
\end{aligned}$$