Queueing and Fluid Analysis of Partial Message Discarding Policy

Parijat Dube and Eitan Altman
Project MISTRAL, INRIA,
2004, Route des Lucioles, B.P. 93, 06902, Sophia Antipolis Cedex, France.
E-mail: Parijat.Dube, Eitan.Altman@sophia.inria.fr

Abstract

A message (or a frame) is a group of consecutive packets. Often, a loss of one packet from the message can result in the loss of the whole message. This is the case in ATM where a transport layer protocol (the AAL layer) is responsible for this grouping. Thus, packets of a frame, that arrive after a packet is lost from the same frame, are useless for the receiver, and an intelligent policy is to discard them so as to avoid the occurrence of future congestion and also to maximize the delivery of complete (*good*) messages. We provide here a queueing (packet level) and a fluid analysis of such a policy. The first part of the paper deals with a packet model and we obtain exact expressions for the stationary queue length distribution and the *Goodput ratio*. The second part deals with a fluid model and the fluid analysis leads to simple closed form expressions for the stationary workload process and the fluid *Goodput ratio*. Our analytical results may also be quite useful in dimensioning the buffer size that is required to achieve a desired Quality of Service (for e.g., bound on the message delay, goodput, throughput etc.).

Keywords: PMD policy, Packet model, Queues with vacation, Goodput, Fluid queue.

1 Introduction

Often, a set of consecutive packets are grouped into a frame, and loss of one packet results in the loss of the whole frame. This is the case in ATM where a transport layer protocol (known as AAL) is responsible for this grouping, see e.g. Chapter 5 in [9].

Thus, packets of a frame, that arrive after a packet is lost from the same frame, are useless at the receiver (in the sense that the original message cannot be recovered successfully from this incomplete set of packets), and it is advantageous to discard them thus achieving the twofold objective of congestion avoidance and goodput maximization. This approach for discarding is known as the Partial Message Discard (PMD) policy. This policy as well as other discarding approaches have been studied in several previous papers [3, 4, 13, 8] and [11].

In [4, 13], the basic performance measure for the study of discarding policies is the *effective throughput*, which is the ratio of good packets on the outgoing link to the total outgoing flow. However, as argued in [11], a more suitable performance measure is the *Goodput*, defined as the ratio of good packets out of the total number of packets that arrive at the network element's input.

The goal of this paper is to present *explicit expressions* for the stationary queue size distribution and the goodput of the PMD policy. Most of the previous works [8, 11], deal with numerical studies of the performance of PMD policy. The first part of the paper considers a Markovian framework: a Poisson process of packet arrivals, geometrically distributed frame size, and exponentially distributed service times of packets. Explicit expressions for the queue size distribution and of the goodput are obtained based on recursions introduced in [11]. The originality of the first part is in providing closed form expressions for the stationary distribution of queue size and the goodput ratio. We also propose an interpretation of our model as equivalent to a dual of a vacation model (queue with vacation servers) and provide an alternate approach for obtaining the limiting (when buffer size approaches infinity) queue size distribution in the heavy traffic regime.

In the second part of the paper we derive fluid approximations which are valid for heavy traffic conditions. The input process (which may be quite general) is approximated by a fluid with a constant rate. We obtain explicit expressions for the workload process distribution and the goodput for the fluid approximation. The fluid analysis yields analytically tractable, simple expressions which will be helpful in analytical study of the sensitivity of the goodput to different parameters for e.g., the message length, the buffer size etc., which were studied numerically in earlier works [11], [8]. Also, the nature of dependence of goodput to various parameters is clear, in particular the *goodput depends on the mean message length and the buffer size only through their product*. Our analytical results may be quite useful in dimensioning the buffer and/or capacity that is required for a given required goodput under PMD policy.

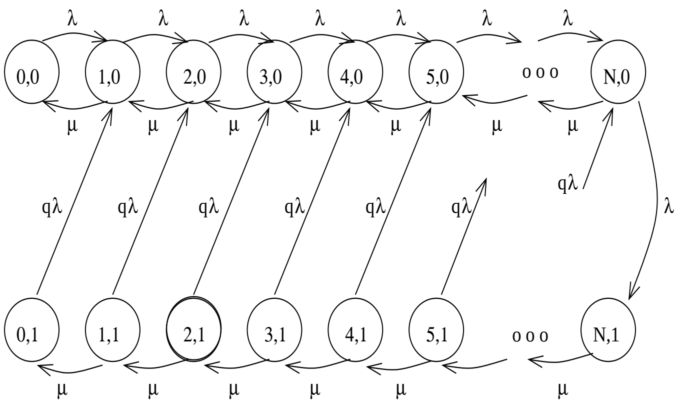

Figure 1: Transition structure under the PMD policy

2 Packet Model

The packet model is the same as the one employed in [11]. In terms of packets the network element is a $M/M/1/N$ queue with the arrival rate λ and the service rate μ and the load $\rho = \lambda/\mu$. Though we do a single node (router) analysis, we would like to comment that analysis by approximating the whole chain of routers (between the source and the destination) by one single router which experiences the maximum losses (*the bottleneck*) has both theoretical and experimental [7, 6] justification (see also [2]). In this sense our single node should be looked on as the bottleneck node. The service time represents the time between the beginning of the transmission of a packet on the bottleneck interface until the beginning of the transmission of the next packet from the same flow.[1] A message length (in terms of packets) is considered to be geometrically distributed with parameter q. In PMD policy, if a packet arrives when the queue is full, it is discarded and all the subsequent packets belonging to the same message are also discarded, irrespective of the state of the queue upon their arrival epochs, until the head-of-message packet (i.e., a new message) arrives. To model the policy, two modes for working of the network element are defined: the *normal mode*, in which packets are admitted, and the *discarding mode*, in which arriving packets are discarded. The state transition diagram for PMD policy under this model is shown in Figure (1). Let $P_{i,j} (0 \leq i \leq N, j = 0, 1)$ be the steady-state probability of having i packets in the system and the system is in mode $j (j = 0$ for normal; $j = 1$ for discarding). Thus, we have the following set of equations for

[1]Although we consider the analysis of a single connection, our model could also be useful for the case of multiplexing. In the latter case, two packets of a flow can be spaced apart by a random number of packets from different flows; we may add this to the service time of a packet and use the exponential distribution as an (approximating) candidate for modelling the service times in an equivalent model with a single flow.

the steady-state probabilities [11] from Figure (1).

$$\begin{aligned}
\rho P_{0,0} &= P_{1,0} && (1)\\
q\rho P_{0,1} &= P_{1,1} && (2)\\
(\rho+1)P_{i,0} &= \rho P_{i-1,0} + P_{i+1,0} + q\rho P_{i-1,1},\ \text{for } 1 \le i \le N-1 && (3)\\
(q\rho+1)P_{i,1} &= P_{i+1,1},\ \text{for } 1 \le i \le N-1 && (4)\\
(\rho+1)P_{N,0} &= \rho P_{N-1,0} + q\rho P_{N-1,1} && (5)\\
P_{N,1} &= \rho P_{N,0} && (6)\\
\sum_{i=0}^{N}(P_{i,0}+P_{i,1}) &= 1 && (7)
\end{aligned}$$

Let $Q_j(z) = \sum_{i=0}^{N} z^i P_{i,j}(j = 0$ for normal mode and $j = 1$ for discarding mode) and $Q(z) = Q_0(z) + Q_1(z) = \sum_{i=0}^{N}(P_{i,0} + P_{i,1})z^i$.

2.1 PGF and distribution of the number of packets in the queue

Proposition 1 *The probability generating function $Q_j(z)$ is given by*

$$Q_0(z) = P_{0,1}q\left[\rho^{-1}\frac{1-(\rho^{-1}(1+\rho q))^N}{1-\rho^{-1}(1+\rho q)} + \sum_{i=1}^{N}\frac{\left[1-(\rho^{-1}(1+\rho q))^{N-i+1}\right](1+\rho q)^{i-1}}{1-\rho^{-1}(1+\rho q)}z^i\right] \quad (8)$$

$$Q_1(z) = P_{0,1}\frac{(1-z-z^{N+1}(q\rho+1)^N q\rho)}{1-z(q\rho+1)} \quad (9)$$

with

$$P_{0,1} = \frac{(1-\rho)(\rho(1-q)-1)}{D}$$

and

$$D = q(1-\rho^{-N}(1+\rho q)^N) + (\rho(1-q)-1)(1-\rho(1+\rho q)^N)$$

The proof is given in the Appendix.

By taking the inverse *z-transform* of Equation (8) and (9) we obtain:

Corollary 1 *The steady state probabilities are given by,*

$$P_{i,0} = \begin{cases} \frac{P_{0,1}q}{\rho} \frac{1-\left(\rho^{-1}(1+\rho q)\right)^N}{1-\rho^{-1}(1+\rho q)} & \textit{for } i=0 \\ \\ P_{0,1}q \frac{\left[1-\left(\rho^{-1}(1+\rho q)\right)^{N-i+1}\right](1+\rho q)^{i-1}}{1-\rho^{-1}(1+\rho q)} & \textit{for } 1 \le i \le N \end{cases}$$

$$\textit{and} \qquad P_{i,1} = \begin{cases} \frac{(1-\rho)(\rho(1-q)-1)}{D} & \textit{for } i=0 \\ \\ q\rho(1+q\rho)^{i-1}P_{0,1} & \textit{for } 1 \le i \le N \end{cases}$$

2.2 An equivalent vacation model

Next, we propose an interpretation of our model (denoted by P) as equivalent to a dual vacation model (denoted by P_d). This interpretation will especially be useful later, in considering the case of large buffers. Let $A_{t,N}$ be the number of packets in the P model at time t. We define a random variable, $Y_{t,N}$ as,

$$Y_{t,N} = N - A_{t,N}.$$

$Y_{t,N}$[2] will be the number of packets in the equivalent vacation model. In other words, the number of packets in P_d equals the number of vacant places in the P. It thus follows that the service times in P_d are i.i.d. exponentially distributed with parameter λ, and that the arrival process to P_d is Poisson distributed with parameter μ. We note, however, that in the original process, arrivals are stopped during the *discarding period*. This *discarding period* will correspond to a *vacation period* in the dual model. More precisely, we define a discarding period in P as the duration from the instant that the buffer fills, till the next time the event B^1 occurs, where $B^1 = \{$ service occurs and the next packet to arrive after that service is not discarded$\}$.

Note that with this definition, the number of packets discarded during a discarding period may be zero. Indeed, with probability $\mu/(\lambda+\mu) = 1/(1+\rho)$, the first packet that arrives after the queue fills, will find the queue non-full and will not be discarded, and there are zero discarded packets in the discarding period.

A particularly important quantity in the equivalent vacation model is the distribution of the number of arrivals during a vacation denoted by V, or alternatively, the number of service during a discarding period in the original model. Note that by definition of the original discarding period, $V \ge 1$.

[2] N in subscript is for the buffer size N

Let T denote the instant of the beginning of a discarding period, and let S denote the instant when the first service completion occurs after T. Let V_1 (V_2, resp.) be a r.v. distributed as V given that at time S, the message that is being transmitted is bad (good, resp.). In other words, $V =_d V_2$ iff the next packet to arrive after time S is not discarded. The latter occurs iff the following event B^2 occurs; $B^2 = \{$either there is no arrival during the interval (T, S), or there is at least one arrival but the last arrival in that interval is the last packet of a message$\}$. Let

$$\beta := \frac{(1-q)\rho}{1+\rho} \text{ thus } 1-\beta = \frac{1+q\rho}{1+\rho}$$

Thus $V =_d V_1$ with probability β and $V =_d V_2 = 1$ with probability $1-\beta$.

Next we study the distribution of V_1. V_1 equals in distribution to one plus $\mathcal{A} :=$ the number of services that occur during the duration of M arrivals, where M is distributed according to a geometric distribution with parameter q. Let $B(L)$ be the number of services in a random duration L. Then $B_L^*(z)$, the PGF of $B(L)$, is given by

$$B_L^*(z) \quad = \quad E[\sum_{i=0}^{\infty} e^{-\mu L} \frac{(\mu L)^i}{i!} z^i] = E[e^{-\mu L(1-z)}] = L^*(\mu(1-z))$$

where $L^*(.)$ is the LST of L. Thus we need to evaluate $L^*(.)$.

$$L^*(s) = E[e^{-sL}] \quad = \quad E[E[e^{-sT_i}]^M] = E[(T^*(s))^M] = \mathcal{N}(T^*(s))$$

where $T^*(.)$ is LST of an interarrival time ($\sim \exp \lambda$) and $\mathcal{N}(.)$ is the PGF of a geometrically distributed r.v. with parameter q. Observe that,

$$T^*(s) = \frac{\lambda}{\lambda + s} \text{ and } \mathcal{N}(z_1) = \frac{q z_1}{1-(1-q)z_1}$$

Thus, if we denote by $\alpha(z)$ the PGF of $\mathcal{A}$, then

$$\begin{aligned} \alpha(z) &= \mathcal{N}(T^*(\mu(1-z))) = \mathcal{N}\left(\frac{\lambda}{\lambda + \mu(1-z)}\right) \\ &= \frac{q}{\rho^{-1}(1-z) + q} \end{aligned}$$

The PGF of V is given by $z((1-\beta) + \beta\alpha(z))$.

2.3 The case of large buffer

We use now the interpretation proposed in Subsection 2.2 as the dual of a vacation model in order to study the behavior of our system as the buffer size N becomes large. Clearly, non-trivial distribution of $Y_{t,N}$ is obtained in the limit $N \to \infty$ only in the heavy traffic regime $\rho > 1$.

Observe that,

$$\lim_{N\to\infty} P(N - A_{t,N} = k) = \lim_{N\to\infty} P(Y_{t,N} = k) = P(Y_t = k).$$

In [5] (see also [12]) the authors have shown that the stationary number of customers present in a M/G/1 queueing system with generalized server vacation is a convolution of the distribution function of two independent positive random variables (*stochastic decomposition*), one of which being the stationary distribution of the number of customers in an ordinary M/G/1 queueing system without server vacations. The other corresponds to the PGF of the number of packets at an arbitrary moment in a vacation. Let $\phi(.)$ and $\pi(.)$ be the p.g.f. for the stationary distribution of the number of customers at a random point in time in the in the vacation system and in the standard M/G/1 queueing system, respectively. Also, let $\hat{\alpha}(.)$ denote the PGF of the random variable V (i.e., the number of customers that arrive during a vacation). Then, with arrival rate μ (service rate in PMD queue) and departure rate λ (arrival rate in PMD queue), and $\rho = \frac{\lambda}{\mu}$, we have from [5], [3]

$$\phi(z) \quad = \quad \frac{1-\hat{\alpha}(z)}{\dot{\hat{\alpha}}(1)(1-z)}\pi(z)$$

with,

$$\pi(z) \quad = \quad \frac{(1-\rho^{-1})(1-z)B^*(\mu-\mu z)}{B^*(\mu-\mu z)-z}$$

where, $B^*(.)$ is the Laplace transform of the service time pdf. For our $M/M/1/V$ case, $\pi(z)$ simplifies to, $\pi(z) = \frac{1-\rho^{-1}}{1-\rho^{-1}z}$. Using the result of the previous subsection, and inverting the PGF $\phi(z)$, we get finally:

$$P(Y=k) \quad = \quad \frac{q(\rho-1)}{(\rho(1-q)-1)}(\rho(1-q)(1+\rho q)^{-(k+1)} - \rho^{-(k+1)})$$

[3] we denote by $\dot{f}(x)$ as the derivative of f with respect to x in all following discussions

One can now check that this is indeed the limit obtained as $\lim_{N\to\infty} P(A_N = N-k)$ from Corollary 1. We have, for $1 \le k \le N-1$,

$$\begin{aligned}
\lim_{N\to\infty} P(A_N = N-k) &= \lim_{N\to\infty} (P_{N-k,0} + P_{N-k,1}) \\
&= \lim_{N\to\infty} \left[\rho(\rho-1)q(1+q\rho)^{N-(k+1)}[\rho(1-q) - \rho^{-(k+1)}(1+\rho q)^{-(k+1)}]\right] D^{-1} \\
&= P(Y = k)
\end{aligned}$$

2.4 Goodput Ratio: $\mathcal{G}$

The goodput is defined in [11] as the ratio between total packets comprising good messages exiting the system and the total arriving packets at its input. Let $\mathcal{W}$ be the length (number of packets) of an arriving message, Q denote the queue length seen by the first packet of the arriving message and $\mathcal{V}$ be the random variable representing the success of a message ($\mathcal{V} = 1$ for a good message, and $\mathcal{V} = 0$ for a message which has one or more dropped packets). Then $\mathcal{G}$ can be expressed (see [11]), with $S_{n,i} \stackrel{\triangle}{=} P(\mathcal{V} = 1|\mathcal{W} = n, \mathcal{Q} = i)$, as

$$\mathcal{G} = q\sum_{n=1}^{\infty} nq(1-q)^{n-1}\sum_{i=0}^{N} P(\mathcal{V} = 1|\mathcal{W} = n, \mathcal{Q} = i)P(\mathcal{Q} = i) \tag{10}$$

In [11], recursions for evaluating these probabilities and hence $\mathcal{G}$ were given. We will present here an explicit expression for $\mathcal{G}$. To do this we will use the multidimensional generating function for probabilities $S_{n,i}$ which was obtained in a different context in [10]. By some abuse of notation let us denote $\bar{S}_i(x)(= \sum_{n=1}^{\infty} S_{n,i}x^{n-1})$ and by $\bar{S}_n(y)(= \sum_{i=0}^{N} S_{n,i}y^i)$ as the generating function for probabilities $S_{n,i}$ ($1 \le n \le \infty$, $0 \le i \le N$) for fixed i and fixed n respectively. We also define the two-dimensional generating function of $S_{n,i}$ as $\bar{S}(x,y)$, i.e., $\bar{S}(x,y) = \sum_{n=1}^{\infty}\sum_{i=0}^{N} S_{n,i}y^i x^{n-1}$

Proposition 2 *The probability generating function $\bar{S}(x,y)$ can be expressed as $\bar{S}(x,y) = \sum_{i=0}^{N} c_i(x)y^i$ where, for $0 \le i \le N-1$,*

$$c_i(x) = \begin{cases} 1 + K_3\left(A_1 - A_2 y_1^{N-(i+1)} - A_3 y_2^{N-(i+1)}\right) + K_4\left(B_1 y_1^{N-i} + B_2 y_2^{N-i}\right), & 0 \le i \le N-1 \\ 0, & i = N \end{cases}$$

with, $y_{1,2} = \frac{1+\rho\pm\sqrt{(1+\rho)^2 - 4\rho x}}{2}$, $K_3 = -x\rho$, $K_4 = \frac{x\rho(y_1^N - y_2^N)}{y_2^{N+1}(y_1-\rho) - y_1^{N+1}(y_2-\rho)}$, $A_1 = \frac{1}{(1-y_1)(1-y_2)}$, $A_2 = \frac{1}{(1-y_1)(y_1-y_2)}$, $A_3 = \frac{1}{(1-y_2)(y_2-y_1)}$ *and* $B_1 = -B_2 = \frac{1}{y_1-y_2}$ [4].

[4] It should be noted that all the apparent constants $y_{1,2}$, $K_{3,4}$, $A_{1,2,3}$ and $B_{1,2}$ are functions of x

Proof:
From [10], we have,

$$[(1-\alpha y)\alpha y - x\rho\alpha^2]\bar{S}(x,y) = \frac{1-y^N}{1-y}(1-\alpha y)\alpha y - x\rho\alpha^2(\alpha y)^{N+1}K_1 + x\alpha^2(y-\rho)K_2$$

with,

$$K_1 = \frac{\alpha^{-(N+1)}(y_1^N - y_2^N)}{y_2^{N+1}(y_1-\rho) - y_1^{N+1}(y_2-\rho)}$$

$$K_2 = \frac{1}{(y_1-\rho)(y_2-\rho)}\left[-1 + y_1^N + \frac{\rho y_1^{N+1}(y_2-\rho)(y_1^N - y_2^N)}{y_2^{N+1}(y_1-\rho) - y_1^{N+1}(y_2-\rho)}\right]$$

where, y_1 and y_2 are the roots of the equation $(1-\alpha y)\alpha y - x\rho\alpha^2 = 0$, i.e.,

$$y_{1,2}(x) = \frac{1+\rho \pm \sqrt{(1+\rho)^2 - 4\rho x}}{2}$$

Also $y_1 + y_2 = 1+\rho$ and $y_1 y_2 = \rho x$. We will now represent $\bar{S}(x,y)$ as $\sum_{i=0}^{N} c_i(x)y^i$. Observe that,

$$\bar{S}(x,y) = G_1(y) - K_3 G_2(y) + K_4 G_3(y) - K_5 G_4(y)$$

where, $G_1(y) = \frac{1-y^N}{1-y}$, $G_2(y) = \frac{(1-y^N)}{(1-y)(y-y_1)(y-y_2)}$, $G_4(y) = \frac{(y-\rho)}{(y-y_1)(y-y_2)}$, $G_3(y) = \frac{y^{N+1}}{(y-y_1)(y-y_2)}$ and $K_3 = -x\rho$, $K_4 = x\rho\alpha^{N+1}K_1$ and $K_5 = -xK_2$. We shall now apply the partial fraction method and express the right-hand side of the last equation in the form of $\frac{y^k - a^k}{(y-a)}$ for some k and a. Thus we can write,

$$\begin{aligned}\bar{S}(x,y) &= (1+K_3A_1)\frac{1-y^N}{1-y} - K_3A_2\frac{y_1^N - y^N}{y_1 - y} - K_3A_3\frac{y_2^N - y^N}{y_2 - y} + K_4B_1\frac{y_1^{N+1} - y^{N+1}}{y_1 - y} \\ &\quad + K_4B_2\frac{y_2^{N+1} - y^{N+1}}{y_2 - y} + \left(K_3A_2(1-y_1^N) + K_4B_1y_1^{N+1} + K_5C_1\right)\frac{1}{y-y_1} \\ &\quad + \left(K_3A_3(1-y_2^N) + K_4B_2y_2^{N+1} + K_5C_2\right)\frac{1}{y-y_2}\end{aligned}$$

where, $A_1 = \frac{1}{(1-y_1)(1-y_2)}$, $A_2 = \frac{1}{(1-y_1)(y_1-y_2)}$, $A_3 = \frac{1}{(1-y_2)(y_2-y_1)}$, $B_1 = \frac{1}{y_1-y_2}$, $B_2 = \frac{1}{y_2-y_1}$, $C_1 = \frac{y_1-\rho}{y_1-y_2}$ and $C_2 = \frac{y_2-\rho}{y_2-y_1}$. But,

$$K_3A_2(1-y_1^N) + K_4B_1y_1^{N+1+K_5C_1} = K_3A_3(1-y_2^N) + K_4B_2y_2^{N+1} + K_5C_2 = 0$$

This is because $\bar{S}(x,y)$ is analytic in y, the left-hand side of Equation (11) vanishes at $y = y_i, i = 1, 2..$ Hence, the above equation can be written as

$$\begin{aligned}\bar{S}(x,y) &= (1+K_3A_1)\frac{1-y^N}{1-y} - K_3A_2\frac{y_1^N - y^N}{y_1 - y} - K_3A_3\frac{y_2^N - y^N}{y_2 - y} + K_4B_1\frac{y_1^{N+1} - y^{N+1}}{y_1 - y} \\ &\quad + K_4B_2\frac{y_2^{N+1} - y^{N+1}}{y_2 - y}\end{aligned}$$

Again, recalling that

$$\frac{a^k - y^k}{a - y} = a^{k-1} + a^{k-2}y + a^{k-3}y^2 + \ldots + ay^{k-2} + y^{k-1}$$

and grouping the coefficients of the same power of y we get, $\bar{S}(x,y) = \sum_{i=0}^{N} c_i(x)y^i$.

Having expressed $\bar{S}(x,y)$ as $\sum_{i=1}^{N} c_i(x)y^i$ (in Proposition 2) we now proceed to obtain the expression for $\mathcal{G}$ using $\bar{S}(x,y)$.

Proposition 3 *The Goodput ratio, $\mathcal{G}$ can be written as,*

$$\mathcal{G} = q^2\sum_{i=0}^{N}\left(\frac{d(xc_i(x))}{dx}\right)_{x=(1-q)} P(Q=i) = q^2\left[\frac{d}{dx}\left(\sum_{i=0}^{N} xc_i(x)P(Q=i)\right)\right]_{x=(1-q)}$$

Proof:
We know by Equation (10),

$$\begin{aligned}\mathcal{G} &= q\sum_{n=1}^{\infty} nq(1-q)^{n-1}\sum_{i=0}^{N} S_{n,i}P(\mathcal{Q}=i) \\ &= q^2\sum_{i=0}^{N}\sum_{n=1}^{\infty} S_{n,i}n(1-q)^{n-1}P(\mathcal{Q}=i)\end{aligned}$$

Also,

$$\bar{S}(x,y) = \sum_{i=0}^{N} y^i \sum_{n=1}^{\infty} S_{n,i}x^{n-1} = \sum_{i=0}^{N} c_i(x)y^i$$

Thus, $c_i(x) = \sum_{n=1}^{\infty} S_{n,i}x^{n-1}$ and

$$\left(\frac{d(xc_i(x))}{dx}\right)_{x=(1-q)} = \sum_{n=1}^{\infty} nS_{n,i}(1-q)^{n-1}$$

Thus,

$$
\begin{aligned}
\mathcal{G} &= q^2 \sum_{i=0}^{N} \sum_{n=1}^{\infty} S_{n,i} n (1-q)^{n-1} P(\mathcal{Q}=i) \\
&= q^2 \sum_{i=0}^{N} \left(\frac{d(xc_i(x))}{dx} \right)_{x=(1-q)} P(Q=i)
\end{aligned}
$$

Thus, we can obtain the exact expression for the *Goodput ratio* by knowing the coefficients $c_i(x)$ and $P(Q=i)(=P_{i,0}+P_{i,1})$, for $0 \le i \le N$ (both being previously obtained in Corollary 1 and Proposition 2 respectively). Since the derivation as well as the final result are complex, we defer these to the appendix.

2.5 Numerical Examples

We now work with numerical examples and plot the $\mathcal{G}$ obtained with our explicit formula from Equation (27) (in appendix) with increasing load ρ. We first keep N fixed at 20 and plot for $1/q = 5, 10, 15, 20, 25, 30$ with ρ varying from 0.1 to 3.0 (in steps of 0.01) in Figure 2. Next we keep $1/q$ fixed at 20 and plot for $N = 5, 10, 15, 20, 25, 30$, again with ρ varying from 0.1 to 3.0 (in steps of 0.01) in Figure 3.

We have observed in Subsection 2.3 that there exists a nontrivial limiting behavior as N becomes large, while keeping all other parameters the same. This observation is supported by Figure 3 which shows that there is a limit of $\mathcal{G}$ as N tends to infinity.

So far we have done packet-level performance evaluation of the PMD policy for a M/M/1/N queue model. The explicit expression for goodput is somewhat complex. In the next section we propose a fluid model for analysing the PMD policy towards obtaining simple approximations for the goodput. The fluid model can be seen as a weak limit of the original packet model through a standard scaling argument.

3 Fluid Approximation

3.1 The workload process

Our fluid source always has messages to send and the capacity of the fluid buffer is finite, say B. The fluid buffer is served with a capacity c. The length of a message is assumed to be exponentially distributed with parameter η. If during the arrival of a message, the workload process $V(t)$ (alternatively the queue length, i.e., amount of fluid in the fluid buffer) reaches

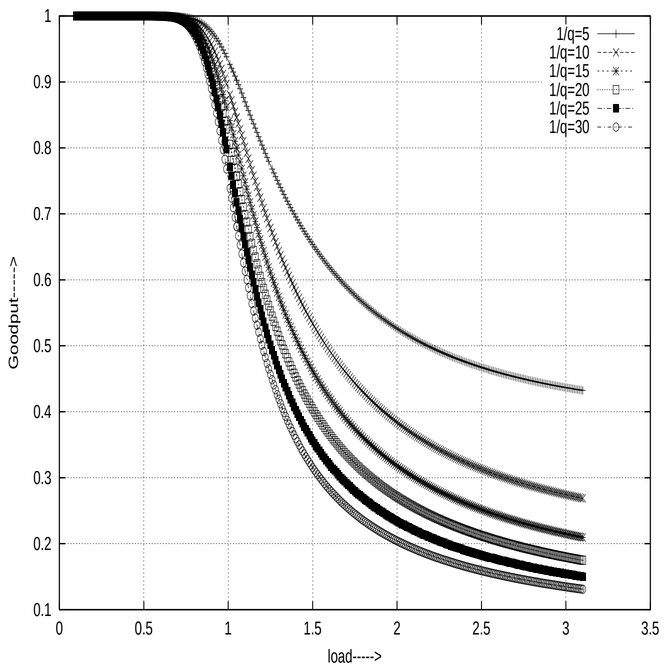

Figure 2: $\mathcal{G}$ Vs ρ for $1/q = 5, 10, 15, 20, 25, 30$ with $N = 20$

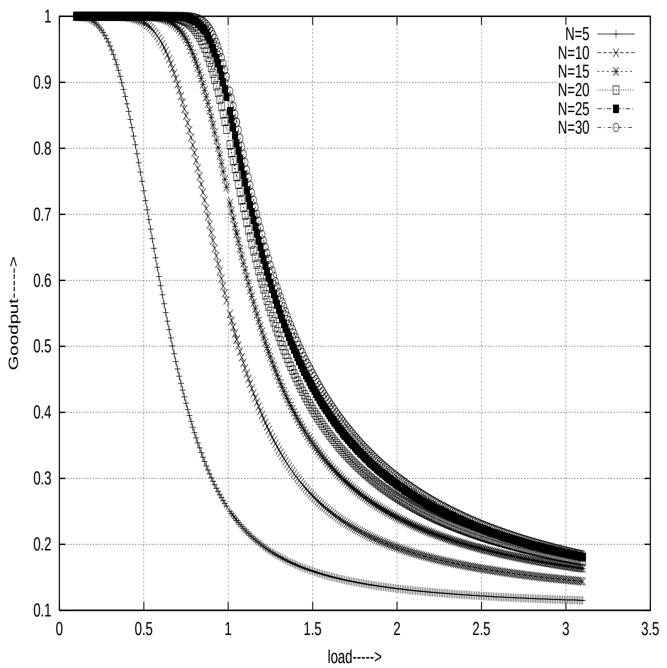

Figure 3: $\mathcal{G}$ Vs ρ for $N = 5, 10, 15, 20, 25, 30$ with $1/q = 20$

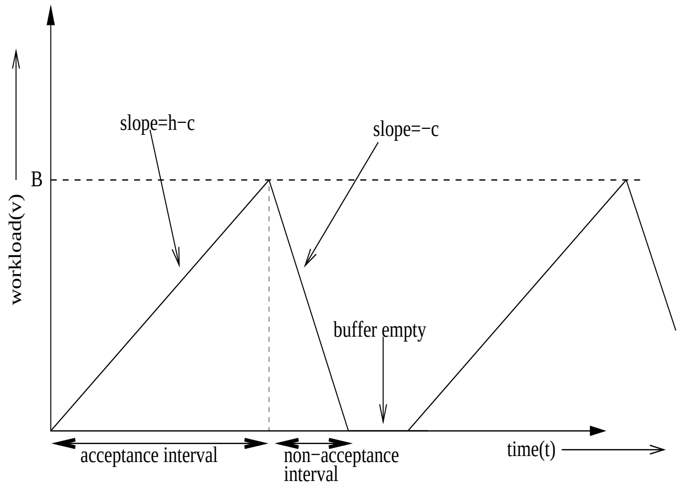

Figure 4: A typical evolution of the workload process, $V(t)$ in our fluid model

B, then all the remaining fluid corresponding to this message is dropped. Let the fluid arrival rate be h.

Remark 1 *The fluid limit can be seen as a weak limit of the original model through a standard scaling. More precisely, consider n models, and add n as a superscript to the parameters of the nth model. Then the scaling is obtained as follows.*

- *Arrival rate:* $\lambda^{(n)} := n\lambda$;
- *Service rate:* $\mu^{(n)} := n\mu$;
- *Size of messages: geometrically distributed with parameter* $q^{(n)} := q/n$;
- *Buffer size:* $N^{(n)} = nN$.

Let $X^{(n)}(t)$ be the queue length process of the nth model. Then, as $n \to \infty$, the process $X^{(n)}(t)/n$ weakly converges to our fluid process $V(t)$, with $h = \lambda$, $c = \mu$, and with $\eta = q\lambda$.

A typical evolution of $V(t)$ in our model is shown in Figure 4. Also, let A be the event that the incoming fluid is accepted. To remove trivialities we assume that $c < h$. Let V_n be the random variable denoting the queue length at the end of nth non-acceptance period. The dynamics of V_n can be written as,

$$V_{n+1} \quad = \quad V_n + (h-c)A_n - cA_n^c \tag{11}$$

where $A_n = (B - V_n)/(h - c)$ is the duration of the $n+1$st acceptance period and $A_n^c = \min(X_n, B/c)$ is the duration upto which the process $V(t)$ will have a negative slope where X_n is the remaining length of the current incoming message at the epoch when $V(t)$ hits B, i.e., at the end of the $n+1$th acceptance period and the start of the $n+1$th non-acceptance period. Let T_n be the epoch of the commencement of the $n+1$st acceptance event. Thus, $T_{n+1} = T_n + A_n + X_n$ and $(V_n, A_n + X_n)$ can be viewed as a marked point process [1]. Thus,

$$V_{n+1} = B - cA_n^c$$

Denote the steady state Laplace Stieltjes Transform (LST) of V_n, i.e., $E[e^{-s(B-cA_n^c)}]$ by $V_p(s)$ [5]. Let ρ_p be the probability density of V_n (the inverse of LST of V_n) in steady state.

Lemma 1 *The LST and the probability density of V_n in steady state are given by:*

$$V_p(s) = e^{-sB}\eta\left[\frac{1 - e^{-\frac{B}{c}(\eta - sc)}}{\eta - sc} + \frac{e^{sB}e^{\frac{-\eta B}{c}}}{\eta}\right]$$

$$\rho_p(v) = \frac{\eta}{c}e^{\frac{\eta}{c}(v-B)} + e^{-\frac{\eta B}{c}}\delta(v)$$

(for $v \in [0, B)$).

In the above Lemma we obtained the workload LST and probability density at the end of non-acceptance periods. Next we shall obtain these quantities at an arbitrary time, i.e. the LST and the probability density of the time stationary workload.

Proposition 4 *The LST $V(s)$ and the probability density ρ of $V(t)$ in stationary regime are given by:*

$$V(s) = \frac{\eta h}{(h - ce^{\frac{-B\eta}{c}})}e^{-sB}\left(\frac{1 - e^{-\frac{B(\eta - sc)}{c}}}{\eta - sc}\right) - \frac{h(1 - e^{\frac{-\eta B}{c}})}{(h - ce^{\frac{-B\eta}{c}})} + 1 \tag{12}$$

For $v \in [0, B)$,

$$\rho(v) = \rho^0(v) + \rho^1(v)$$

where, $\rho^0(v) = \left(1 - \frac{h(1-e^{\frac{-\eta B}{c}})}{(h-ce^{\frac{-B\eta}{c}})}\right)\delta(v)$, $\rho^1(v) = \frac{\eta h}{(h-ce^{\frac{-B\eta}{c}})}\left(\frac{e^{\frac{\eta}{c}(v-B)}}{c}\right)$
and $\rho(v) = 0$ for $v \geq B$. Finally, the mean stationary workload M_f is given by

$$M_f = \frac{h}{h - ce^{\frac{-B\eta}{c}}}\left(B - \frac{c}{\eta}(1 - e^{\frac{-B\eta}{c}})\right). \tag{13}$$

[5] The subscript p indicates the point process

Proof:
We shall now use the following inversion formula (see e.g., [1], Ch.1, Sec.4) to obtain the LST for the workload process $V(t)$ which we will then invert to get the probability density function.

$$E[e^{-sV(t)}] = \frac{E^0[\int_0^{T_1} e^{-sV(t)}]}{E^0[T_1]}$$

Thus,

$$\begin{aligned} V(s) &= \frac{E^0[\int_0^{\frac{B-V_0}{h-c}} e^{-s(V_0+(h-c)t)}dt}{E^0[\frac{B-V_0}{h-c}+X_0]} + \frac{\int_{\frac{B-V_0}{h-c}}^{\frac{B-V_0}{h-c}+A_0^c} e^{-s(B-c(t-\frac{B-V_0}{h-c}))}dt + X_0 - A_0^c]}{E^0[\frac{B-V_0}{h-c}+X_0]} \\ &= \frac{(h-c)\eta}{(h-ce^{\frac{-B\eta}{c}})}\left[\frac{E^0[e^{-sV_0}]-e^{-sB}}{s(h-c)} + \frac{e^{-sB}(E^0[e^{scA_0^c}]-1)}{sc} + E[X_0] - E[A_0^c]\right] \end{aligned}$$

Observe that we have expressions for $E[e^{-sV_0}]$ (i.e., $V_p(s)$), $E[V_0]$ (can be obtained from $V_p(s)$) and the expression for $E[e^{scA_0^c}]$ and hence for $E[A_0^c]$ can be easily obtained by the definition of A_0^c. Thus, we get after some calculations (12). The inverse of the LST of the last equation gives $\rho(v)$. Finally, M_f is obtained by the integration:

$$M_f = \int_0^B v\rho(v)dv = \int_0^B v\frac{\eta h}{(h-ce^{\frac{-B\eta}{c}})}\left(\frac{e^{\frac{\eta}{c}(v-B)}}{c}\right)dv$$

which implies (13).

3.2 The Goodput Ratio $\mathcal{G}_f$

We proceed with the model from the previous subsection and in particular, we continue to assume that $c < h$ in order to avoid trivialities. We define, the fluid analog of the *Goodput Ratio*, $\mathcal{G}_f$ as the ratio of the total fluid comprising good messages (i.e., messages which do not suffer any fluid loss due to buffer overflow) exiting the node to the total arriving fluid at its input. Let $\mathcal{V}_f$ be the random variable representing the success of a message, $\mathcal{V}_f = 1$ for a good message, and $\mathcal{V}_f = 0$ for a message which has lost some fluid. Let us define the sub-distribution function $F(w,1)$ as the probability that a message is of length $\leq w$ and is good, i.e, $F(w,1) = P(W \leq w, \mathcal{V}_f = 1)$. Then we can write the goodput ratio as

$$\mathcal{G}_f = \frac{\int_0^\infty wdF(w,1)}{\int_0^\infty wdF(w)}$$

where, $F(w)$ is the message length distribution ($\sim \exp \eta$). Again, writing $F(w,1)$ as,

$$\begin{aligned} F(w,1) &= P(\mathcal{V}_f = 1|W \leq w)P(W \leq w) \\ &= \int_0^B P(\mathcal{V}_f = 1|W \leq w, V = v)\rho(v)dv \int_0^w f(u)du \end{aligned}$$

where, $\rho(v)$ is the queue length density and V is the queue length at the epoch of the arrival of the message [6] and $f(x)$ is the message length density.

Proposition 5 *The Goodput is given by*

$$\mathcal{G}_f = \frac{c}{(h - ce^{\frac{-B\eta}{c}})}\left[e^{\frac{-B\eta h}{(h-c)c}}(1-\frac{c}{h}) + (\frac{c}{h} - e^{\frac{-B\eta}{c}})\right]$$

Proof:
Observe that, for $w \in [0, \frac{B-v}{h-c}]$,

$$P(\mathcal{V}_f = 1|W \leq w, V = v) = 1$$

and, for $w > \frac{B-v}{h-c}$,

$$P(\mathcal{V}_f = 1|W \leq w, V = v) = P(W < \frac{B-v}{h-c}|W < w)$$

Or in other words, for $w \in [0, \frac{B}{h-c}]$, if $v \in [0, B - w(h-c)]$,

$$P(\mathcal{V}_f = 1|W \leq w, V = v) = 1$$

else,

$$P(\mathcal{V}_f = 1|W \leq w, V = v) = \frac{P(W < \frac{B-v}{h-c})}{P(W < w)}$$

And for $w > \frac{B}{h-c}$,

$$P(\mathcal{V}_f = 1|W \leq w, V = v) = \frac{P(W < \frac{B-v}{h-c})}{P(W < w)}$$

[6] due to PASTA the queue length distribution at the arrival epochs of messages, which come as a Poisson stream, is same as the stationary queue length distribution

Thus we write, for $w \in [0, \frac{B}{h-c}]$, $F(w,1) = F_1(w) + F_2(w)$, where,

$$
\begin{aligned}
F_1(w) &= (1-e^{-\eta w}) \int_0^{B-w(h-c)} \rho(v)dv \\
&= (1-e^{-\eta w}) \left[1 - \frac{h(1-e^{\frac{-w\eta(h-c)}{c}})}{(h-ce^{\frac{-B\eta}{c}})}\right]
\end{aligned}
$$

Further,

$$
\begin{aligned}
F_2(w) &= \int_{B-w(h-c)^+}^{B} (1-e^{-\frac{B-v}{h-c}\eta})\rho^1(v)dv \\
&= \frac{h}{(h-ce^{\frac{-B\eta}{c}})} \left[(1-e^{\frac{-\eta w(h-c)}{c}}) - \frac{(h-c)}{h}(1-e^{\frac{-\eta h w}{c}})\right]
\end{aligned}
$$

and for $w > B/(h-c)$, $F(w) = \int_0^B (1-e^{-\frac{(B-v)\eta}{h-c}})\rho(v)dv$. Thus we get

$$
\begin{aligned}
dF_1(w) &= \frac{\eta h}{(h-ce^{\frac{-B\eta}{c}})} \left((e^{\frac{-w\eta h}{c}} - \frac{c}{h}e^{-\frac{\eta B}{c}}e^{-\eta w}) + (1-\frac{h}{c})(e^{\frac{-w\eta(h-c)}{c}} - e^{\frac{-w\eta h}{c}})\right) dw \\
dF_2(w) &= \frac{h\eta(h-c)}{c(h-ce^{\frac{-B\eta}{c}})} \left[e^{\frac{-\eta(h-c)w}{c}} - e^{\frac{-\eta h w}{c}}\right] dw
\end{aligned}
$$

Thus, for $w \in [0, \frac{B}{h-c})$,

$$
\begin{aligned}
dF(w,1)) &= dF_1(w) + dF_2(w) \\
&= \frac{\eta h}{(h-ce^{\frac{-B\eta}{c}})} \left((e^{\frac{-w\eta h}{c}} - \frac{c}{h}e^{-\frac{\eta B}{c}}e^{-\eta w})\right) dw
\end{aligned}
$$

and for $w > \frac{B}{h-c}$, $dF(w,1) = 0$. Hence we obtain

$$
\mathcal{G}_f = \frac{\eta^2 h}{(h-ce^{\frac{-B\eta}{c}})} \int_0^{\frac{B}{h-c}} w(e^{\frac{-w\eta h}{c}} - \frac{c}{h}e^{-\frac{\eta B}{c}}e^{-\eta w})dw
$$

from which we obtain Equation (14).

Let us now observe the behavior of $\mathcal{G}_f$ for extreme values of η, keeping all other parameters fixed. As η tends to zero we see from expression (14) that $\mathcal{G}_f$ tends to zero. This can be explained by the fact that small η corresponds to very long frames, so that the probability that the queue will fill during the arrival of a message tends to one (since $h > c$).

For the other extreme, i.e. $\eta \to \infty$, the length of a message is very short; one could then expect that the Goodput would be equal to the relative amount of fluid that is lost, since a message corresponds to an infinitesimal amount of fluid. This would give a Goodput of c/h. This is however not the real limiting value of the Goodput: we see, in fact, that as $\eta \to \infty$, we get $\mathcal{G}_f \to c^2/h^2$ from expression (14). The reason that one could expect to have a goodput of c/h is that this indeed is the fraction of fluid that could be served. So this could give an expression for throughput. But even for a huge buffer, this does not take into account the fact that part of the fluid that is already in the queue corresponds to *bad* packets: they belong to messages in which some packets are dropped. In fact all the queued fluid of a message that arrives when the amount of fluid hits the boundary is lost. We next provide an intuitive argument through an example that may justify this limiting behavior.

Consider $\eta = 1, h = 10$ and $c = 6$. We will show the approximate limit achievable for an expected behavior of our queue and see that it is close to c^2/h^2. We look at the case when a message is discarded, then the expected length of the subsequent discarding period will be $1(= 1/\eta)$ (the expected remaining length of a $exp(1)$ distributed random variable). During this period there will be an approximate [7] expected reduction of $6(= c/\eta)$ in the fluid level. Then a new message, call it message a of expected length 1 starts arriving. The expected queue length at the end of the arrival of this message will approximately $B-6+(h-c) = B-2$. However the expected amount of good fluid that was injected in the queue by this message is approximately 10. Then message b starts arriving, whose expected remaining length is again 1. Thus had message b been completely accepted the expected amount of fluid injected into the queue would have been 10, but because of buffer overflow the expected amount of fluid that can be accepted is approximately $\frac{h\times 2}{(h-c)} = 5$ units. Then the next discarding period starts , whose expected length will again be 1. And the expected amount of fluid that will be discarded in this discarding period will be approximately 10. Thus we have,

$$G_f = \frac{10 + 10 + \ldots}{10 + 5 + 10 + 10 + 5 + 10 + \ldots}$$

which gives,

$$G_f = \frac{10}{25} = 0.4 \approx c^2/h^2 = 0.36$$

Another interesting observation from the expression for G_f is that *the dependence of G_f on different parameters is only through two ratios, c/h and $B\eta/c$.* In particular *G_f is dependent*

[7] Note that the reduction cannot be greater than B units of fluid, hence in fact the reduction is $\simeq min(B, Xc)$, where $X \simeq exp(\eta)$. The use of word *approximate/approximately* in subsequent discussion is to highlight the fact that we are approximating a restricted exponential distribution as an exponential distribution

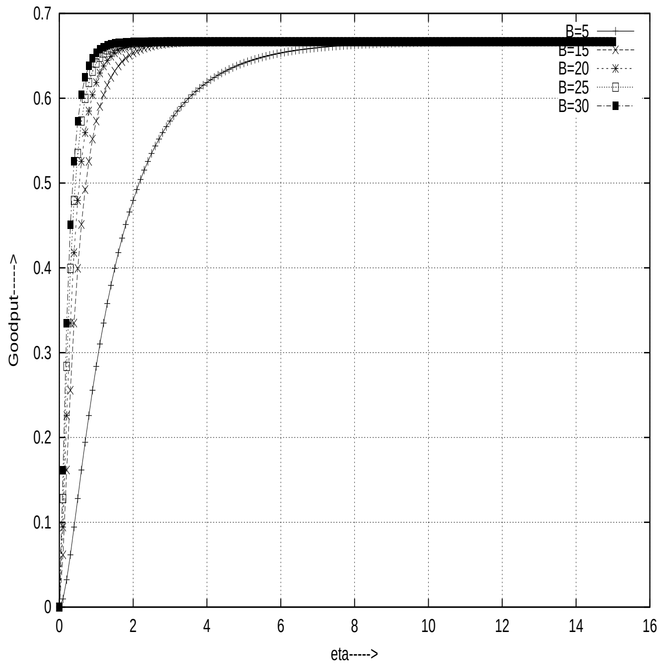

Figure 5: $\mathcal{G}_f$ Vs η for $\eta \in [0, 15]$ for different values of B with $c = 8$ and $h = 12$

on η and B only through their product. Also, observe that as B tends to 0, $\mathcal{G}_f$ tends to 0 and when B tends to ∞, $\mathcal{G}_f$ tends to c^2/h^2.

3.3 Numerical Examples

We now work with some numerical examples and show that behavior of $\mathcal{G}_f$ as we increase η. We take $c = 8$, $h = 12$ and observe the behavior of $\mathcal{G}_f$ for $B = 5, 15, 20, 25, 30$ as η increases from 0 to 15 in Figure 5. The limiting behavior ($\lim_{\eta \to \infty} \mathcal{G}_f = c^2/h^2 = 0.6667$) as η increases is seen even at low values of η as B increases.

4 Conclusion

We have provided explicit expressions for the queue size distribution and for the goodput for the packet model based on recursions introduced in [11]. We then provided an alternative fluid approximation for studying the PMD policy and obtained the queue size distribution and goodput in this framework of fluid queue. Our analytical results will be quite useful in dimensioning the buffer size that should be used for a given required goodput under the PMD policy. Also the explicit expressions will be helpful in analytically studying the sensitivity of the goodput to various parameters. We are currently studying the potentials of the fluid model as an alternative to the packet model. Our ongoing work include generalizing the fluid model to include Markovian fluids and experimental validations of our models/findings.

5 Appendix

5.1 Proof of Proposition 1

From Equation (3) and Equation (5) by applying *z-transform* we get

$$Q_0(z)[(\rho+1)-\rho z-z^{-1}]-q\rho z Q_1(z) = P_{0,0}[1+\rho-z^{-1}]-P_{1,0}-\rho z^{N+1}[P_{N,0}+qP_{N,1}] \quad (14)$$

Similarly, by applying *z-transform* to Equation (4), we get

$$Q_1(z)=P_{0,1}\frac{(1-z-z^{N+1}(q\rho+1)^N q\rho)}{1-z(q\rho+1)} \quad (15)$$

Now from Equations (1) (2) (6) and Equation (14), we get on solving for $Q_0(z)$,

$$Q_0(z)=\frac{P_{0,0}[1-z^{-1}]-\rho z^{N+1}P_{N,0}[1+\rho q]+Q_1(z)q\rho z}{[(\rho+1)-\rho z-z^{-1}]} \quad (16)$$

The zeros of the denominator of Equation (16) are $z=1,\rho^{-1}$. At these values of z, the numerator of Equation (16) should be equal to 0 because of the analyticity of $Q_0(z)$ (being a polynomial in z of degree less than or equal to N). Thus, substituting $z=1$ in the numerator of Equation (16) and equating it to 0 we get an equation

$$P_{N,0}=(1+q\rho)^{N-1}qP_{0,1} \quad (17)$$

Also, substituting $z=\rho^{-1}$ in the numerator of Equation (16) and equating it to 0 we get another equation

$$P_{0,0}(1-\rho)-\rho^{-N}P_{N,0}(1+\rho q)+Q_1(\rho^{-1})q=0 \quad (18)$$

From Equations (15), (17) and (18), we get,

$$P_{0,0}=P_{0,1}\frac{q(1-(p^{-1}+q)^N)}{[p(1-q)-1]} \quad (19)$$

We shall interpret Equation (7) as

$$Q_0(1) + Q_1(1) = 1 \tag{20}$$

From Equation (15)

$$Q_1(1) = (q\rho + 1)^N P_{0,1} \tag{21}$$

and

$$\dot{Q}_1(1) = \frac{P_{0,1}\left[1 + (\rho q + 1)^N (N\rho q - 1)\right]}{q\rho} \tag{22}$$

From Equation (16) differentiating the numerator and denominator and taking limit as $z \to 1$, we get

$$Q_0(1) = \lim_{z\to 1} Q_0(z) = \lim_{z\to 1} \frac{z^{-2}P_{0,0} - \rho(N+1)z^N P_{N,0}(1+\rho q) + q\rho(z\dot{Q}_1(z) + Q_1(z))}{(-\rho + z^{-2})} \tag{23}$$

Thus from Equations (21) (22) (17) (19) and Equation (23) we get

$$Q_0(1) = \frac{P_{0,1}}{(1-\rho)}\left[\frac{q(1-(p^{-1}+q)^N)}{[p(1-q)-1]} + 1 - (1+\rho q)^N\right] \tag{24}$$

Substituting Equation (24) and (21) in Equation (20), and solving for $P_{0,1}$, we get

$$P_{0,1} \quad = \quad \frac{(1-\rho)(\rho(1-q)-1)}{q(1-\rho^{-N}(1+\rho q)^N) + (\rho(1-q)-1)(1-\rho(1+\rho q)^N)} \tag{25}$$

Knowing $P_{0,1}$ we have obtained the generating functions $Q_0(z)$ and $Q_1(z)$. However, we can further modify the expression for $Q_1(z)$ to a more meaningful form. From Equation (16) we write, after some algebraic manipulations,

$$Q_0(z) \quad = \quad P_{0,1}q\Big[(1+\rho q)^{N-1}\left(\frac{z^{N+1}-\rho^{-(N+1)}}{z-\rho^{-1}}\right) + (1+\rho q)^{N-2}\left(\frac{z^N-\rho^{-N}}{z-\rho^{-1}}\right) + \ldots + \frac{z^2-\rho^{-2}}{z-\rho^{-1}}\Big]$$

Observe that, each fraction inside the bracket on the right hand side of the last equation is of the form $(x^k - a^k)/(x-a)$ which simplifies to

$$\frac{x^k - a^k}{x-a} = x^{k-1} + x^{k-1}a + x^{k-2}a^2 + \ldots + xa^{k-2} + a^{k-1}$$

thus we get,

$$Q_0(z) \quad = \quad P_{0,1}q\rho\Big[(1+\rho q)^{N-1}z^{N+1}\sum_{j=1}^{N+1}\left(\frac{1}{\rho z}\right)^j + (1+\rho q)^{N-2}z^N\sum_{j=1}^{N}\left(\frac{1}{\rho z}\right)^j + \ldots + z^2\sum_{j=1}^{2}\left(\frac{1}{\rho z}\right)^j\Big]$$

Grouping the coefficients of the powers of z we get (8).

5.2 Exact Expression for $\mathcal{G}$

We shall first obtain an expression for $\sum_{i=0}^{N} c_i P(Q=i)$.

$$\sum_{i=0}^{N} c_i(x)P(Q=i) = c_0(x)P(Q=0) + \sum_{i=1}^{N-1} c_i(x)P(Q=i)$$

$$\begin{aligned}
&= (1+K_3A_1) + (c_0 - (1+K_3A_1))P(Q=0) + [K_4B_1y_1^N - K_3A_2y_1^{N-1}]\sum_{i=1}^{N-1} P(Q=i)y_1^{-i} \\
&\quad +[K_4B_2y_2^N - K_3A_3y_2^{N-1}]\sum_{i=1}^{N-1} P(Q=i)y_2^{-i} - (1+K_3A_1)P(Q=N)
\end{aligned} \tag{26}$$

Observe that,

$$\begin{aligned}
\sum_{i=1}^{N-1} P(Q=i)y_1^{-i} &= Q_0(y_1^{-1}) + Q_1(y_1^{-1}) - P(Q=0) - P(Q=N)y_1^{-N} \\
\sum_{i=1}^{N-1} P(Q=i)y_2^{-i} &= Q_0(y_2^{-1}) + Q_1(y_2^{-1}) - P(Q=0) - P(Q=N)y_2^{-N}
\end{aligned}$$

Writing $Q(z) = Q_1(z) + Q_2(z)$, from the above Equation the expression for $\sum_{i=1}^{N} c_i P(Q=i)$ simplifies to

$$\begin{aligned}
\sum_{i=0}^{N} c_i(x)P(Q=i) &= (1+K_3A_1)(1-P(Q=N)) + [K_4B_1y_1^N - K_3A_2y_1^{N-1}]Q(y_1^{-1}) \\
&\quad +[K_4B_2y_2^N - K_3A_3y_2^{N-1}]Q(y_2^{-1}) + K_3(A_2y_1^{-1} + A_3y_2^{-1})P(Q=N)
\end{aligned}$$

And by Proposition 3 we write,

$$\mathcal{G} = q^2\left[(1-q)\left(\frac{d}{dx}\left(\sum_{i=0}^{N} c_i(x)P(Q=i)\right)\right)_{x=(1-q)} + \sum_{i=0}^{N} c_i(1-q)P(Q=i)\right]$$

Thus we need to evaluate $\frac{d}{dx}\left(\sum_{i=0}^{N} c_i(x)P(Q=i)\right)$. From the expression for $\sum_{i=0}^{N} c_i(x)P(Q=i)$ from Equation (26), we write

$$\frac{d}{dx}\left(\sum_{i=0}^{N} c_iP(Q=i)\right)$$

$$
\begin{aligned}
= \quad & \frac{d}{dx}K_3A_1(1-P(Q=N)) + [K_4B_1y_1^N - K_3A_2y_1^{N-1}]\frac{d}{dx}Q(y_1^{-1}) \\
& + Q(y_1^{-1})\frac{d}{dx}[K_4B_1y_1^N - K_3A_2y_1^{N-1}] + [K_4B_2y_2^N - K_3A_3y_2^{N-1}]\frac{d}{dx}Q(y_2^{-1}) \\
& + Q(y_2^{-1})\frac{d}{dx}[K_4B_2y_2^N - K_3A_3y_2^{N-1}] \\
& + \frac{d}{dx}\left(K_3(A_2y_1^{-1} + A_3y_2^{-1})\right)P(Q=N)
\end{aligned}
$$

Thus, we need to obtain the derivative terms on the right side of the last equation. We have obtained these terms. The final expressions are provided here.

$$
\begin{aligned}
\frac{d}{dx}K_3A_1 &= \frac{1}{q^2} \\
(K_4B_1y_1^N - K_3A_2y_1^{N-1}) &= \frac{y_1^{N-1}x\rho}{(1-y_1)\phi_{N+1}\delta}\left(\rho\phi_N - \delta_N(1-y_1)\right) \\
(K_4B_2y_2^N - K_3A_2y_2^{N-1}) &= -\frac{y_2^{N-1}x\rho}{(1-y_2)\phi_{N+1}\delta}\left(\rho\phi_N - \delta_N(1-y_2)\right)
\end{aligned}
$$

$$
\begin{aligned}
& \frac{d}{dx}\left(K_4B_1y_1^N - K_3A_2y_1^{N-1}\right) \\
& \quad = \quad \frac{y_1^{N-1}x\rho}{(1-y_1)\phi_{N+1}\delta}\left[\rho\left(x(N-1)\beta_{N-2}\frac{dy_1}{dx} + \delta_{N-1} - N\beta_{N-1}\frac{dy_1}{dx}\right)\right. \\
& \qquad -(1-y_1)N\beta_{N-1}\frac{dy_1}{dx} + \delta_N\frac{dy_1}{dx} + (\rho\phi_N - \delta_N(1-y_1)) \times \\
& \qquad \left.\left(\frac{1}{x} + \frac{1}{(1-y_1)}\frac{dy_1}{dx} + \frac{(N-1)}{y_1}\frac{dy_1}{dx} - \frac{1}{\delta}2\frac{dy_1}{dx} - \frac{1}{\phi_{N+1}}\frac{d\phi_{N+1}}{dx}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
& \frac{d}{dx}\left(K_4B_2y_2^N - K_3A_3y_2^{N-1}\right) \\
& \quad = \quad -\frac{y_2^{N-1}x\rho}{(1-y_2)\phi_N\delta}\left[\rho\left(x(N-1)\beta_{N-2}\frac{dy_2}{dx} + \delta_{N-1} - N\beta_{N-1}\frac{dy_2}{dx}\right)\right. \\
& \qquad -(1-y_2)N\beta_{N-1}\frac{dy_2}{dx} + \delta_N\frac{dy_2}{dx} + (\rho\phi_N - \delta_N(1-y_2)) \times \\
& \qquad \left.\left(\frac{1}{x} + \frac{1}{(1-y_2)}\frac{dy_2}{dx} + \frac{(N-1)}{y_2}\frac{dy_2}{dx} - \frac{1}{\delta}2\frac{dy_2}{dx} - \frac{1}{\phi_{N+1}}\frac{d\phi_{N+1}}{dx}\right)\right]
\end{aligned}
$$

$$Q(y) = (P_{0,0}+P_{0,1}) + \frac{qP_{0,1}}{(1+q\rho)(1-\rho^{-1}(1+\rho q))}\left[\rho(1-q)(1+\rho q)y \right.$$
$$\left. \frac{1-(1+\rho q)^N y^N}{1-(1+\rho q)y} - \frac{(1+\rho q)^{N+1}}{\rho^N} y \frac{1-(y\rho)^N}{1-(y\rho)}\right]$$

and finally,

$$\frac{dy_1}{dx} = \frac{-dy_2}{dx} = \frac{\rho}{\sqrt{(1+\rho)^2 - 4\rho x}}$$

Thus, having obtained all the terms in Equation (27) we have the explicit expression for $\mathcal{G}$.

References

[1] F. Baccelli and Pierre Bremaud. *Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurrence.* Springer-Verlag, 1994.

[2] O.J. Boxma. Sojourn Times in Cyclic Queues - the Influence of the Slowest Server. In G. Iazeolla, P. J. Courtois, and O. J. Boxma, editors, *Computer Performance and Reliability*, pages 84–88. Elsevier Science Publishers B.V. (North-Holland), 1988.

[3] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1(4):25–39, 1993.

[4] S. Floyd and A. Romanow. Dynamics of TCP Traffic over ATM networks. *in Proc. of ACM SIGCOMM'94*, pages 79–88, September 1994.

[5] S. W. Fuhrmann and R. B. Cooper. Stochastic Decomposition in the M/G/1 Queue with General Vacations. *Operations Research*, 33:1117–1129, 1985.

[6] J.C.Bolot. End-to-end Delay and Loss Behavior in the Internet. *In Proc. of ACM Sigcomm '93*, pages 289–298, Sept. 1993.

[7] J.C.Bolot. Analysis of Audio Packet loss in the Internet. *NOSSDAV*, 1995.

[8] Y. H. Kim and S. Q. Li. Performance Analysis of Data Packet Discarding in ATM Networks. *IEEE/ACM Transactions on Networking*, 7(2):216–227, 1999.

[9] D. Kofman and M. Gagnaire. Réseaux Haut Débit: réseaux ATM, réseaux locaux et réseaux tout-optiques. InterEditions (groupe Masson), May 1996.

[10] Omar Ait-Hellal Eitan Altman Alan Jean-Marie Irina A. Kurkova. On Loss Probabilities in Presence of Redundant Packets and Several Traffic Sources. *Performance Evaluation*, 36-37:485–518, 1999.

[11] Yael Lapid Raphael Rom and Moshe Sidi. Analysis of Discarding Policies in High-Speed Networks. *IEEE Journal on Selected Areas in Communications*, 16(5):764–777, June 1998.

[12] J. George Shanthikumar. On Stochastic Decomposition in M/G/1 Type Queues with Generalized Server Vacations. *Operations Research*, 36(4):566–569, July-August 1988.

[13] J. S. Turner. Maintaining high throughput during overload in ATM switches . *in Proc. of INFOCOM '96*, pages 287–295, April 1996.