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

Abstract

In this paper, we analyze the delay performance for a Markovian source transported over a wireless channel with time-varying error characteristics. To improve the reliability of the channel, the end points of the wireless link implement a selective-repeat (SR) ARQ error control protocol. We provide an approximate discrete-time analysis of the end-to-end mean packet delay, which consists of transport and resequencing delays. The transport delay, in turn, consists of queueing, transmission/retransmission, and propagation delays. In contrast to previous studies, our analysis accommodates the inherent autocorrelations in *both* the input traffic and the channel state. Our approximation of the mean transport delay is based on decoupling the dependence of the queueing behavior from the past 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, our analysis is performed under heavy-traffic assumptions, hence providing an upper bound on the actual mean resequencing delay. Numerical results and simulations indicate that our approximate analysis is sufficiently accurate for a wide range of parameter values.

1 Introduction

Automatic repeat request (ARQ) protocols are used to provide reliable data transfer in wireless communications [1, 9, 11]. 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. There are three major types of ARQ protocols: stop-and-wait (SW), go-back-N (GBN), and selective-repeat (SR). In SW ARQ, after sending a packet the transmitter stays idle while waiting for the ACK/NACK message of that packet. The channel capacity is wasted if meanwhile other packets are waiting to be transmitted. 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.

In this study, we consider a wireless link that provides sequential delivery of packets and that uses SR ARQ for error control. This scenario arises in various wireless transport technologies, including wireless asynchronous transfer mode (ATM), where the in-sequence delivery of cells must be maintained to ensure seamless interface with the wireline ATM network [2]. In such a scenario, the transmitter assigns each packet a unique identifier (see Fig. 1). Packets are transmitted according to their identifiers. Following its transmission, a packet is temporarily buffered until its ACK arrives back at the transmitter after some feedback delay. Once the ACK arrives, the packet is removed from the waiting buffer and a new packet is transmitted. If instead a NACK arrives, the packet is retransmitted. Although the transmitter sends packets in the proper order, the order of correctly received 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 other packets with lower identifiers are correctly received.

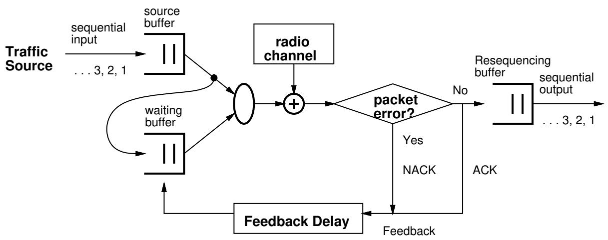

Figure 1: Packet transmission over a wireless link.

Fig. 2 shows the various delay components that a packet undergoes when transported over a wireless link. The end-to-end delay consists of transport and resequencing delays. The transport delay is again subdivided into queueing and

* This research was supported by the National Science Foundation under CAREER Grant ANI-9733143.

retransmission delays. The queueing delay is defined as the time taken by a packet from its arrival at the transmitter buffer until its first transmission attempt. The retransmission delay is defined as the time from a packet's first transmission until its *successful* arrival at the receiver. The resequencing delay is defined as the waiting time of the packet in the resequencing buffer.

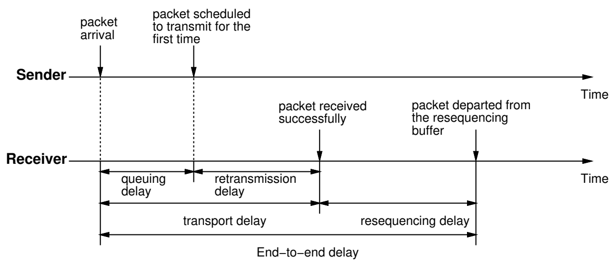

Figure 2: Time diagram for packet transmission process.

Most previous studies of the delay performance under SR ARQ error control were based on renewal traffic models (e.g., Poisson, Bernoulli) [4, 5, 10]. Such models do not capture the bursty and highly autocorrelated nature of actual network traffic, and can lead to inaccurate performance predictions. In addition to traffic correlations, accurate evaluation of the performance over a wireless link also requires capturing the time-varying nature of the wireless link and its highly correlated error behavior.

In this study, we investigate the mean end-to-end delay for a general Markovian source that is transported over a wireless link with SR ARQ error control. To capture the time-varying and correlated nature of the radio channel, we model it using Gilbert-Elliot's model. We divide the end-to-end delay into queueing, retransmission, and resequencing delays (see Fig. 2). For the queueing delay, we simplify the analysis by eliminating the dependence of the queueing process on the past packet transmission process. That is, we assume that the feedback message for a packet transmitted in the past is treated (statistically) as the one for the packet at the head-of-line (HOL) of the queue. This simplifying assumption becomes exact as the feedback delay approaches zero. Fantacci in [5] refers to this scenario as the *ideal* SR ARQ. We derive the exact probability generating function (PGF) for the ideal SR ARQ and obtain the mean delay using Little's law. The mean retransmission delay is easily obtained since it only depends on the channel parameters and the round-trip delay. Finally, we derive an expression for the mean resequencing delay under heavy-traffic conditions, i.e., packets are always supplied to the system. This gives an upper bound on the actual mean resequencing delay. The adequacy of our analytical results are verified by contrasting them against more realistic simulation results.

An extensive amount of literature exists on analyzing the performance of SR ARQ protocol in terms of throughput, mean queue length, and mean delay. In order to analyze other measures of performance such as buffer distribution and packet delay, Konheim used the system state vector considering a feedback delay [7]. Anagnostou and Protonotarios proposed an alternative approximate approach that reduces the computational complexity of the analytical results [1]. One problem with these approaches is that their computational complexity increases dramatically with the feedback delay. Furthermore, many of them rely on renewal traffic models and/or overly simplified channel models (e.g., *i.i.d.* bit errors). Fantacci used a 2-state Markovian radio channel, yet employing a Bernoulli process for packet arrivals [5]. Rosberg and Shacham analyzed the resequencing delay and the buffer occupancy at the resequencing buffer assuming heavy-traffic conditions and static radio channel [9]. Rosberg and Sidi analyzed the joint distribution of buffer occupancy at the transmitter and receiver [10]. In addition, they derived the mean transmission and resequencing delays. However, they assumed a renewal arrival process and independent packet errors. Schachum and Towsley investigated the buffer occupancy and resequencing delay for the situation in which a single transmitter and multiple receivers communicate, assuming heavy-traffic conditions and independent packet errors [11].

The rest of the paper is organized as follows. In Section 2 we present the queueing model for a general Markovian source transmitted over a radio channel. The mean resequencing delay is studied in Section 3. Numerical results and simulations are reported in Section 4, followed by concluding remarks in Section 5.

2 Queueing and Retransmission Delays

Consider the queueing system at the transmitter side of a wireless link. Our queueing 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. A time slot corresponds to a packet transmission time. We assume that ACK/NACK messages are always error free. 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. The wireless channel is modeled by Gilbert-Elliot's model, in which the channel alternates between *Good* and *Bad* states, with corresponding bit error probabilities P_{eg} and P_{eb}, respectively. It is assumed that state transitions occur at the end of time slots. Since the packet transmission time is very short compared to the sojourn time of a channel state, the inaccuracy due to this assumption is negligible. The packet error probability when the channel state is in state j is denoted by e_j, $j = 0, 1$. The packet error probability in Good ($j = 0$) and Bad ($j = 1$) 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.

Our analytical approach is based on the approximation in [1, 5], where the transport delay is divided into two parts: queueing and retransmission delays (see Fig. 2). In order to obtain the queueing delay, the authors in [1, 5] approximate the behavior of a real SR ARQ by ignoring the dependence of ACK/NACK arrivals on the system's past history. The packet at the HOL of the transmitter buffer will be transmitted only if an ACK message arrives. At a given time t, the feedback message that arrives at the transmitter corresponds to 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 that is associated with the packet to be transmitted at time t, i.e., as if the feedback delay is zero. This simplification is referred to as the *ideal* SR ARQ case [5]. Note that this assumption does not mean the feedback delay is ignored, but that its impact on the queueing process is not incorporated. In the following, 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.
$q(k\|i,j)$:	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 buffer.
$\boldsymbol{R}$:	Transition probability matrix for the process that describes the state of the radio channel.

The transition probabilities for the arrival process are defined as $\boldsymbol{P} = [p_{i,j}]$, where

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

Also, the transition probabilities for the channel process are defined as $\boldsymbol{R} = [r_{i,j}]$, where

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

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 in the previous slot, the number of packets that arrive during slot k, 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 $q(k|\cdot,\cdot) + a(k) > 0$, then

$$q(k+1|l,j) = \begin{cases} q(k|i,j) + i - 1, \\ \quad \text{with probability } p_{i,l} \cdot (1-e_j) \cdot r_{j,j} \\ q(k|i,j) + i, \\ \quad \text{with probability } p_{i,l} \cdot e_j \cdot r_{j,j} \\ q(k|i,1-j) + i - 1, \\ \quad \text{with probability } p_{i,l} \cdot (1-e_{1-j}) \cdot r_{1-j,j} \\ q(k|i,1-j) + i, \\ \quad \text{with probability } p_{i,l} \cdot e_{1-j} \cdot r_{1-j,j} \end{cases} \quad (5)$$

and if $q(k|\cdot,\cdot) + a(k) = 0$, then

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

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:

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

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

If $n > 0$,

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

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

And, if $n = 0$

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

Let $Q_{i,j}(z)$ denote the PGF of the queue length:

$$Q_{i,j}(z) \triangleq \sum_{n=0}^{\infty} q_{i,j}(n) z^n.$$

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

$$\begin{aligned} Q_{i,j}(z) = &\sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n+1)} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}(n-l+1) \\ &+ r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}(n-l+1)) z^n + \sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n)} \\ &(r_{j,j} e_j p_{l,i} q_{l,j}(n-l) + r_{\bar{j},j} e_{\bar{j}} p_{l,i} q_{l,\bar{j}}(n-l)) z^n \\ &+ \sum_{l=0}^{1} (r_{j,j}\bar{e}_j p_{l,i} q_{l,j}(1-l) + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{l,i} q_{l,\bar{j}}(1-l)) \\ &+ p_{0,i}(r_{j,j} q_{0,j}(0) + r_{\bar{j},j} q_{0,\bar{j}}(0)). \end{aligned} \quad (10)$$

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

$$\begin{aligned} &\sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n)} p_{l,i} q_{l,j}(n-l) z^n = \\ &\quad \sum_{l=0}^{N} \sum_{n=l}^{\infty} p_{l,i} q_{l,j}(n-l) z^n - p_{0,i} q_{0,j}(0) \end{aligned} \quad (11)$$

and

$$\begin{aligned} &\sum_{n=1}^{\infty} \sum_{l=0}^{\min(N,n+1)} p_{l,i} q_{l,j}(n-l+1) z^n \\ &\quad = \sum_{l=2}^{N} \sum_{n=l-1}^{\infty} p_{l,i} q_{l,j}(n-l+1) z^n \\ &\quad + \sum_{l=0}^{1} \sum_{n=1}^{\infty} p_{l,i} q_{l,j}(n-l+1) z^n. \end{aligned} \quad (12)$$

Using (11) and (12), we obtain:

$$\begin{aligned} Q_{i,j}(z) = &\, r_{j,j}\bar{e}_j p_{0,i} q_{0,j}(0) + r_{\bar{j},j}\bar{e}_{\bar{j}} p_{0,i} q_{0,\bar{j}}(0) \\ &+ r_{j,j}\bar{e}_j \sum_{l=2}^{N} p_{l,i} z^{l-1} Q_{l,j}(z) + r_{j,j}\bar{e}_j (z^{-1} p_{0,i}(Q_{0,j}(z) \\ &- q_{0,j}(0)) + p_{1,i} Q_{1,j}(z)) + r_{\bar{j},j}\bar{e}_{\bar{j}} \sum_{l=2}^{N} p_{l,i} z^{l-1} Q_{l,\bar{j}}(z) \\ &+ r_{\bar{j},j}\bar{e}_{\bar{j}} (z^{-1} p_{0,i}(Q_{0,\bar{j}}(z) - q_{0,\bar{j}}(0)) + p_{1,i} Q_{1,\bar{j}}(z)) \\ &+ r_{j,j} e_j \sum_{l=0}^{N} p_{l,i} z^l Q_{l,j}(z) + r_{\bar{j},j} e_{\bar{j}} \sum_{l=0}^{N} p_{l,i} z^l Q_{l,\bar{j}}(z) \end{aligned} \quad (13)$$

Arranging the previous equation, we obtain:

$$\begin{aligned}
&Q_{i,j}(z) - r_{j,j}\eta_j(z)\sum_{l=0}^{N} p_{l,i} z^l Q_{l,j}(z) \\
&\quad -r_{\bar{j},j}\eta_{\bar{j}}(z)\sum_{l=0}^{N} p_{l,i} z^l Q_{l,\bar{j}}(z) = r_{j,j}(1-\eta_j(z))p_{0,i}q_{0,j}(0) \\
&\quad +r_{\bar{j},j}(1-\eta_{\bar{j}}(z))p_{0,i}q_{0,\bar{j}}(0) \qquad (14)
\end{aligned}$$

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,

$$\begin{aligned}
&\sum_{l=0}^{N} p_{l,i} z^l (r_{j,j}\eta_j(z) Q_{l,j}(z) - r_{\bar{j},j}\eta_{\bar{j}}(z) Q_{l,\bar{j}}(z)) \\
&\quad = \sum_{l=0}^{N} p_{l,i} z^l [\boldsymbol{R}^T \boldsymbol{E}(z)]_{(j)} \boldsymbol{Q}_l \\
&\quad = [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]_{(2i+j)} \boldsymbol{Q} \qquad (15)
\end{aligned}$$

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

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

Using (15), we can arrange (14) in the following matrix form:

$$\begin{aligned}
&[\boldsymbol{I} - \boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)] \boldsymbol{Q}(z) = \\
&\qquad [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T [\boldsymbol{I} - \boldsymbol{E}(z)]] \boldsymbol{Q}_0 \qquad (16)
\end{aligned}$$

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

With some algebraic manipulation of (16), we obtain:

$$\begin{aligned}
\boldsymbol{Q}(z) &= [\boldsymbol{I} - \boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{-1} [\boldsymbol{P}^T \mathrm{diag}[z^i] \\
&\qquad \otimes \boldsymbol{R}^T [\boldsymbol{I} - \boldsymbol{E}(z)]] \boldsymbol{Q}_0 \\
&= \sum_{l=0}^{\infty} [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{l+1} \\
&\qquad [\boldsymbol{I} \otimes [\boldsymbol{E}(z)^{-1} - \boldsymbol{I}]] \boldsymbol{Q}_0. \qquad (17)
\end{aligned}$$

In (17), $\boldsymbol{Q}_0$ 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 the closed unit disk [8]. Finding these poles is facilitated by the following diagonalization of the matrix $\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)$:

$$\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z) = \boldsymbol{G}(z) \boldsymbol{\Lambda}(z) \boldsymbol{G}^{-1}(z) \qquad (18)$$

where $\boldsymbol{\Lambda}(z)$ is a diagonal matrix given by:

$$\boldsymbol{\Lambda}(z) = \mathrm{diag}[\lambda_0(z), \lambda_1(z), \cdots, \lambda_{2N+1}(z)].$$

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 (18) given by:

$$\boldsymbol{G}(z) = [g_0(z), g_1(z), \cdots, g_{2N+1}(z)] \qquad (19)$$

and

$$\boldsymbol{G}^{-1}(z) = [h_0(z), h_1(z), \cdots, h_{2N+1}(z)]^T. \qquad (20)$$

By spectral decomposition, we obtain:

$$\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z) = \sum_{l=0}^{2N+1} \lambda_l(z) g_l(z) h_l(z). \qquad (21)$$

Each eigenvalue and eigenvector in the RHS of the previous equation can be obtained by using properties of Kronecker products. Based on the previous equation, we can simplify (17) into:

$$\begin{aligned}
\boldsymbol{Q}(z) &= \sum_{l=0}^{\infty} [\boldsymbol{P}^T \mathrm{diag}[z^i] \otimes \boldsymbol{R}^T \boldsymbol{E}(z)]^{l+1} [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0 \\
&= \sum_{l=0}^{\infty} \sum_{i=0}^{2N+1} \lambda_i^{l+1}(z) g_i(z) h_i(z) [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0 \\
&= \sum_{i=0}^{2N+1} \frac{\lambda_i(z)}{1-\lambda_i(z)} g_i(z) h_i(z) [\boldsymbol{I} \otimes \boldsymbol{E}(z)^{-1} - \boldsymbol{I}] \boldsymbol{Q}_0. \qquad (22)
\end{aligned}$$

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

$$\begin{aligned}
Q(z) &= \sum_{i=0}^{2N+1} \frac{\lambda_i(z)}{1-\lambda_i(z)} \sum_{l=0}^{2N+1} g_{li}(z) \\
&\quad \left(h_{i0}(z) \frac{1-\eta_0(z)}{\eta_0(z)} q_{0,0} + h_{i1}(z) \frac{1-\eta_1(z)}{\eta_1(z)} q_{0,1} \right). \qquad (23)
\end{aligned}$$

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

$$\Delta(z) \triangleq \prod_{i=0}^{2N+1} (1 - \lambda_i(z)). \qquad (24)$$

The poles of (23) are equal to the roots of this characteristic function. We need to determine two unknown variables $q_{0,0}$ and $q_{0,1}$ using two conditions. First, since $Q(z)$ is analytic for each root z_i, $|z_i| < 1$, we can set up the following boundary condition:

$$h_{i0}(z_i) \frac{1-\eta_0(z_i)}{\eta_0(z_i)} q_{0,0} + h_{i1}(z_i) \frac{1-\eta_1(z_i)}{\eta_1(z_i)} q_{0,1} = 0 \qquad (25)$$

Secondly, we use the relation:

$$\lim_{z \to 1} Q(z) = 1. \qquad (26)$$

Solving (25) and (26), we can determine the values of the unknown variables $q_{0,0}$ and $q_{0,1}$. Thus, the mean queue length $\bar{q}$ is given by

$$\bar{q} = Q'(1). \qquad (27)$$

Also, using Little's law, we obtain the mean packet delay $\bar{d}$ given by

$$\bar{d} = \frac{\bar{q}}{\rho_s} \qquad (28)$$

where ρ_s is the mean arrival rate. Recall that we approximate the *queueing delay* under SR ARQ error control by $\bar{d}$

in (28), which is the mean queueing delay under an ideal SR ARQ system with zero feedback delay.

To obtain the retransmission delay for a real SR ARQ, we use the results in [6], where the mean number of transmission attempts per correctly received packet $\bar{n}$ was given by

$$\bar{n} = 1 + \boldsymbol{U}_r(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{V} \tag{29}$$

where $\boldsymbol{U}_r = [1\ \ 1]$ and

$$\boldsymbol{S} = \begin{bmatrix} r_{0,0}^{(s)}e_0 & r_{1,0}^{(s)}e_0 \\ r_{0,1}^{(s)}e_1 & r_{1,1}^{(s)}e_1 \end{bmatrix} \tag{30}$$

$$\boldsymbol{V} = \begin{bmatrix} \pi_{r,0}e_0 \\ \pi_{r,1}e_1 \end{bmatrix} \tag{31}$$

where $r_{i,j}^{(s)}$ corresponds to the (i,j)th element of 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.

Combining the queueing delay in (28) and the retransmission delay in (29), we obtain the normalized mean transmission delay T:

$$T = \bar{d} + s\bar{n} - \frac{s}{2}. \tag{32}$$

In the previous equation, $\bar{d}$ and $s\bar{n} - \frac{s}{2}$ correspond to the mean queueing and retransmission delays, respectively. The term $\frac{s}{2}$ is subtracted because the time to deliver an ACK to the transmitter does not contribute to the transport delay. Using Little's law, the mean number of packets in the queue at the transmitter for a real SR ARQ is given by:

$$E[q] = \bar{q} + s\bar{n}\rho_s. \tag{33}$$

3 Resequencing Delay

In this section, we derive an upper bound on the mean resequencing delay obtained under a heavy-traffic scenario, i.e., packets are always supplied. Given a feedback delay of s slots, the feedback message from the receiver is delivered to the transmitter s slots after the packet is transmitted. We adapt the analytical approach in [9], which assumes *i.i.d.* packet error probabilities, to the underlying case where packet errors are correlated in a Markovian manner. Let $\boldsymbol{X}(t) \triangleq (X_1(t), X_2(t), \cdots, X_s(t))$ denote the set of identifiers of the packets which are transmitted during window t. We assume that packet identifiers are numbered in an increasing order. This assumption affects the accuracy of this analytical approach since the packet error probability is dependent on the location of a slot. However, the error caused by this assumption is acceptable in most practical situations except when the sojourn time of a channel state is 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 $W_k(t)$ be defined as follows:

$$D_i(t) \triangleq X_{i+1}(t) - X_i(t), \quad i = 1, 2, \cdots, s \tag{34}$$

$$W_j(t) \triangleq \sum_{i=j}^{s} D_i(t), \quad j = 1, 2, \cdots, s \tag{35}$$

with $D_s(t) \triangleq 1$. 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, $\boldsymbol{X}(2) = (2,4,5,6,9,10,11,12)$. Again, if transmission of packets 4, 5, 9, 10, and 12 fail, $\boldsymbol{X}(3) = (4,5,9,10,12,13,14,15)$. In this example, the corresponding $D_i(t)$ and $W_i(t)$ are given in Table 1. The size of the resequencing buffer at windows

	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: An example of the evolution of the occupancy of the resequencing buffer.

1, 2, and 3 is 0, 3, and 4, respectively. Rosberg and Shacham [9] observed that the buffer occupancy at window t, $B(t)$, is given by:

$$B(t) = W_1(t) - s. \tag{36}$$

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

- If there were fewer than $s-i$ NACK's during window t, then $W_{s-i}(t+1) = i+1$
- If there were $s-i+l$ NACK's, $0 \leq l \leq i$, and if the $(s-i)$th NACK was 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 Gilbert-Elliot's channel. First, let $W_i(t|g)$ and $W_i(t|b)$ denote the value defined in (35) given that the state of the radio channel before the beginning of window $t-1$ is Good (g) and Bad (b), respectively. The distribution of $W_{s-i}(t+1|g)$ is given by:

$$W_{s-i}(t+1|g) = \begin{cases} i+1, \\ \quad \text{with probability } \sum_{m=0}^{s-i-1} p(s,m|g) \\ W_k(t) + (i-l), \\ \quad \text{with probability} P_{t,g}(i,k,l) \end{cases} \tag{37}$$

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

$$W_{s-i}(t+1|b) = \begin{cases} i+1, \\ \quad \text{with probability } \sum_{m=0}^{s-i-1} p(s,m|b) \\ W_k(t) + (i-l), \\ \quad \text{with probability} P_{t,b}(i,k,l) \end{cases} \tag{38}$$

where

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

Taking the z-transform, we have

$$W_{s-i}(z|g) = \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} (p(k-1, s-i-1, g|g)$$
$$\cdot(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b))$$
$$+p(k-1, s-i-1, b|g)(r_{1,0}e_0 p(s-k, l|g)$$
$$+r_{1,1}e_1 p(s-k, l|b)))W_k(z)z^{i-l}$$
$$+ \sum_{m=0}^{s-i-1} p(s, m|g)z^{i+1}. \tag{39}$$

Similarly, for (38) we have

$$W_{s-i}(z|b) = \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} (p(k-1, s-i-1, g|b)$$
$$\cdot(r_{0,0}e_0 p(s-k, l|g) + r_{0,1}e_1 p(s-k, l|b))$$
$$+p(k-1, s-i-1, b|b)(r_{1,0}e_0 p(s-k, l|g)$$
$$+r_{1,1}e_1 p(s-k, l|b)))W_k(z)z^{i-l}$$
$$+ \sum_{m=0}^{s-i-1} p(s, m|b)z^{i+1}. \tag{40}$$

Multiplying (39) and (40) by $\pi_{r,0}$ and $\pi_{r,1}$, respectively, and summing each, we obtain the following equation:

$$W_{s-i}(z) = \sum_{m=0}^{s-i-1} \mathbf{\Pi}\mathcal{P}[s,m]\boldsymbol{U}z^{i+1} + \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} \mathbf{\Pi}$$
$$\cdot\mathcal{P}[k-1, s-i-1]\boldsymbol{RE}\mathcal{P}[s-k, l]\boldsymbol{U}W_k(z)z^{i-l}. \tag{41}$$

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

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

Exploiting the recursive structure, we obtain the following difference equation:

$$\mathcal{P}[n,k] = \boldsymbol{R\bar{E}}\mathcal{P}[n-1,k] + \boldsymbol{RE}\mathcal{P}[n-1,k-1] \tag{42}$$

with the boundary conditions

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

where $\boldsymbol{\bar{E}} = \mathrm{diag}[\bar{e}_0, \bar{e}_1]$. The solution to the above difference equation is obtained numerically. To obtain the mean buffer size, we differentiate (41) with respect to z and evaluate at $z = 1$. Let μ_{s-i} be defined as:

$$\mu_{s-i} \triangleq \frac{dW_{s-i}(z)}{dz} |_{z=1} . \tag{43}$$

Thus, we have

$$\mu_{s-i} = (i+1)f_1(i)$$
$$+ \sum_{l=0}^{i} \sum_{k=s-i}^{s-l} ((i-l)f_2(i,k,l) + f_2(i,k,l)\mu_k) \tag{44}$$

where

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

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

$$\mu_{s-i} = ((i+1)f_1(i) + \sum_{l=0}^{i} \sum_{k=s-i+1}^{s-l} f_2(i,k,l)(i-l+\mu_k)+$$
$$\sum_{l=0}^{i} f_2(i, s-i, l)(i-l))(1 - \sum_{l=0}^{i} f_2(i, s-i, l))^{-1} \tag{45}$$

and $\mu_s = 1$. The mean buffer occupancy is $\mu_1 - s$. Using Little's law, we obtain the mean resequencing delay T_r:

$$T_r = \frac{\mu_1 - s}{\pi_{r,0}\bar{e}_0 + \pi_{r,1}\bar{e}_1}. \tag{46}$$

4 Numerical Results

We now give numerical examples based on the previously presented analysis and contrast them against more realistic simulation results. We consider a single discrete-time on-off source in which one packet is generated in a time slot during the on periods. Transitions between on and off states are governed by the transition probability matrix $\boldsymbol{P} = [p_{i,j}], 0 \le i,j \le 1$. The characteristics of the on-off source are represented by the mean arrival rate (ρ_s) and the mean length of the on periods (T_{on}). We also define three parameters for the Gilbert-Elliot's radio channel: the average packet error rate (ϵ), the duty cycle of the Bad period (ρ_r), and the ratio of the mean packet error rate during a Bad state to the mean packet error rate during a Good state:

$$\epsilon \triangleq \frac{r_{0,1}e_1 + r_{1,0}e_0}{r_{0,1} + r_{1,0}} \tag{47}$$

$$\rho_r \triangleq \frac{r_{0,1}}{r_{0,1} + r_{1,0}} \tag{48}$$

$$\theta \triangleq \frac{e_1}{e_0} \tag{49}$$

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

Fig. 3 shows the mean queue length as a function of the load ρ_s for three values of feedback delay, $s = 10, 50, 100$. Note that the queue length also includes already transmitted packets that are waiting for acknowledgments. A good agreement is observed between simulation and analysis.

Fig. 4 shows the mean transport delay as a function of ρ_s for $s = 10, 50, 100$ (time is in slot units). 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

Parameter	Symbol	Range of values (default value)
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	$0.05 - 0.2(0.1)$
Transition probability from Good to Bad	$r_{0,1}$	$0.005 - 0.1(0.03)$
Ratio of mean packet error rate of Bad state over Good state	θ	variable

Table 2: Parameters used in the numerical results.

fix the average duration of the Bad period at 10% of the average duration of the Good period, it is hard to notice any significant queueing delay up to medium input load. As ρ_s 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%.

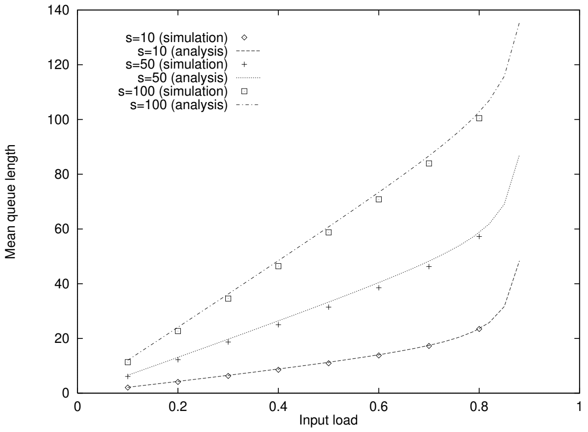

Figure 3: Mean queue length versus input load.

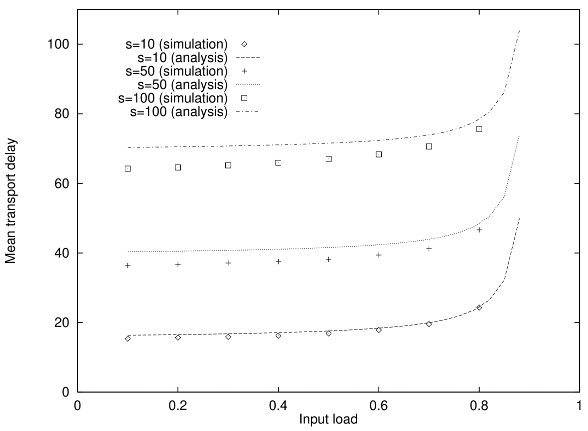

Figure 4: Mean transport delay versus input load.

Fig. 5 shows the mean transport delay as a function of the transition probability from on to off states ($p_{1,0}$) for $s = 10, 50$ ($\rho_s = 0.5$). As $p_{1,0}$ decreases (on period increases), the mean transport delay increases abruptly. For $p_{1,0} > 0.03$ the performance is almost insensitive to the average duration of the on periods.

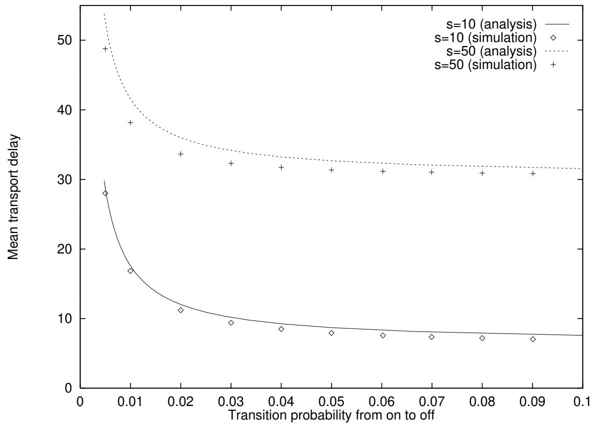

Figure 5: Mean transport delay versus the transition probability from on to off.

Fig. 6 shows the mean resequencing delay as a function of the mean packet error rate (ϵ) for $s = 50, 100$. As discussed earlier, our analysis of the resequencing delay is appropriate only under heavy traffic, i.e., packets are always supplied. For comparison purposes, the figure also includes the curves for the on-off source with $\rho_s = 0.3$ and 0.7 along with the heavy traffic results. As expected, both analytical and simulation results match well as ρ_s increases. It is observed that the gap between the analytical and simulation results increases as the feedback delay and mean packet error rate increase. Interestingly, the shape of the curves is concave.

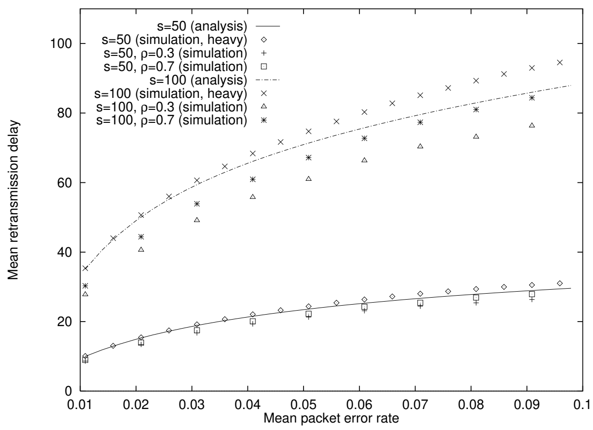

Figure 6: Mean resequencing delay versus ϵ.

Fig. 7 and 8 show the mean end-to-end delay as a function of ϵ for $s = 50$ and $s = 100$, respectively. When $s = 50$ the transmission delay is always greater than the resequencing delay, whereas a crossing point is observed in the case of $s = 100$. In particular, the resequencing delay becomes more significant as the mean packet error rate increases.

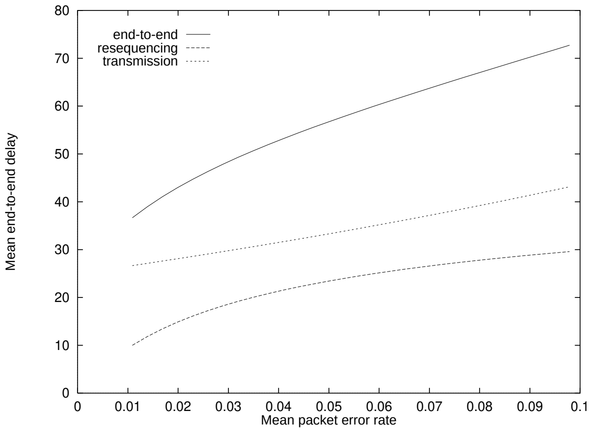

Figure 7: Mean end-to-end delay versus ϵ ($s = 50$).

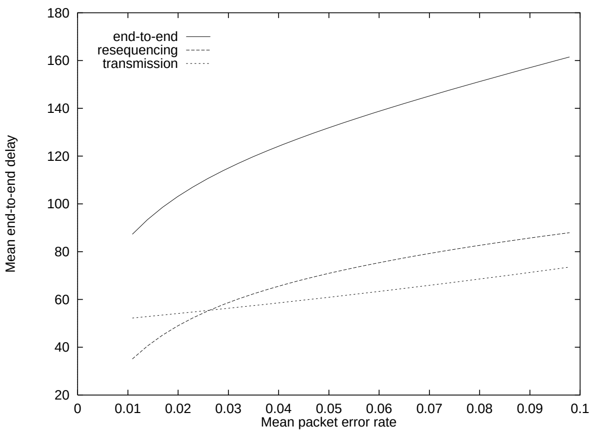

Figure 8: Mean end-to-end delay versus ϵ ($s = 100$).

5 Conclusions

In this paper, we investigated the mean end-to-end delay for a general Markovian source transported over a wireless channel with time-varying error characteristics. An SR ARQ error control protocol was assumed between the transmitter and the receiver. The state of the channel was represented using a two-state Gilbert-Elliot's model. We obtained an approximation for each component of the total mean delay, which consists of queueing, transmission/retransmission, propagation, and resequencing delays. For the queueing delay, our approximation is based on decoupling the dependence of the queueing process at the transmitter from the past history of the channel state. Such approximation becomes exact in the case of "ideal" SR ARQ with zero feedback delay. We obtained the exact probability generating function (PGF) of the queue length under ideal SR ARQ and combined it with the retransmission delay to obtain the mean transport delay. For the resequencing delay, our analysis was performed under heavy-traffic assumptions, hence providing an upper bound on the actual mean resequencing delay. Numerical examples based on the analysis indicate good agreement with simulation results obtained under less stringent assumptions. It was observed that the mean resequencing delay becomes a more significant portion of the total mean delay as the channel conditions deteriorate.

References

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

[2] 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.

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

[4] 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.

[5] 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.

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

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

[8] 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, Jul. 1991.

[9] 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.

[10] 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, Sep. 1990.

[11] 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, Jun. 1991.