Delay Analysis of Selective Repeat ARQ for a Markovian Source Over a Wireless Channel*

Jeong Geun Kim and Marwan Krunz

Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721
{jkkim, krunz}@ece.arizona.edu

Last Revised: September 8, 1999

Abstract

In this paper, we analyze the mean delay experienced by a Markovian source over a wireless channel with time-varying error characteristics. The wireless link implements the selective-repeat ARQ scheme for retransmission of erroneous packets. We obtain good approximations of the total delay, which consists of transport and resequencing delays. The transport delay, in turn, consists of queueing and transmission delays. In contrast to previous studies, our analysis accommodates *both* the inherent autocorrelations in the input traffic and the time-varying nature of channel conditions. The exact probability generating function (PGF) of the queue length under "ideal" SR ARQ is obtained and is combined with the retransmission delay to obtain the mean transport delay. For the resequencing delay, the analysis is performed under the assumptions of heavy traffic and small window sizes (relative to the channel sojourn times). The inaccuracy due to these assumptions is observed to be negligible. We show that ignoring the autocorrelations of the incoming traffic or the time-varying nature of the channel state can lead to significant underestimation of the delay performance, particularly at high channel error rates. Some interesting effects of key system parameters on the delay performance are observed.

*This work was supported by the National Science Foundation under Career Award ANI-9733143. An abridged version of this paper was presented at the *Second ACM Int. Workshop on Wireless Mobile Multimedia (WoWMoM '99), Aug. 1999.*

1 Introduction

Automatic repeat request (ARQ) protocols are used to provide reliable data transfer in wireless communications. In these protocols, the transmitter sends a packet that consists of payload bits and error detection code. The receiver checks the integrity of the packet by decoding the error detection code. Depending on the outcome of the decoder, a positive acknowledgment (ACK) or a negative acknowledgment (NACK) is sent back to the sender. The sender retransmits the packet upon the receipt of the NACK message, whereas it transmits a new packet if an ACK is received.

In general, ARQ protocols are variants of three basic schemes: stop-and-wait (SW), go-back-N (GBN), and selective-repeat (SR). In SW ARQ, the transmitter must receive the ACK of a packet before transmitting the next packet. This scheme preserves the order of packets but it results in low channel utilization if the feedback delay is large. In GBN ARQ, packets are transmitted continuously without waiting for ACKs/NACKs. If a NACK is received, the transmitter retransmits the negatively acknowledged packet and all subsequent packets regardless of their acknowledgments. In SR ARQ, packets are transmitted continuously as in GBN ARQ, but only negatively acknowledged packets are retransmitted. Of the three schemes, SR ARQ achieves the highest throughput. Note that as the round-trip time (RTT) of a packet goes to zero, SR ARQ and SW ARQ become identical; a situation that is referred to as *ideal* SR ARQ [1, 2].

In this study, we consider a wireless link protocol that provides sequential delivery of packets and that uses SR ARQ for error control. An example of this protocol is used in wireless asynchronous transfer mode (WATM) networks [3]. In such a protocol, the transmitter assigns each packet a unique identifier (see Fig. 1). Packets are transmitted according to their identifiers. A copy of each transmitted packet is temporarily kept in a "waiting buffer" until the ACK message of that packet is received. If a NACK is received, it triggers a retransmission of the erroneous packet. Once an ACK is received, the packet is removed from the waiting buffer and a new packet is transmitted. Although the transmitter sends packets in order, packets at the receiver may be out of sequence due to the random occurrence of packet transmission errors. Thus, correctly received packets with higher identifiers must wait in a buffer (called resequencing buffer) until packets with lower identifiers are correctly received.

Fig. 2 describes all the delay components that a packet undergoes when transported over a wireless link. The total delay consists of transport and resequencing delays. The transport delay is subdivided into queueing and transmission delays. The queueing delay is the duration from the time a packet arrives at the transmitting node until its first transmission attempt. The transmission delay is defined as the time from a packet's first transmission until its *successful* arrival at the receiver (i.e., it includes all retransmission delays). The resequencing delay is defined as the waiting time of the packet in the resequencing buffer.

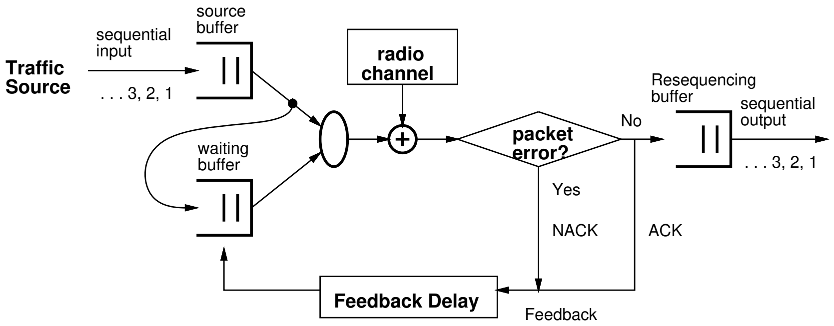

Figure 1: Packet transmission over a wireless link with in-sequence delivery.

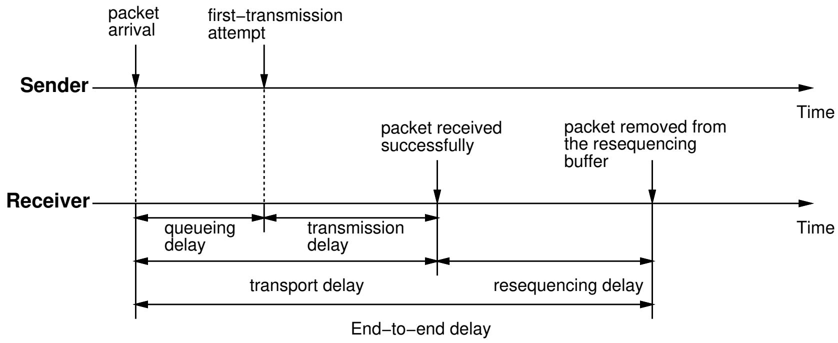

Figure 2: Time diagram for packet transmission process.

Related Work

Several previous studies have been conducted on the delay performance of SR ARQ over a wireless channel [4, 1, 5, 6, 7]. In [5] Konheim derived the probability generating functions for the transport delay and the queue length in GBN and SR ARQ protocols assuming a renewal traffic source and a "static" radio channel (i.e., errors are independent with a constant error rate). However, the complexity of his approach increases exponentially with the feedback delay. An alternative exact analysis of the delay performance was provided by Anagnostou and Protonotarios [4] for the special case of a Bernoulli renewal process. But similar to [5], the computational complexity of this analysis grows rapidly with the RTT and becomes quite prohibitive for moderate RTTs. In the same paper, the authors proposed another approach that is based on the *ideal* SR ARQ approximation, i.e., the dependence between the queueing process and the history of the transmission process is ignored. The ideal SR ARQ approximation was also used by Fantacci to analyze the delay performance under a Bernoulli arrival process and a time-varying radio channel [1]. Rosberg and Shacham analyzed the resequencing delay and buffer occupancy at the resequencing buffer assuming heavy-traffic conditions and a static radio channel [8]. Rosberg and Sidi [7] analyzed the joint distribution of buffer occupancy at the transmitter and receiver, and derived the mean transmission and resequencing delays assuming a renewal arrival process and a static channel. Schachum and Towsley investigated the buffer occupancy and resequencing delay in a wireless environment in which a single transmitter and multiple receivers communicate, assuming heavy-traffic conditions and a static channel [9].

Contributions and Paper Organization

The previous works were conducted using overly simplifying assumptions on the incoming traffic (e.g., Bernoulli arrival process), the channel errors (e.g., static channel model), or both. As shown in this paper, such simplifications can lead to significantly underestimating the delay performance. Accordingly, we provide refined analysis of the mean transport delay that accounts for traffic correlations and for the time-varying channel behavior. Correlations are known to exist in all types of network traffic, and their profound impact on the queueing performance has been widely reported in the context of wireline networks [10, 11, 12, 13, 14]. We show that such impact extends to the wireless environment, where the mean transport delay is seen to rise rapidly with traffic burstiness and with the channel error rate. In the wireless case, the delay is also impacted by the error behavior of the channel. More specifically, we show that the use of a time-varying error model can result in a notable increase in the mean transport delay. Relaxing the traditional assumptions on the source and channel models enables us to provide accurate predictions of the delay performance, which can be used in efficient resource allocation and admission control subject to guaranteed delay performance.

We capture the time-varying nature of the channel via a two-state Gilbert-Elliot (GE) model. To account for traffic correlations, we represent the arrival process by an N-state Markov process, where $N \geq 2$. Of particular interest here is the case of $N = 2$ as it represents the common on-off behavior of network traffic. Markovian processes have been extensively used in wired networks to characterize various types of traffic (see [15] for details). While other, non-Markovian classes of models, including self-similar (e.g., [16, 12, 17]) and subexponential models (e.g., [18]), have also been proposed as a means of capturing the persistent correlations in network traffic, the jury is still out on whether such models provide better traffic characterization than conventional Markovian models [19, 20, 21]. In fact, resource allocation in wired networks is still being carried out under the assumption of Markovian models. Given the popularity, tractability, and widespread use of Markov models in teletraffic studies, we made them the basis for our analysis.

In order to obtain the mean queueing delay, we simplify the analysis by eliminating the dependence between the queueing process and the history of the packet transmission process. This so-called ideal SR ARQ approximation, which was also used in [1, 4] under a Bernoulli arrival process, becomes exact as the feedback delay approaches zero. We derive the exact probability generating function (PGF) for the ideal SR ARQ and obtain the mean queueing delay using Little's law. The mean transmission delay is easily obtained since it only depends on the channel parameters and the RTT. We also derive an approximate expression for the mean resequencing delay. The goodness of our approximate results is verified by contrasting them against more realistic simulation results. Through numerical examples, we study the effects of key system parameters on the various delay components, providing insight on some of their interesting interactions.

The rest of the paper is organized as follows. In Section 2 we analyze the queueing and transmission delays. The mean resequencing delay is studied in Section 3. Numerical results and discussion are given in Section 4, followed by concluding remarks in Section 5.

2 Queueing and Transmission Delays

In this section, we analyze the queueing and transmission delays over a wireless link assuming a Markovian source and a time-varying wireless channel. Consider the queueing system at the transmitter side of the link. The arrival process is N-state Markovian that is governed by a transition probability matrix $\boldsymbol{P}$, where at each state i, $i = 0, \cdots, N$, i packets are generated in one time slot. Our model is based on an embedded Markov chain in which the number of packets in the queue is observed at the beginning of each time slot, just before the arrival of a new packet or of an ACK/NACK message. We assume ACK/NACK messages are always error free. This common assumption is justified by the fact that these important messages are often protected by FEC [1]. The wireless channel is characterized by the GE model, in which the channel alternates between *Good* and *Bad* states with corresponding bit error probabilities P_{eg} and P_{eb}, respectively (see Fig. 3). It is assumed that state transitions occur at the end of time slots, where a time slot corresponds to a packet transmission time. Since this time is very small compared to the sojourn time of a channel state, the inaccuracy due to our assumption is negligible. The packet error probabilities during Good and Bad channel states are given by:

$$e_0 = 1 - (1 - P_{eg})^L \quad (1)$$

$$e_1 = 1 - (1 - P_{eb})^L \quad (2)$$

for a packet size of L bits.

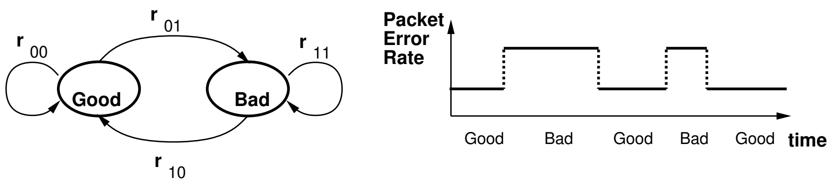

Figure 3: Wireless channel model.

At time t, the feedback message that arrives at the transmitter is for a packet that was transmitted at time $t - s$, where s is the feedback delay. However, for simplicity we assume that this feedback message has the same probabilistic nature as the feedback message associated with the packet to be transmitted at time t, i.e., as if the feedback delay is zero. This so-called ideal SR ARQ approximation was also used in [4, 1] to study the delay performance under a Bernoulli arrival process. Note that this approximation does not mean the feedback delay is ignored, but that its impact on the queueing process is not incorporated.

First, we derive the PGF for the queue length in the ideal SR ARQ case. We assume that packets are served on a FCFS basis and that the buffer capacity is infinite. Key notations are summarized as follows:

$a(k)$:	Number of new arrivals during the kth slot.
$r(k)$:	Channel state at the beginning of the kth slot.
$q_{i,j}(k)$:	Queue length at the beginning of the kth slot when the source is in state i and the channel is in state j.
$\boldsymbol{P}$:	Transition probability matrix for the arrival process at the transmitter.
$\boldsymbol{R}$:	Transition probability matrix for the process that describes the state of the radio channel.

The elements of $\boldsymbol{P} = [p_{i,j}]$ and $\boldsymbol{R} = [r_{i,j}]$ are defined as follows:

$$p_{i,j} \triangleq \Pr[a(k+1) = j \mid a(k) = i], \quad 0 \le i, j \le N \tag{3}$$

$$r_{i,j} \triangleq \Pr[r(k+1) = j \mid r(k) = i], \quad i, j \in \{0, 1\} \tag{4}$$

where states 0 and 1 denote Good and Bad channel states, respectively.

The size of the queue at the beginning of slot k is a function of its size at the previous slot, the number of new arrivals, and the state of the feedback message. Thus, the queue size at the beginning of the $(k+1)$th slot is obtained as follows: If $q_{\cdot,\cdot}(k) + a(k) > 0$, then

$$q_{l,j}(k+1) = \begin{cases} q_{i,j}(k) + i - 1, & \text{with probability } p_{i,l} \cdot (1 - e_j) \cdot r_{j,j} \\ q_{i,j}(k) + i, & \text{with probability } p_{i,l} \cdot e_j \cdot r_{j,j} \\ q_{i,i-j}(k) + i - 1, & \text{with probability } p_{i,l} \cdot (1 - e_{1-j}) \cdot r_{1-j,j} \\ q_{i,1-j}(k) + i, & \text{with probability } p_{i,l} \cdot e_{1-j} \cdot r_{1-j,j}. \end{cases} \tag{5}$$

And if $q_{\cdot,\cdot}(k) + a(k) = 0$, then

$$q_{l,j}(k+1) = 0, \quad \text{with probability } p_{i,l} \cdot (r_{j,j} + r_{1-j,j}) \tag{6}$$

where $0 \le i, l \le N$ and $0 \le j \le 1$. In (5), the last two cases correspond to the state of the radio channel going from $1-j$ to j, whereas no transition occurs in the other two cases. Furthermore, the first and third cases correspond to a successful packet transmission, whereas in the other cases the transmitted packet is in error. The steady state probability $q_{i,j}[n]$ is defined as:

$$q_{i,j}[n] \triangleq \lim_{k\to\infty} \Pr[q_{i,j}(k) = n]. \tag{7}$$

¿From (5) and (6), the state balance equation is obtained as follows:

If $n > 0$,

$$\begin{aligned} q_{i,j}[n] &= \sum_{l=0}^{\min(N,n+1)} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}[n-l+1] + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}[n-l+1]) \\ &\quad + \sum_{l=0}^{\min(N,n)} (r_{j,j} e_j p_{l,i} q_{l,j}[n-l] + r_{\bar{j},j} e_{\bar{j}} p_{l,i} q_{l,\bar{j}}[n-l]) \end{aligned} \tag{8}$$

where $\bar{x}$ denotes $1-x$.

And, if $n = 0$,

$$\begin{aligned} q_{i,j}[n] &= \sum_{l=0}^{\min(N,n+1)} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}[n-l+1] + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}[n-l+1]) \\ &\quad + p_{0,i}(r_{j,j} q_{0,j}[0] + r_{\bar{j},j} q_{0,\bar{j}}[0]). \end{aligned} \tag{9}$$

The following proposition gives an expression for the PGF of the queue length.

Proposition 2.1 *Let $Q_{i,j}(z)$ denote the PGF of the queue length when the source is in state i and the channel is in state j. Then,*

$$\begin{aligned} \boldsymbol{Q}(z) &= [\boldsymbol{I} - \boldsymbol{P}^T diag[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{-1} [\boldsymbol{P}^T diag[z^i] \otimes \boldsymbol{R}^T [\boldsymbol{I} - \boldsymbol{E}(z)]] \boldsymbol{Q}_0 \\ &= \sum_{l=0}^{\infty} [\boldsymbol{P}^T diag[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{l+1} [\boldsymbol{I} \otimes [\boldsymbol{E}(z)^{-1} - \boldsymbol{I}]] \boldsymbol{Q}_0. \end{aligned} \tag{10}$$

where

$$\begin{aligned} \boldsymbol{Q}(z) &\triangleq [Q_{0,0}(z), Q_{0,1}(z), Q_{1,0}(z), Q_{1,1}(z), \cdots, Q_{N,0}(z), Q_{N,1}(z)]^T \\ \boldsymbol{Q}_l(z) &\triangleq [Q_{l,0}(z) \quad Q_{l,1}(z)]^T, \quad l = 0, 1, \ldots, N \\ \boldsymbol{E}(z) &\triangleq diag[\eta_0(z), \eta_1(z)] \\ \eta_j(z) &\triangleq e_j + \bar{e}_j z^{-1}, \quad j = 0, 1 \\ diag[z^i] &\triangleq diag[1, z, z^2, \cdots, z^N] \end{aligned}$$

Proof: See Appendix B.

Let $Q(z) \triangleq \sum_{i,j} Q_{i,j}(z)$ be the PGF of the queue length. In (10), $\boldsymbol{Q}_0$ contains the unknown terms $q_{0,0}[0]$ and $q_{0,1}[0]$. Since for a stable system $Q(z)$ is analytic in a closed unit disk, these unknown terms can be obtained by finding all the poles of $Q(z)$ in a closed unit disk [22]. The following diagonalization of the matrix $\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)$ may facilitate finding these poles:

$$\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z) = \boldsymbol{G}(z)\boldsymbol{\Lambda}(z)\boldsymbol{G}^{-1}(z) \tag{11}$$

where $\boldsymbol{\Lambda}(z) = \mathrm{diag}[\lambda_0(z), \lambda_1(z), \cdots, \lambda_{2N+1}(z)]$. For each $\lambda_l(z)$, $l = 0, 1, \cdots, 2N+1$, let $g_l(z)$ and $h_l(z)$ denote the respective left column and right row eigenvectors of (11). The matrices $\boldsymbol{G}(z)$ and $\boldsymbol{G}^{-1}(z)$ are given by:

$$\begin{aligned} \boldsymbol{G}(z) &\triangleq [g_0(z), g_1(z), \cdots, g_{2N+1}(z)] & (12)\\ \boldsymbol{G}^{-1}(z) &\triangleq [h_0(z), h_1(z), \cdots, h_{2N+1}(z)]^T. & (13) \end{aligned}$$

By spectral decomposition, we obtain:

$$\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z) = \sum_{l=0}^{2N+1} \lambda_l(z) g_l(z) h_l(z). \tag{14}$$

Each eigenvalue and eigenvector in the right-hand side (RHS) of the previous equation can be obtained by using properties of Kronecker products. An example for an on-off source is shown in Appendix A. Using the previous equation, we can simplify (10) into:

$$\begin{aligned} \boldsymbol{Q}(z) &= \sum_{l=0}^{\infty} [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{l+1} [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0 \\ &= \sum_{l=0}^{\infty} \sum_{i=0}^{2N+1} \lambda_i^{l+1}(z) g_i(z) h_i(z) [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0 \\ &= \sum_{i=0}^{2N+1} \frac{\lambda_i(z)}{1-\lambda_i(z)} g_i(z) h_i(z) [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0. \qquad (15) \end{aligned}$$

Substituting $\boldsymbol{Q}_0$ into the previous equation, we obtain:

$$Q(z) = \sum_{i=0}^{2N+1} \frac{\lambda_i(z)}{1-\lambda_i(z)} \sum_{l=0}^{2N+1} g_{li}(z) \left(h_{i0}(z) \frac{1-\eta_0(z)}{\eta_0(z)} q_{0,0}[0] + h_{i1}(z) \frac{1-\eta_1(z)}{\eta_1(z)} q_{0,1}[0] \right). \tag{16}$$

Let $\Delta(z)$ denote the characteristic function of the system:

$$\Delta(z) \triangleq \prod_{i=0}^{2N+1} (1 - \lambda_i(z)). \tag{17}$$

The poles of (16) are equal to the roots of this characteristic function. We need to determine the unknown variables $q_{0,0}[0]$ and $q_{0,1}[0]$. First, since $Q(z)$ is analytic for each root z_i, $|z_i| < 1$, we can set up the following boundary equation:

$$h_{i0}(z_i)\frac{1-\eta_0(z_i)}{\eta_0(z_i)}q_{0,0}[0] + h_{i1}(z_i)\frac{1-\eta_1(z_i)}{\eta_1(z_i)}q_{0,1}[0] = 0 \tag{18}$$

Secondly, we use the relation:

$$\lim_{z \to 1} Q(z) = 1. \tag{19}$$

By solving (18) and (19), one can determine the values of the unknown variables $q_{0,0}[0]$ and $q_{0,1}[0]$, which completes the solution for $Q(z)$. Thus, the mean queue length $\bar{q}$ is given by

$$\bar{q} = Q'(1). \tag{20}$$

Using Little's law, we obtain the mean queueing delay $\bar{d}$ for the ideal SR ARQ scenario:

$$\bar{d} = \frac{\bar{q}}{\rho_s} \tag{21}$$

where ρ_s is the mean arrival rate.

To obtain the transmission delay, we use the results in [23], where the mean number of transmission attempts per correctly received packet $\bar{n}$ was given by

$$\bar{n} = 1 + \boldsymbol{U}_r(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{V} \tag{22}$$

where $\boldsymbol{U}_r = [1 \;\; 1]$. The vector $\boldsymbol{S}$ and $\boldsymbol{V}$ are given by:

$$\boldsymbol{S} = \begin{bmatrix} r_{0,0}^{(s)}e_0 & r_{1,0}^{(s)}e_0 \\ r_{0,1}^{(s)}e_1 & r_{1,1}^{(s)}e_1 \end{bmatrix} \tag{23}$$

$$\boldsymbol{V} = \begin{bmatrix} \pi_{r,0}e_0 \\ \pi_{r,1}e_1 \end{bmatrix} \tag{24}$$

where $r_{i,j}^{(s)}$ corresponds to the (i,j)th element of the s-step transition matrix $\boldsymbol{R}$, and $\pi_{r,0}$ and $\pi_{r,1}$ are the steady-state probabilities that the channel is in Good and Bad states, respectively. Accordingly, the mean transmission delay is given by $s\bar{n} - \frac{s}{2}$ ($\frac{s}{2}$ is subtracted because the time to deliver an ACK to the transmitter does not contribute to the transmission delay). Combining the queueing delay in (21) and the transmission delay, we obtain the normalized mean transport delay T:

$$T = \bar{d} + s\bar{n} - \frac{s}{2}. \tag{25}$$

Using Little's law, we obtain the *total* mean number of packets at the transmitter for the SR ARQ with non-zero feedback delay (which includes packets in the source and waiting buffers):

$$E[q] = \bar{q} + s\bar{n}\rho_s. \tag{26}$$

3 Resequencing Delay

In this section, we give an approximate analysis of the mean resequencing delay under heavy traffic, i.e., packets are always supplied. This assumption has also been used in previous studies [8, 9], but for a static channel. We adapt the analytical approach in [8], which assumes *i.i.d.* packet error probabilities, to the underlying case where packet errors are correlated in a Markovian manner. Let $\boldsymbol{X}(t) \stackrel{\triangle}{=} (X_1(t), X_2(t), \cdots, X_s(t))$ denote the set of identifiers for packets transmitted during window t. We assume that packet identifiers are numbered in an increasing order. This assumption slightly affects the accuracy of our analytical results since the packet error probability is dependent on the location of a slot. However, the inaccuracy caused by this assumption is negligible in most practical situations except when the sojourn time of a channel state is very small relative to the window size (or the feedback delay).

The process $\{\boldsymbol{X}(t), t = 1, 2, \cdots\}$ governs the evolution of the occupancy of the resequencing buffer. Let $D_i(t)$ and $W_j(t)$ be defined as follows:

$$D_i(t) \stackrel{\triangle}{=} X_{i+1}(t) - X_i(t), \quad i = 1, 2, \cdots, s \tag{27}$$

$$W_j(t) \stackrel{\triangle}{=} \sum_{i=j}^{s} D_i(t), \quad j = 1, 2, \cdots, s \tag{28}$$

with $D_s(t) \stackrel{\triangle}{=} 1$. As an example, let $\boldsymbol{X}(1) = (1, 2, 3, 4, 5, 6, 7, 8)$ and $s = 8$. If transmissions of packets $2, 4, 5$, and 6 fail, then $\boldsymbol{X}(2) = (2, 4, 5, 6, 9, 10, 11, 12)$. Again, if transmission of the packet $4, 5, 9, 10$, and 12 fail, then $\boldsymbol{X}(3) = (4, 5, 9, 10, 12, 13, 14, 15)$. The corresponding $D_i(t)$ and $W_i(t)$ in this example are given in Table 1. The size of the resequencing buffer at windows $1, 2$, and 3 is 0, 3, and 4, respectively. Rosberg and Shacham [8] observed that the buffer occupancy at window t, $B(t)$, is given by:

$$B(t) = W_1(t) - s. \tag{29}$$

Furthermore, they observed that the system state $W_{s-i}(t)$, $t \geq 1$, $1 \leq i < s - 1$ is governed by

	Slots							
$\boldsymbol{X}(1)$	1	2	3	4	5	6	7	8
$D(1)$	1	1	1	1	1	1	1	1
$W(1)$	8	7	6	5	4	3	2	1
$\boldsymbol{X}(2)$	2	4	5	6	9	10	11	12
$D(2)$	2	1	1	3	1	1	1	1
$W(2)$	11	9	8	7	4	3	2	1
$\boldsymbol{X}(3)$	4	5	9	10	12	13	14	15
$D(3)$	1	4	1	2	1	1	1	1
$W(3)$	12	11	7	6	4	3	2	1

Table 1: Example showing the evolution of the occupancy of the resequencing buffer.

the following:

- If there are fewer than $s-i$ NACKs during window t, then $W_{s-i}(t+1) = i+1$.
- If there are $s-i+l$ NACKs, $0 \leq l \leq i$, and if the $(s-i)$th NACK is for the packet $X_k(t)$, $s-i \leq k \leq s-l$, then $W_{s-i}(t+1) = W_k(t) + (i-l)$.

In the following, we extend the previous analysis to the case of GE channel model. First, let $W_i^g(t)$ and $W_i^b(t)$ denote the value defined in (28) given that the state of the radio channel just before the beginning of window $t-1$ is Good (g) and Bad (b), respectively. The distribution of $W_{s-i}^g(t+1)$ is given by:

$$W_{s-i}^g(t+1) = \begin{cases} i+1, & \text{with probability } \sum_{m=0}^{s-i-1} p(s,m|g) \\ W_k(t) + (i-l), & \text{with probability} P_{t,g}(i,k,l) \end{cases} \tag{30}$$

where

$$\begin{aligned} P_{t,g} \;\; = \;\; & \sum_{k=s-i}^{s-l} (p(k-1, s-i-1, g|g)(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b)) \\ & + p(k-1, s-i-1, b|g)(r_{1,0}e_0 p(s-k, l|g) + r_{1,1}e_1 p(s-k, l|b)). \end{aligned}$$

In the previous equation, $p(n,k|r_1)$ denotes the probability of k unsuccessful transmissions in n consecutive slots given that the channel state at the beginning of a window is r_1. Similarly, $p(n,k,r_2|r_1)$ denotes the probability of k unsuccessful transmissions in n consecutive slots and the channel state in the last slot is r_2 given that the channel state just before the start of a window is r_1. In a similar way to the case of $W_{s-i}^g(t+1)$, the distribution of $W_{s-i}^b(t+1)$ is given by:

$$W_{s-i}^b(t+1) = \begin{cases} i+1, & \text{with probability } \sum_{m=0}^{s-i-1} p(s,m|b) \\ W_k(t) + (i-l), & \text{with probability} P_{t,b}(i,k,l) \end{cases} \tag{31}$$

where

$$
\begin{aligned}
P_{t,b}(i,k,l) \quad = \quad & \sum_{k=s-i}^{s-l} (p(k-1, s-i-1, g|b)(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b)) \\
& + p(k-1, s-i-1, b|b)(r_{1,0}e_0 p(s-k, l|g) + r_{1,1}e_1 p(s-k, l|b)).
\end{aligned}
$$

Letting $t \rightarrow \infty$, we obtain the steady-state probabilities for $W_{s-i}(t)$ and $W_{s-i}^{j}(t)$, where $j = g, b$:

$$
\Pr[W_{s-i} = k] \triangleq \lim_{t\rightarrow\infty} \Pr[W_{s-i}(t) = k] \tag{32}
$$

$$
\Pr[W_{s-i}^{j} = k] \triangleq \lim_{t\rightarrow\infty} \Pr[W_{s-i}^{j}(t) = k]. \tag{33}
$$

Let $\mathcal{W}_{s-i}(z)$ be the PGF of W_{s-i}. An expression for this PGF is given below.

Proposition 3.1

$$
\begin{aligned}
\mathcal{W}_{s-i}(z) \quad = \quad & \sum_{m=0}^{s-i-1} \boldsymbol{\Pi}\mathcal{P}[s,m]\boldsymbol{U}z^{i+1} \\
& + \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} \boldsymbol{\Pi}\mathcal{P}[k-1, s-i-1]\boldsymbol{R}\boldsymbol{E}\mathcal{P}[s-k, l]\boldsymbol{U}\mathcal{W}_k(z)z^{i-l}.
\end{aligned} \tag{34}
$$

where $\boldsymbol{\Pi}$ *is the steady-state probability vector of the radio state, i.e.,* $\boldsymbol{\Pi} = [\pi_{r,0}, \pi_{r,1}]$, $\boldsymbol{U} = [1 \;\; 1]^T$, $\boldsymbol{E} = diag[e_0, e_1]$, *and*

$$
\mathcal{P}[n,k] \triangleq \begin{bmatrix} p(n,k,g|g) & p(n,k,b|g) \\ p(n,k,g|b) & p(n,k,b|b) \end{bmatrix}.
$$

Proof: See Appendix C.

Exploiting the recursive structure of $\mathcal{P}[n,k]$, we arrive at the following difference equation:

$$
\mathcal{P}[n,k] = \boldsymbol{R}\bar{\boldsymbol{E}}\mathcal{P}[n-1,k] + \boldsymbol{R}\boldsymbol{E}\mathcal{P}[n-1,k-1] \tag{35}
$$

with the boundary conditions

$$
\begin{aligned}
\mathcal{P}[0,0] \quad &= \quad \boldsymbol{I} \\
\mathcal{P}[n,k] \quad &= \quad \boldsymbol{O}, \text{ if } n < k \text{ or } n,k < 0.
\end{aligned}
$$

where $\bar{\boldsymbol{E}} = \mathrm{diag}[\bar{e}_0, \bar{e}_1]$. The solution to (35) is obtained recursively for various values of n and k. To obtain the mean buffer size, we differentiate (34) with respect to z and evaluate at $z = 1$. Let

μ_{s-i} be defined as:

$$\mu_{s-i} \triangleq \frac{d\mathcal{W}_{s-i}(z)}{dz} |_{z=1} \ . \tag{36}$$

Thus, we have

$$\mu_{s-i} = (i+1)f_1(i) + \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} ((i-l)f_2(i,k,l) + f_2(i,k,l)\mu_k) \tag{37}$$

where

$$\begin{aligned} f_1(i) &= \sum_{m=0}^{s-i-1} \mathbf{\Pi}\mathcal{P}[s,m]\boldsymbol{U} \\ f_2(i,k,l) &= \mathbf{\Pi}\mathcal{P}[k-1,s-i-1]\boldsymbol{R}\boldsymbol{E}\mathcal{P}[s-k,l]\boldsymbol{U}. \end{aligned}$$

Arranging the previous equation, we obtain for $1 \leq i \leq s-1$:

$$\mu_{s-i} \;=\; \frac{(i+1)f_1(i) + \sum_{l=0}^{i}\sum_{k=s-i+1}^{s-l} f_2(i,k,l)(i-l+\mu_k) + \sum_{l=0}^{i} f_2(i,s-i,l)(i-l)}{1-\sum_{l=0}^{i} f_2(i,s-i,l)} \tag{38}$$

and $\mu_s = 1$. Accordingly, μ_1 is obtained recursively from (38) with the initial condition $\mu_s = 1$. The mean buffer occupancy is $\mu_1 - s$. Using Little's law, we obtain the mean resequencing delay T_r:

$$T_r = \frac{\mu_1 - s}{\pi_{r,0}\bar{e}_0 + \pi_{r,1}\bar{e}_1}. \tag{39}$$

4 Numerical Results and Discussion

In this section, we present numerical examples for the mean delay performance. The objectives of these examples are to: (1) test the accuracy of our approximate analysis, and (2) investigate the impacts of traffic correlations and the time-varying error behavior on the delay performance. The second objective will enable us to gauge the importance of employing a Markovian source model (as opposed to a renewal Bernoulli process) and a two-state time-varying channel model (as opposed to a one-state "static" model). This, in effect, establishes the true contribution of the paper. For the model validation part, our analytical results are contrasted against more realistic simulations that are obtained under no approximations. In the following, we present the results for the special yet commonly encountered case of a two-state on-off source. Note, however, that our analysis applies, in general, to an N-state source, where $N \geq 2$.

For an on-off Markov source, one packet is generated per time slot during on periods and no packets are generated during off periods. Transitions between on and off states are governed by the transition probability matrix $\boldsymbol{P} = [p_{i,j}], 0 \leq i, j \leq 1$. The characteristics of the on-off source are represented by the mean arrival rate (ρ_s), which is also the traffic load, and the mean length of the on periods (T_{on}). Note that for a fixed ρ_s, T_{on} measures the burstiness of the traffic. In practice, the value of T_{on} depends on the application and the underlying packetization mechanism. For example, for 53-byte ATM cells T_{on} ranges from few tens for voice with silence detection to few hundreds for compressed video.

Two parameters are defined for the GE channel model: the mean packet error rate (ϵ) and the duty cycle of the Bad period (ρ_r):

$$\epsilon \triangleq \frac{r_{0,1}e_1 + r_{1,0}e_0}{r_{0,1} + r_{1,0}} \tag{40}$$

$$\rho_r \triangleq \frac{r_{0,1}}{r_{0,1} + r_{1,0}} \tag{41}$$

Table 2 gives the parameter values used in our experiments. When varying the value of one parameter, the other parameters are set to their default values, unless indicated otherwise.

Parameter	Symbol	Value (default)
Mean arrival rate	ρ_s	$0.3 - 0.7(0.5)$
Mean on period	T_{on}	$10 - 300(100)$
Mean packet error rate	ϵ	$0.01 - 0.3(e_1 = 0.9, e_0 = 0.001)$
Duty cycle of Bad period	ρ_r	$0.05 - 0.4(0.1)$
Transition probability from Good to Bad	$r_{0,1}$	$0.005 - 0.1(0.03)$

Table 2: Parameters used to obtain the numerical results.

Fig. 4 shows the mean queue length $E[q]$, obtained using (26), as a function of ρ_s for three values of feedback delay: $s = 10, 50, 100$. Here, the mean queue length includes packets in the source buffer as well as those in the waiting buffer. A good agreement is observed between simulation and analysis, with the analytical results being slightly conservative when s is large. For $\rho_s < 0.8$, $E[q]$ increases almost linearly with the load indicating almost no change in the mean transport delay. This can be justified as follows. Since at most one packet is generated per time slot, queueing delay occurs only when a packet is retransmitted from the waiting buffer and the source buffer is not empty. The former condition is the result of a returned NACK, which occurs mostly during Bad channel periods. Given that the channel is Bad roughly 10% of the time (by default, $\rho_r = 0.1$), the likelihood that the channel is in a Bad state and, simultaneously, the source is on is small, resulting in infrequent queueing delays. In this case, the transport delay is dominated by the retransmission part, which is independent of the input load. As ρ_s increases beyond 0.8, queueing delay in the source buffer becomes more probable, causing a nonlinear increase in the *total* mean queue length (note that to maintain a stable queue, ρ_s must not exceed 0.9 since the channel is "clear" about 90% of the time).

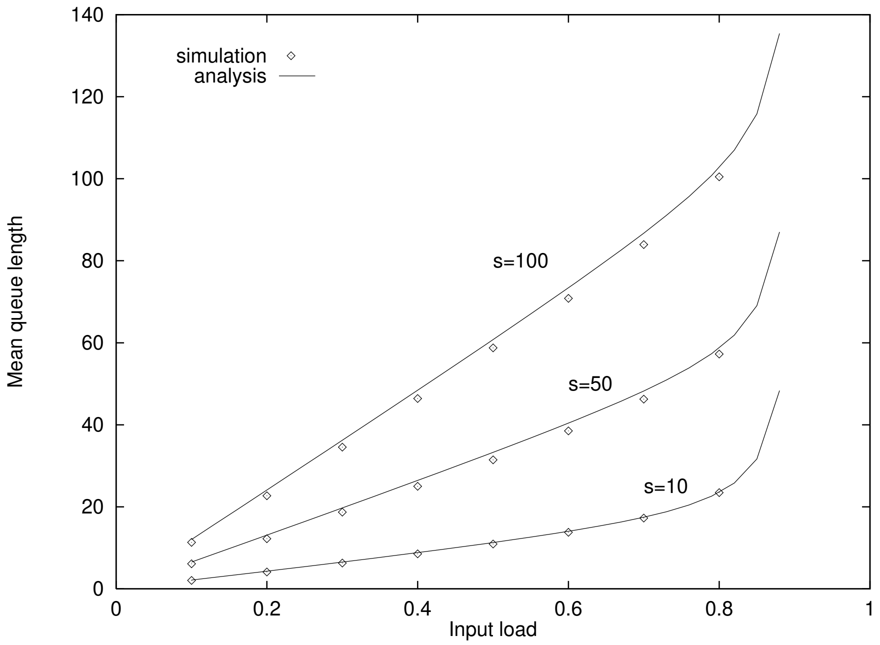

Figure 4: Mean queue length versus input load.

Fig. 5 shows the mean transport delay as a function of ρ_s for three values of feedback delay, $s = 10, 50, 100$. Time is normalized by a slot time. As shown in the figure, the mean transport delay is less sensitive to the input load. Since one packet is generated per slot, no queueing delay is occurred unless a NACK is returned. Since we fix the average duration of the Bad period at 10% of the average duration of the Good period, it is hard to notice significant queueing delay up to medium input load. As input load increases, it is more likely that the on state and Bad state occur simultaneously which may cause a higher queueing delay. In this figure, the analytical results tend to overestimate the simulations by less than 10%.

The impact of the mean packet error rate (ϵ) on the transport delay is illustrated in Fig. 6. Here, we take $\rho_s = 0.8$ and set the other parameters to their default values. We vary ϵ by varying the error rate during Bad periods (e_1) with e_0 set to 0.001. The figure indicates a good agreement between the approximate analysis and the simulations. The gap, however, tends to slightly increase with ϵ. It is interesting to note that as ϵ increases, the corresponding *rate* of increase in the mean transport delay is almost the same for different values of s (i.e., the gap between, say, the plots for $s = 1$ and $s = 10$ does not change with ϵ). At first, we expected this gap to increase with ϵ, since we speculated that increasing ϵ should result in an increase in *both* the queueing and retransmission delays. While the queueing delay does not depend on s (since it is obtained under the ideal SR ARQ assumption), the retransmission delay equals to $s(\bar{n} - 1/2)$, and hence depends on s. If the retransmission delay were to increase with ϵ, through an increase in the mean number of transmission attempts $\bar{n}$, then the gap between, say, $s = 1$ and $s = 10$ in Fig. 6 should also increase. Surprisingly, it turns out that the increase in the transport

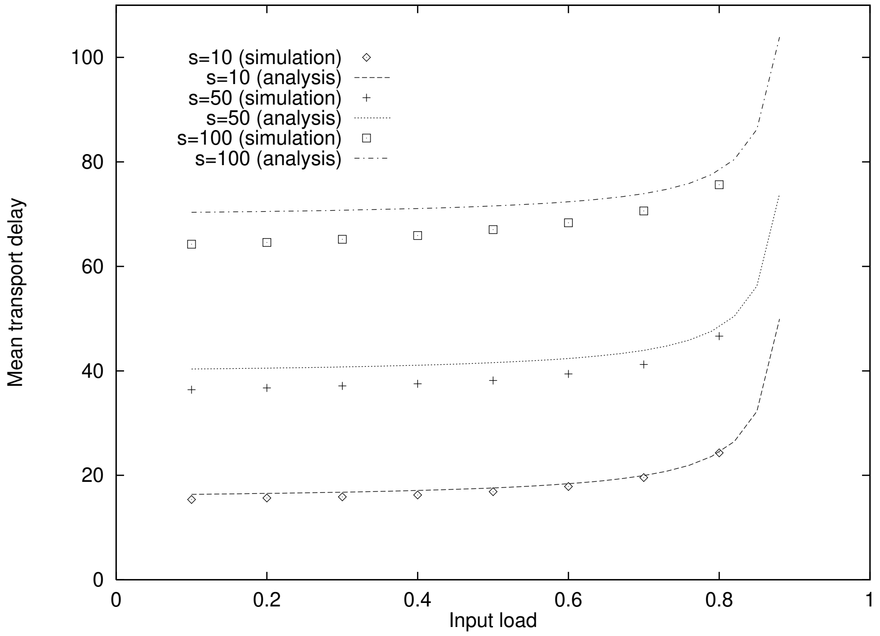

Figure 5: Mean transport delay versus input load.

delay shown in Fig. 6 is mainly caused by the queueing delay. We found that as s increases at a fixed ϵ, $\bar{n}$ *decreases* causing the product $s\bar{n}$ to be almost constant. This unexpected interaction can be justified as follows. Retransmissions are mainly caused by errors occurring during the Bad periods. When s is small compared to the mean length of a Bad period, it is likely that a retransmission of an erroneous packet will take place during the same Bad period in which the packet was last (re)transmitted, and hence will probably encounter channel errors. This, in effect, increases the value of $\bar{n}$. As s increases, the impact of error correlations fades away, resulting in a decrease in $\bar{n}$.

Fig. 7 shows the mean transport delay as a function of the duty cycle of a Bad period (ρ_r) for $s = 50, 100$. As expected, the mean transport delay increases with the duty cycle of a Bad period.

Fig. 8 shows the mean transport delay as a function of the transition probability from on to off states ($p_{1,0}$) for $s = 10, 50$. For the input load (ρ_s) of 0.5, it is observed that mean transport delay increases abruptly for the longer on period. However, it becomes much insensitive from the value 0.03.

Fig. 9 demonstrates the significance of traffic correlations by contrasting the mean transport delay under a Markovian model ($T_{on} = 16$ and 32) with that of a renewal Bernoulli model. All results are based on analysis. The mean error rate is varied through e_1 with $\rho_s = 0.8$. Note that in this case the value of T_{on} reflects the burstiness of the Markovian source. For a fixed ϵ, the Markovian model always gives a higher mean transport delay than the Bernoulli model. In fact, even when T_{on} is set to its smallest possible value (i.e., $T_{on} = 2$), we observed that the

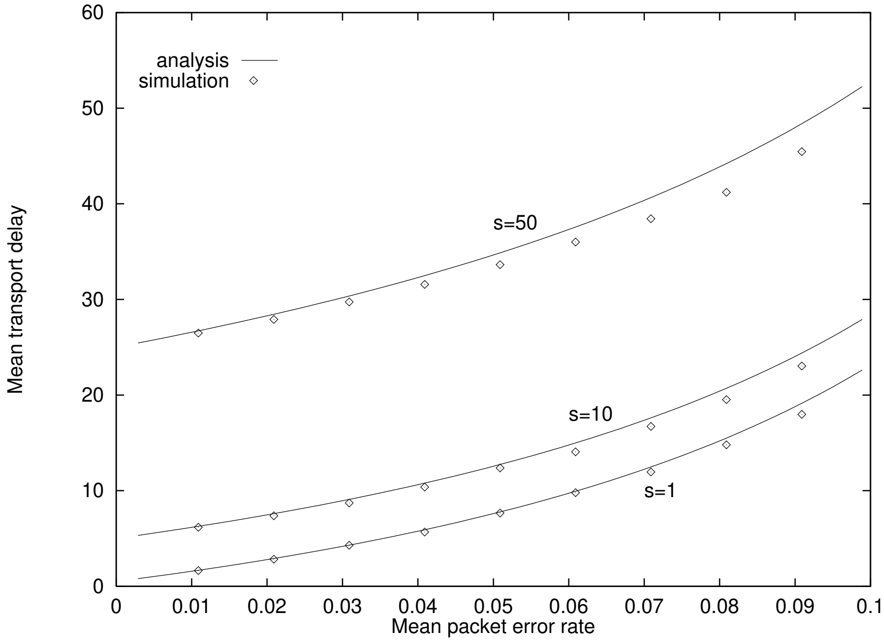

Figure 6: Mean transport delay versus mean packet error rate ($\rho_s = 0.8$).

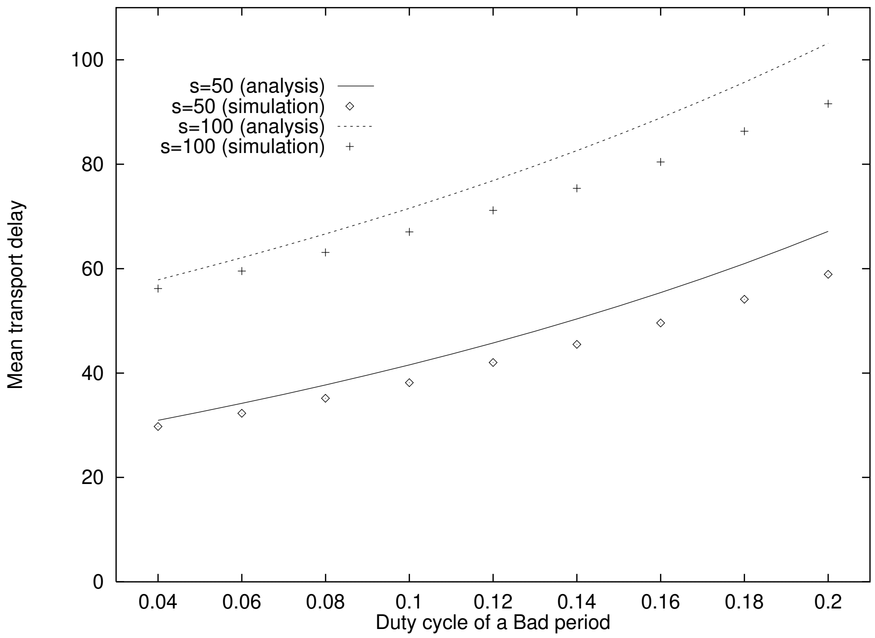

Figure 7: Mean transport delay versus the duty cycle of a Bad period.

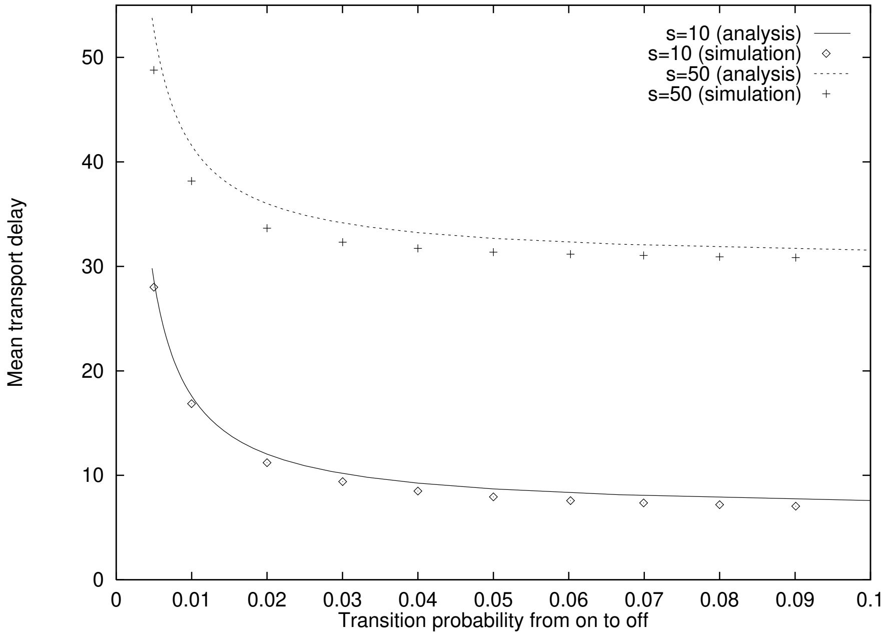

Figure 8: Mean transport delay versus the transition probability from on to off.

Markovian model still results in higher mean delay value 10. For a fixed T_{on}, as ϵ increases the gap between the Bernoulli and Markovian results also increases, indicating more profound impact for correlations in this regime. This trend is justified as follows. When ϵ is small, the queue is empty most of the time. In this case, the traffic characteristics have little impact on the queueing delay (i.e., traffic burstiness is absorbed by the clear channel). As the noise level increases, so does the mean queue length at the source buffer. But with more frequent buffer buildup, burstiness starts to have a more profound impact on the mean queueing delay by increasing the likelihood that an arriving packet will see a nonempty queue. As a matter of fact, if we only consider the mean queueing delay (i.e., $s = 0$), the gap between the Markovian and Bernoulli results is even more acute than what is depicted in Fig. 9 (See 10). The figure also shows a fixed gap between the curves for $s = 1$ and those for $s = 10$ in each source model, which we previously commented on.

The significance of employing a time-varying channel-error model as opposed to a static one is illustrated in Fig. 11. In this figure, we depict the mean transport delay versus ϵ under Markov and Bernoulli sources assuming static (one state) and time-varying (two states) channel models. As before, ϵ for the 2-state channel model is varied through e_1 with $\rho_r = 0.4$. We set $\rho_s = 0.4$, $s = 10$, and use the default values for the other fixed parameters. From this figure, one can make the following remarks. First, for a fixed mean error rate and a given source model, a time-varying error model almost always results in a larger mean transport delay compared to a static model. This trend can be attributed to the fact that a time-varying channel acts as a "bursty" server whose alternating Good-Bad pattern adversely impacts the effective service rate. So during Good

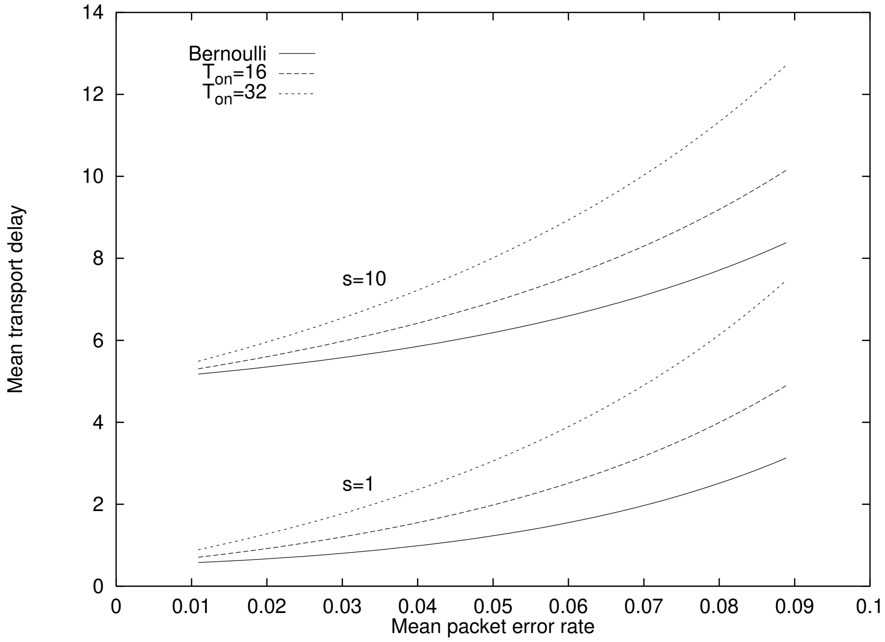

Figure 9: Mean transport delay versus ϵ under Markovian and Bernoulli sources ($\rho_s = 0.8$).

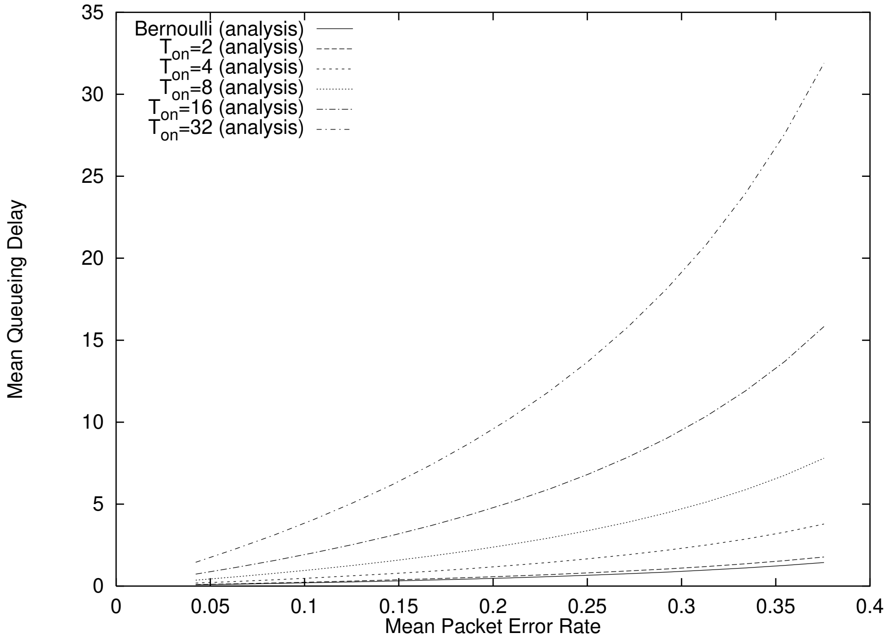

Figure 10: Mean queueing delay versus ϵ under Markovian and Bernoulli sources ($\rho_s = 0.8, \rho_r = 0.41667, r_{01} = 0.5$).

periods, if no traffic is being generated and the queue is empty, the capacity is wasted for an extended amount of time. By evently distributing the effective service rate over time, the static model undermines the amount of wasted capacity, hence providing overly optimistic predictions of the mean delay. Secondly, the gap between the time-varying and static channel models increases with ϵ. This is attributed to a corresponding increase in the variance of the mean error rate in the former model (this variance is zero in the static model). Finally, it is observed that in the case of the Markovian source model, the effect of the time-varying channel starts to show at smaller values of ϵ compared to the Bernoulli model. So by employing a Bernoulli model, one may mistakenly undermine the impact of the time-varying error behavior. For large ϵ, the considerable gap between the uppermost and lowermost curves in Fig. 11 demonstrates the amount of inaccuracy caused by employing simplified source and channel models.

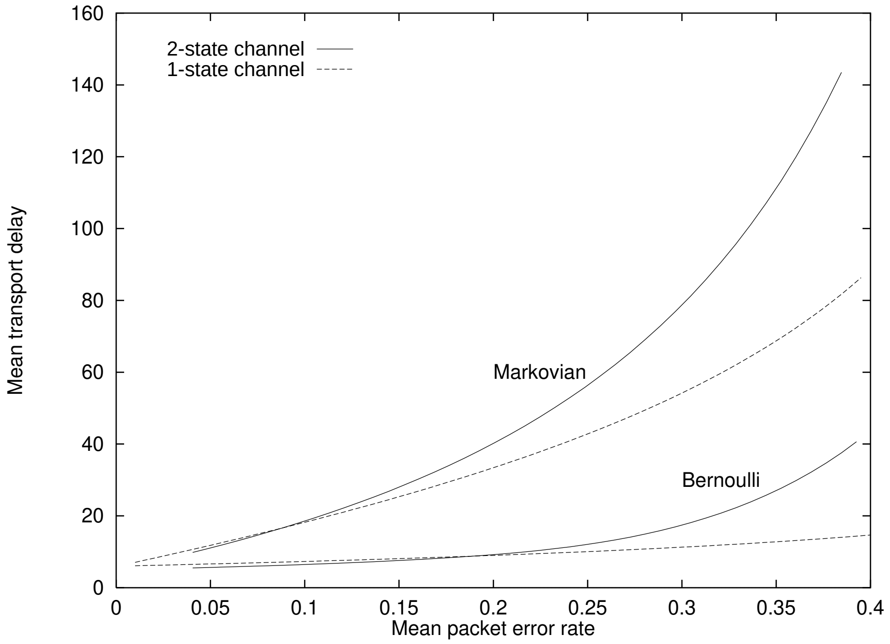

Figure 11: Mean transport delay versus ϵ under Markovian and Bernoulli sources for static and time-varying channels ($\rho_s = 0.4$, $\rho_r = 0.4$, $s = 10$).

Fig. 12 and 13 are related to the mean resequencing delay. The objective of both figures is to examine the accuracy of our approximate analysis in Section 3. Recall that the analysis was carried out under two assumptions: heavy-traffic load (i.e., packets are always supplied) and a small window size relative to the channel sojourn times. Fig. 12 shows the mean resequencing delay versus ϵ for $s = 50, 100$. With such large values of s, we can examine the worst-case inaccuracy of the analysis with respect to the second assumption. The analysis is contrasted with *exact* simulation results obtained under heavy load and under 70% load. If the resequencing analysis were to be conducted without the second assumption, then we would expect a match between the analytical results and the heavy-traffic simulations. But Fig. 12 shows that the heavy-traffic simulations upper bound the analytical results, indicating opposite effects for the above two assumptions. For $s = 50$, the analytical results are sufficiently close to both types of simulations. In general, we observed that at small values of s the mean resequencing delay is somehow insensitive to the input load. For $s = 100$, the analytical results lie between the two types of simulations, being closer to the heavy-traffic simulations when ϵ is small and to the 70%-load simulations when ϵ is large. In Fig. 13 we compare the analytical results with simulations obtained at 80% load using two values for e_1 (the error rate during Bad periods). A good agreement is observed over the range of s. As s increases, e_1 starts to have a more significant impact on the mean resequencing delay.

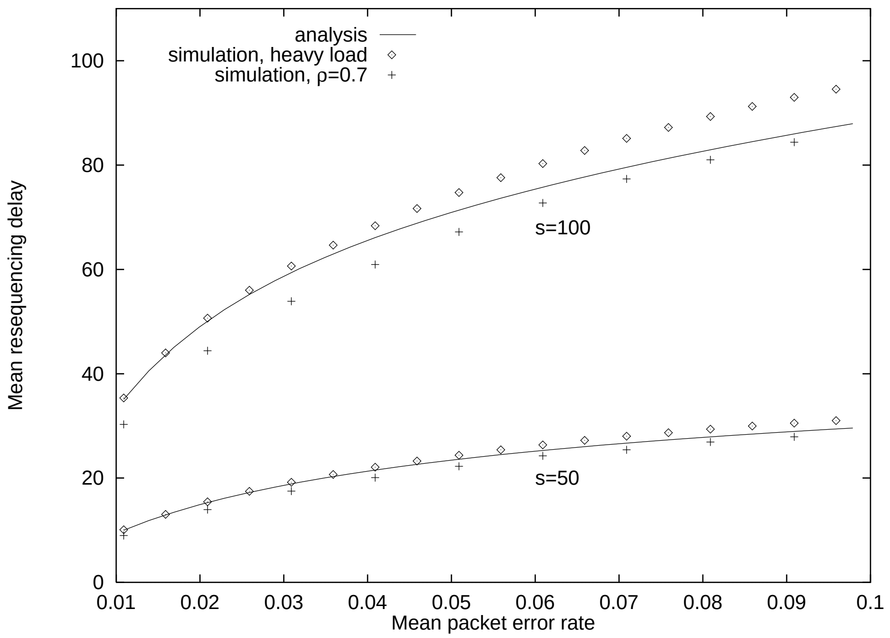

Figure 12: Mean resequencing delay versus mean packet error rate.

In contrast to the convex plots in Fig. 6 and 9, Fig. 12 indicates that the mean resequencing delay increases sublinearly with ϵ. This says that as ϵ increases, the mean transport delay becomes more sensitive to incremental changes in ϵ, while the opposite effect is observed for the mean resequencing delay. Fig. 14 further illustrates this point by depicting both types of delay as a function of ϵ when $s = 50,\ 100$ ($\rho_s = 0.8$, $\rho_r = 0.2$). Because of the contrasting sensitivities to changes in ϵ, two crossover points in the delay curves are observed when $s = 100$.

Fig. 15 shows the mean resequencing delay versus the duty cycle of a Bad period for the cases of feedback delay, $s = 50, 100$. As shown in the figure, the mean resequencing delay shows a gradual increase as ρ_r increases.

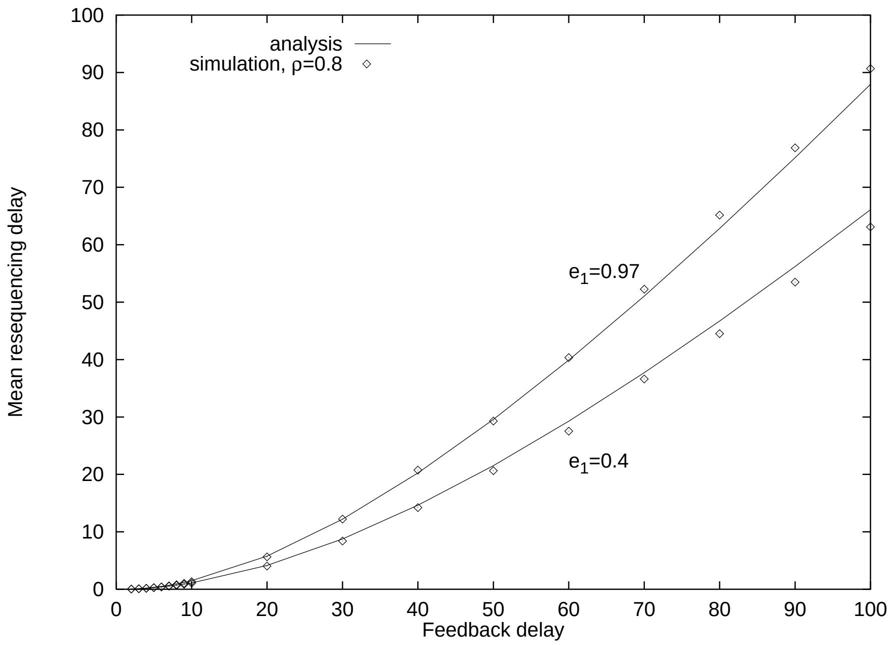

Figure 13: Mean resequencing delay versus feedback delay.

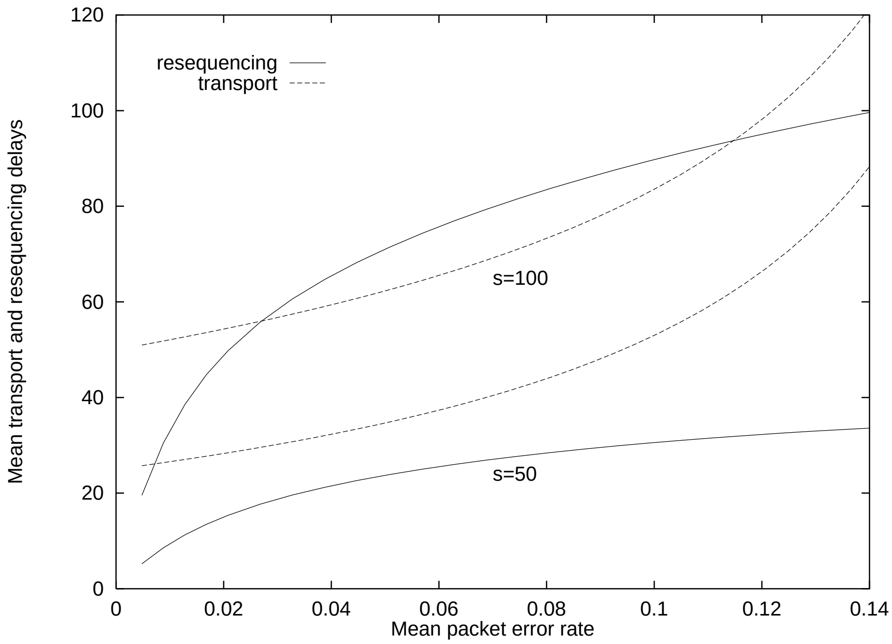

Figure 14: Mean transport and resequencing delays versus ϵ.

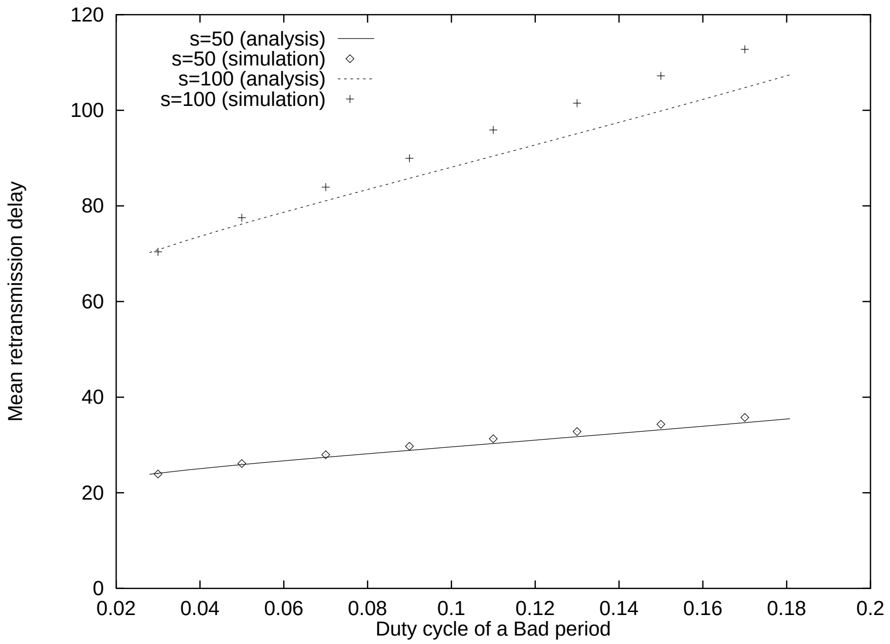

Figure 15: Mean resequencing delay versus the duty cycle of a Bad period.

5 Conclusions

In this study, we analyzed the total mean delay (queueing, transmission, and resequencing) experienced by a packet over a time-varying wireless channel with SR ARQ error control. The goal of such analysis is to enable *online* assessment of the delay performance to be used in resource allocation and admission control subject to guaranteed quality of service. In contrast to previous studies, ours was conducted under more realistic source and channel models. By employing a Markovian source model, our analysis accommodates the inherent correlations in network traffic. We showed that ignoring such correlations through the use of a Bernoulli renewal model can lead to significantly underestimating the true mean transport delay, particularly when the channel error rate is high. As for the channel model, we conducted our analysis assuming a two-state Gilbert-Elliot model, which captures the time-varying and correlated nature of channel errors. We found that at high channel error rates the mean transport delay obtained under the Gilbert-Elliot model is much larger than the one obtained under a one-state "static" model. This, again, points to the overly optimistic predictions that one can get when ignoring the time-varying error behavior of the channel. Overall, the implications of ignoring traffic and channel correlations are most apparent at high channel error rates.

Interestingly, we found that the feedback delay (s) has two opposite effects on the transmission part of the mean transport delay. While a larger s expectedly increases the RTT per transmission, it also reduces the mean number of retransmissions per packet by increasing the separation between two successive retransmissions. Since retransmissions are mainly caused by errors occurring during Bad states, a large s allows a retransmission of an erroneous packet to "escape" the same Bad state of the previous retransmission, hence improving its chance of error-free delivery. In other words, the negative impact of channel correlations tends to fade away as s increases.

We observed that the mean resequencing delay is somehow insensitive to the input load for small to moderate values of s and moderate to high input loads. This delay increases sublinearly with ϵ (the mean error rate), displaying concave functionality over the range of ϵ. In contrast, the mean transport delay increases with ϵ as a convex function. As a result, the transport delay tends to dominate the overall delay for small and large ϵ. For medium values of ϵ, the dominant delay component depends on the value of s.

Our analysis of the both transport (specifically, queueing) and resequencing delays was conducted under certain simplifying assumptions. By contrasting the analytical results against more realistic simulations, we verified that the inaccuracy of such assumptions is negligible for most practical cases. This inaccuracy tends to slightly increase with both s and ϵ.

APPENDIX

A Single On-Off Source

In Section 2, we outlined a spectral decomposition approach for obtaining the PGF of the queue length under a general Markovian arrival process. We now apply this approach to the case of a single on-off source. The eigenvalues and eigenvectors of $\boldsymbol{P}^T \text{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)$ can be obtained by employing some properties of Kronecker products [24]; namely, the eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices, are obtained by element-wise multiplication of the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$. Also, the eigenvectors of $\boldsymbol{A} \otimes \boldsymbol{B}$ are given by the Kronecker products of the individual eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Let $\alpha_{1,2}(z)$ and $\beta(z)$ denote the eigenvalues and eigenvectors of $\boldsymbol{P}^T \text{diag}[z^i]$, respectively, which are given by:

$$\alpha_{1,2}(z) = \frac{1}{2}(p_{0,0} + p_{1,1} z \pm \xi(z)) \tag{42}$$

$$\beta(z) = \begin{bmatrix} 1 & 1 \\ \frac{\alpha_1(z) - p_{0,0}}{(1-p_{1,1})z} & \frac{\alpha_2(z) - p_{0,0}}{(1-p_{1,1})z} \end{bmatrix} \tag{43}$$

where $\xi(z) = \sqrt{(p_{0,0} + p_{1,1} z)^2 + 4(1 - p_{0,0} - p_{1,1})z}$.

Let $\delta_{1,2}(z)$ and $\gamma(z)$ denote the eigenvalues and eigenvectors of $\boldsymbol{R}^T \boldsymbol{E}(z)$, respectively, which are given by:

$$\delta_{1,2}(z) = \frac{1}{2}(r_{0,0}\eta_0(z) + r_{1,1}\eta_1(z) \pm \sigma(z)) \tag{44}$$

$$\gamma(z) = \begin{bmatrix} 1 & 1 \\ \frac{\delta_1(z)-r_{0,0}\eta_0(z)}{(1-r_{1,1})\eta_1(z)} & \frac{\delta_2(z)-r_{0,0}\eta_0(z)}{(1-r_{1,1})\eta_1(z)} \end{bmatrix} \quad (45)$$

where $\sigma(z) = \sqrt{(r_{0,0}\eta_0(z) + r_{1,1}\eta_1(z))^2 + 4(1 - r_{0,0} - r_{1,1})\eta_0(z)\eta_1(z)}$. In addition, the inverse of each eigenvector is given by:

$$\beta^{-1}(z) = \begin{bmatrix} \frac{\alpha_2(z)-p_{0,0}}{\alpha_2(z)-\alpha_1(z)} & \frac{-(1-p_{1,1})z}{\alpha_2(z)-\alpha_1(z)} \\ \frac{-(\alpha_1(z)-p_{0,0})}{\alpha_2(z)-\alpha_1(z)} & \frac{(1-p_{1,1})z}{\alpha_2(z)-\alpha_1(z)} \end{bmatrix} \quad (46)$$

$$\gamma^{-1}(z) = \begin{bmatrix} \frac{\delta_2(z)-r_{0,0}\eta_0(z)}{\delta_2(z)-\delta_1(z)} & \frac{-(1-r_{1,1})\eta_1(z)}{\delta_2(z)-\delta_1(z)} \\ \frac{-(\delta_1(z)-r_{0,0}\eta_0(z))}{\delta_2(z)-\delta_1(z)} & \frac{(1-r_{1,1})\eta_1(z)}{\delta_2(z)-\delta_1(z)} \end{bmatrix}. \quad (47)$$

Thus, the right-hand side of (11) for a single on-off source is given by:

$$\begin{aligned} \mathbf{\Lambda}(z) &= \mathrm{diag}[\alpha_1(z)\delta_1(z), \alpha_1(z)\delta_2(z), \alpha_2(z)\delta_1(z), \alpha_2(z)\delta_2(z)] \\ \boldsymbol{G}(z) &= \boldsymbol{\beta} \otimes \boldsymbol{\gamma}(z) \\ &= [\beta_1 \otimes \gamma_1(z), \beta_1 \otimes \gamma_2(z), \beta_2 \otimes \gamma_1(z), \beta_2 \otimes \gamma_2(z)] \\ \boldsymbol{G}^{-1}(z) &= (\boldsymbol{\beta} \otimes \boldsymbol{\gamma}(z))^{-1} \\ &= [\beta_1^{-1} \otimes \gamma_1^{-1}(z), \beta_1^{-1} \otimes \gamma_2^{-1}(z), \beta_2^{-1} \otimes \gamma_1^{-1}(z), \beta_2^{-1} \otimes \gamma_2^{-1}(z)]. \end{aligned}$$

B Proof of Proposition 2.1

¿From (8) and (9), we can obtain $Q_{i,j}(z)$:

$$\begin{aligned} Q_{i,j}(z) = & \sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n+1)} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}[n-l+1] + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}[n-l+1]) z^n \\ & + \sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n)} (r_{j,j} e_j p_{l,i} q_{l,j}[n-l] + r_{\bar{j},j} e_{\bar{j}} p_{l,i} q_{l,\bar{j}}[n-l]) z^n \\ & + \sum_{l=0}^{1} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}[1-l] + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}[1-l]) \\ & + p_{0,i}(r_{j,j} q_{0,j}[0] + r_{\bar{j},j} q_{0,\bar{j}}[0]). \end{aligned} \quad (48)$$

In order to simplify the previous equation, we use the following relations:

$$\sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n)} p_{l,i} q_{l,j}[n-l] z^n = \sum_{l=0}^{N} \sum_{n=l}^{\infty} p_{l,i} q_{l,j}[n-l] z^n - p_{0,i} q_{0,j}[0] \quad (49)$$

and

$$\sum_{n=1}^{\infty}\sum_{l=0}^{\min(N,n+1)} p_{l,i}q_{l,j}[n-l+1]z^n =$$
$$\sum_{l=2}^{N}\sum_{n=l-1}^{\infty} p_{l,i}q_{l,j}[n-l+1]z^n + \sum_{l=0}^{1}\sum_{n=1}^{\infty} p_{l,i}q_{l,j}[n-l+1]z^n. \tag{50}$$

Using (49) and (50), we obtain:

$$\begin{aligned} Q_{i,j}(z) &= r_{j,j}\bar{e}_j p_{0,i}q_{0,j}[0] + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{0,i}q_{0,\bar{j}}[0] \\ &+ r_{j,j}\bar{e}_j \sum_{l=2}^{N} p_{l,i}z^{l-1}Q_{l,j}(z) + r_{j,j}\bar{e}_j(z^{-1}p_{0,i}(Q_{0,j}(z)-q_{0,j}[0]) + p_{1,i}Q_{1,j}(z)) \\ &+ r_{\bar{j},j}\bar{e}_{\bar{j}} \sum_{l=2}^{N} p_{l,i}z^{l-1}Q_{l,\bar{j}}(z) + r_{\bar{j},j}\bar{e}_{\bar{j}}(z^{-1}p_{0,i}(Q_{0,\bar{j}}(z)-q_{0,\bar{j}}[0]) + p_{1,i}Q_{1,\bar{j}}(z)) \\ &+ r_{j,j}e_j \sum_{l=0}^{N} p_{l,i}z^{l}Q_{l,j}(z) + r_{\bar{j},j}e_{\bar{j}} \sum_{l=0}^{N} p_{l,i}z^{l}Q_{l,\bar{j}}(z) \end{aligned} \tag{51}$$

Arranging the previous equation, we obtain:

$$\begin{aligned} Q_{i,j}(z) - r_{j,j}\eta_j(z)\sum_{l=0}^{N} p_{l,i}z^l Q_{l,j}(z) - r_{\bar{j},j}\eta_{\bar{j}}(z)\sum_{l=0}^{N} p_{l,i}z^l Q_{l,\bar{j}}(z) = \\ r_{j,j}(1-\eta_j(z))p_{0,i}q_{0,j}[0] + r_{\bar{j},j}(1-\eta_{\bar{j}}(z))p_{0,i}q_{0,\bar{j}}[0] \end{aligned} \tag{52}$$

where $\eta_j(z) = e_j + \bar{e}_j z^{-1}$ and $\eta_{\bar{j}}(z) = e_{\bar{j}} + \bar{e}_{\bar{j}} z^{-1}$. In the previous equation,

$$\begin{aligned} \sum_{l=0}^{N} p_{l,i}z^l (r_{j,j}\eta_j(z)Q_{l,j}(z) + r_{\bar{j},j}\eta_{\bar{j}}(z)Q_{l,\bar{j}}(z)) &= \sum_{l=0}^{N} p_{l,i}z^l [\boldsymbol{R}^T\boldsymbol{E}(z)]_{(j)}\boldsymbol{Q}_l(z) \\ &= [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T\boldsymbol{E}(z)]_{(2i+j)}\boldsymbol{Q}(z) \end{aligned} \tag{53}$$

where $[\boldsymbol{A}]_{(i)}$ denotes the (i)th row of $\boldsymbol{A}$ and

$$\begin{aligned} \boldsymbol{Q}_l(z) &\triangleq [Q_{l,0}(z) \quad Q_{l,1}(z)]^T \\ \boldsymbol{E}(z) &\triangleq \mathrm{diag}[\eta_0(z), \eta_1(z)] \\ \mathrm{diag}[z^i] &\triangleq \mathrm{diag}[1, z, z^2, \cdots, z^N] \\ \boldsymbol{Q}(z) &\triangleq [Q_{0,0}(z), Q_{0,1}(z), Q_{1,0}(z), Q_{1,1}(z), \cdots, Q_{N,0}(z), Q_{N,1}(z)]^T. \end{aligned}$$

Using (53), we can arrange (52) in the following matrix form:

$$[\boldsymbol{I} - \boldsymbol{P}^T\mathrm{diag}[z^i] \otimes \boldsymbol{R}^T\boldsymbol{E}(z)]\boldsymbol{Q}(z) = [\boldsymbol{P}^T\mathrm{diag}[z^i] \otimes \boldsymbol{R}^T[\boldsymbol{I} - \boldsymbol{E}(z)]]\boldsymbol{Q}_0 \tag{54}$$

where $\boldsymbol{Q}_0 = [q_{0,0}[0], q_{0,1}[0], 0, \cdots, 0]^T$.

With some algebraic manipulation of (54), we obtain:

$$\begin{aligned}\boldsymbol{Q}(z) &= [\boldsymbol{I} - \boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{-1}[\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T [\boldsymbol{I} - \boldsymbol{E}(z)]]\boldsymbol{Q}_0 \\ &= \sum_{l=0}^{\infty}[\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{l+1}[\boldsymbol{I} \otimes [\boldsymbol{E}(z)^{-1} - \boldsymbol{I}]]\boldsymbol{Q}_0. \end{aligned} \tag{55}$$

This completes the proof of Proposition 2.1.

C Proof of Proposition 3.1

For $j = g,\ b$, let $\mathcal{W}^j_{s-i}(z)$ be the PGF of W^j_{s-i}. Taking the z-transform for (30), we have

$$\begin{aligned}\mathcal{W}^g_{s-i}(z) &= \sum_{l=0}^{i}\sum_{k=s-i}^{s-l}(p(k-1, s-i-1, g|g)(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b)) \\ &\quad + p(k-1, s-i-1, b|g)(r_{1,0}e_0 p(s-k, l|g) + r_{1,1}e_1 p(s-k, l|b)))\mathcal{W}_k(z) z^{i-l} \\ &\quad + \sum_{m=0}^{s-i-1} p(s, m|g) z^{i+1}. \end{aligned} \tag{56}$$

Similarly, for (31) we have

$$\begin{aligned}\mathcal{W}^b_{s-i}(z) &= \sum_{l=0}^{i}\sum_{k=s-i}^{s-l}(p(k-1, s-i-1, g|b)(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b)) \\ &\quad + p(k-1, s-i-1, b|b)(r_{1,0}e_0 p(s-k, l|g) + r_{1,1}e_1 p(s-k, l|b)))\mathcal{W}_k(z) z^{i-l} \\ &\quad + \sum_{m=0}^{s-i-1} p(s, m|b) z^{i+1}. \end{aligned} \tag{57}$$

Multiplying (56) and (57) by $\pi_{r,0}$ and $\pi_{r,1}$, respectively, and summing each, we obtain the following equation:

$$\begin{aligned}\mathcal{W}_{s-i}(z) &= \sum_{m=0}^{s-i-1} \boldsymbol{\Pi}\mathcal{P}[s, m]\boldsymbol{U} z^{i+1} \\ &\quad + \sum_{l=0}^{i}\sum_{k=s-i}^{s-l} \boldsymbol{\Pi}\mathcal{P}[k-1, s-i-1]\boldsymbol{R}\boldsymbol{E}\mathcal{P}[s-k, l]\boldsymbol{U}\mathcal{W}_k(z) z^{i-l}. \end{aligned} \tag{58}$$

This completes the proof of Proposition 3.1.

References

[1] R. Fantacci, "Queueing analysis of the selective repeat automatic repeat request protocol

wireless packet networks," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 258–264, May 1996.

[2] R. Fantacci, "Mean packet delay analysis for selective repeat automatic repeat request protocol with correlated arrivala and deterministic and nondeterministic acknowledgement delays," *Telecommunication Systems.*, vol. 9, pp. 41–57, 1998.

[3] E. Ayanoglu, K. Y. Eng, and M. J. Karol, "Wireless ATM: limits, challenges, and protocols," *IEEE Pers. Commun.*, vol. 3, no. 4, pp. 18–34, 1996.

[4] M. E. Anagnostou and E. N. Protonotarios, "Performance analysis of the selective repeat ARQ protocol," *IEEE Trans. Commun.*, vol. 34, no. 2, pp. 127–135, Feb. 1986.

[5] A. G. Konheim, "A queueing analysis of two ARQ protocols," *IEEE Trans. Commun.*, vol. 28, no. 7, pp. 1004–1014, July 1980.

[6] J. Chang and T. Yang, "End-to-end delay of an adaptive selective repeat ARQ protocol," *IEEE Trans. Commun.*, vol. 42, no. 11, pp. 2926–2928, Nov. 1994.

[7] Z. Rosberg and M. Sidi, "Selective-repeat ARQ: the joint distribution of the transmitter and the receiver resequencing buffer occupancies," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1430–1438, Sept. 1990.

[8] Z. Rosberg and N. Shacham, "Resequencing delay and buffer occupancy under the selective-repeat ARQ," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 166–173, Jan. 1989.

[9] N. Shacham and D. Towsley, "Resequencing delay and buffer occupancy in selective repeat ARQ with multiple receivers," *IEEE Trans. Commun.*, vol. 39, no. 6, pp. 928–936, June 1991.

[10] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. SAC-4, no. 6, pp. 856–868, Sept. 1986.

[11] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 833–846, Sept. 1986.

[12] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[13] B. Hajek and L. He, "On variations of queue response for inputs with identical means and autocorrelation functions," *IEEE/ACM Trans. Networking*, vol. 6, no. 5, pp. 588–598, Oct. 1998.

[14] San qi Li and Chia-Lin Hwang, "Queue response to input correlation functions: Discrete spectral analysis," *IEEE/ACM Trans. Networking*, vol. 1, no. 5, pp. 522–533, Oct. 1993.

[15] Victor S. Frost and Benjamin Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, vol. 32, no. 3, pp. 70–81, Mar. 1994.

[16] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566–1579, 1995.

[17] Mark W. Garrett and Walter Willinger, "Analysis, modeling, and generation of self-similar VBR video traffic," in *Proc. of the SIGCOMM '94 Conference*, Sept. 1994, pp. 269–280.

[18] Minothi Parulekar and Armand Makowski, "$M/G/\infty$ input processes: A versatile class of models for network traffic," in *Proceedings of IEEE INFOCOM '97*, Apr. 1997.

[19] M. Grossglauser and Jean-Chrysostome Bolot, "On the relevance of long-range dependence in network traffic," in *Proceedings of the ACM SIGCOMM '96 Conference*, 1996.

[20] Bong Ryu and Anwar Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," in *Proceedings of the ACM SIGCOMM '96 Conference*, Aug. 1996, pp. 3–14.

[21] D. Heyman and T.V. Lakshman, "What are the implications of long-range dependence for VBR video traffic engineering?," *IEEE/ACM Trans. Networking*, vol. 4, pp. 301–317, June 1996.

[22] S. Li, "A general solution technique for discrete queueing analysis of multimedia traffic on ATM," *IEEE Trans. Commun.*, vol. 39, no. 7, pp. 1115–1132, July 1991.

[23] S. Kallel, "Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel," *IEEE Trans. Commun.*, vol. 40, no. 9, pp. 1474–1480, Sept. 1992.

[24] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Automat. Contr.*, vol. CAS-25, no. 9, pp. 772–780, Sept. 1978.