# Fluid Analysis of Delay and Packet Discard Performance for QoS Support in Wireless Networks*

Jeong Geun Kim and Marwan Krunz

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

**Abstract**

Providing quality of service (QoS) guarantees over wireless links requires thorough understanding and quantification of the interactions among the traffic source, the wireless channel, and the underlying error control mechanisms. In this paper, we account for such interactions in a network-layer model that we use to investigate the delay and packet discard performance of a wireless channel. We consider a single ON/OFF traffic stream transported over a wireless link. The capacity of this link fluctuates according to a fluid version of Gilbert-Elliot's model. We derive the packet delay distribution via two different approaches: uniformization and Laplace transform. Numerical aspects of both approaches are compared. Using our analytical model, we investigate the packet discard rate at the receiver. This metric is particularly important for delay-sensitive traffic, where the due-date of a packet is translated into a limit on the maximum number of packet retransmissions. The delay distribution is further used to quantify the wireless effective bandwidth under a given delay guarantee. Numerical results and simulations are used to verify the adequacy of our analysis and to study the impact of error control and bandwidth allocation on the packet delay and loss performance.

**keywords:** Wireless networks, QoS, delay distribution, fluid analysis.

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

## 1 Introduction

Current trends in wireless networks indicate a desire to provide a flexible broadband wireless infrastructure that can support emerging multimedia services as well as traditional data services [1, 2, 3]. In such a multi-service wireless environment, quality-of-service (QoS) guarantees are critical for real-time voice and video. Compared to its wireline counterpart, the provisioning of QoS guarantees over wireless links is a more challenging problem whose difficulty stems from the need to explicitly consider the harsh radio-channel transmission characteristics and the underlying link-layer error control mechanisms. This difficulty is further compounded by host mobility and how it impacts the available bandwidth capacity. These difficulties indicate a clear need for a general QoS framework in the wireless environment.

The provisioning of QoS has been extensively explored in wireline networks. Typically, traffic control and resource allocation strategies in these networks are designed with the assumption that the underlying physical media are highly reliable. Since this assumption does not hold for wireless links, applying the same resource allocation strategies of wireline networks to the wireless environment often results in poor performance [4, 5]. For instance, in transporting TCP traffic over wireless links, TCP makes the implicit assumption that packet losses are caused by congestion, although such losses may, in fact, be caused by packet discarding due to channel errors [4, 6]. Eventually, TCP times out and invokes its congestion control mechanism, which unnecessarily reduces the throughput of the wireless channel. This example illustrates how the performance of a end-to-end connection spanning wireline and wireless segments is heavily dependent on the performance of the wireless link.

Two classes of link-level error control are commonly used to improve the performance of a wireless channel: automatic repeat request (ARQ) and forward error correction (FEC) [4]. In general, ARQ is used to deliver data requiring higher reliability, whereas FEC is more suitable for delay-sensitive traffic [4]. Recent studies suggest that hybrid ARQ/FEC might be more appropriate for a wireless network that carries traffic with diverse characteristics and QoS requirements [4, 7]. For instance, data connections with relaxed time constraints can use ARQ, while voice and video connections that require low delay, delay jitter, and minimal packet loss may need a combination of FEC and ARQ with time-constrained retransmission [8].

QoS guarantees in wireless networks can be provided through a coordination between connection-level bandwidth reservation and packet-level scheduling. Most previous research on QoS over wireless networks has mainly focused on these issues. Levine et al. proposed the shadow cluster concept to estimate the bandwidth requirements of a wireless connection [9]. They indicated that this concept can be used in connection admission control to provide a QoS that is specified in terms of the call dropping probability. Reininger et al. identified the high variability of traffic dynamics of mobile multimedia applications as a function of time and space, and proposed a soft QoS control which allows bandwidth renegotiation according to the varying traffic conditions [10]. Capone and Stavrakakis investigated the region of supportable QoS vectors expressed in terms of packet dropping probability [11]. Their work provided insight into the resource management aspects for handling diverse QoS constraints, although the study was limited to unbuffered services. Lu et al. proposed a fair scheduling algorithm with adaptation to wireless networks that take into account bursty and location-dependent channel errors [12]. Although their work identified many practical issues, it did not address the interaction between packet scheduling and error control. In [13], the authors proposed the concept of wireless effective bandwidth that corresponds to the amount of bandwidth required to provide a guaranteed packet loss rate.

The general goals of the underlying work are twofold. First, we study the delay performance over a wireless link and investigate its implication on optimal bandwidth allocation under delay guarantees. Second, we analyze the packet discard rate *at the receiver* under time-constrained retransmissions. This rate is important for delay-sensitive applications that can tolerate some degree of packet loss but that require prior quantification of this loss for use in error concealment and video/voice compression. Our investigations are carried out for a single stream that is transported over a time-varying wireless link. If the link is used to transport more than one connection, then each connection is guaranteed a constant service rate during its active period (i.e., TDMA style). The outcome of packet transmission is determined by the state of the wireless channel and the error control schemes. This scenario encompasses point-to-point connections between mobile terminals (MT) and a base station (BS) in the cellular communication systems.

To achieve our goals, we follow a fluid-based approach whereby the traffic source is modeled by an on-off fluid process and the channel is modeled by a fluid variant of Gilbert-Elliott's model. Using fluid-flow analysis, we compute the delay distribution for a single stream, which is a function of the traffic source, the service rate, the wireless channel, and the error control schemes. To obtain the delay distribution at the transmitter buffer, we first evaluate the queue length distribution taking into account the channel behavior and the underlying error control schemes. We provide two alternative approaches for obtaining the delay distribution via the uniformization and Laplace transform techniques. The two techniques differ in how the accumulative amount of service is determined. Numerical aspects of both approaches are compared. In particular, we derive the closed-form expression for the delay distribution using the uniformization technique. In addition, we derive the packet discard rate which results from imposing a limit on the number of transmissions to meet the delay requirement of the transmitted packet. The analytical results are used to obtain the wireless effective bandwidth under the delay constraints and to investigate the optimal error control strategy that minimizes the use of bandwidth while guaranteeing the QoS. Extensive simulations are conducted to verify the goodness of our analytical results.

The rest of the paper is organized as follows. In Section 2, we describe the wireless link model. Analysis of the delay performance is provided in Section 3. In Section 4, the packet discard

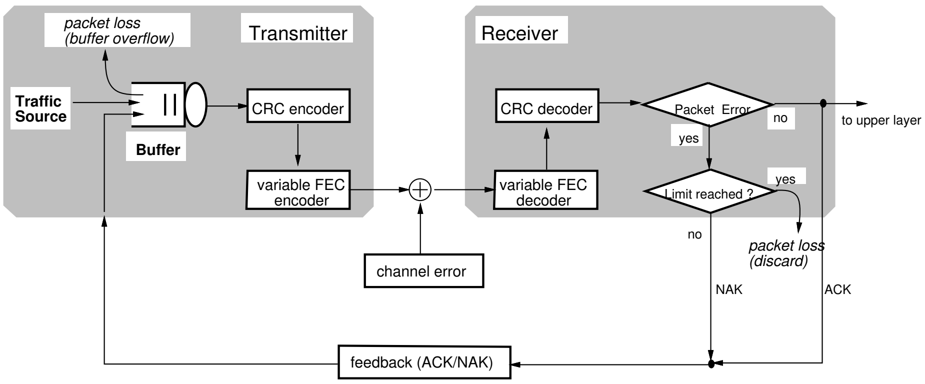


Figure 1: Framework for analyzing the performance over a wireless link.

performance is analyzed. Numerical results and simulations are reported in Section 5, followed by concluding remarks in Section 6.

# 2 Wireless Link Model

## 2.1 Framework

In order to analyze the packet-level performance of a wireless link, we consider the framework shown in Figure 1. In this framework, traffic streams from one or more connections are fed into a finite-size FIFO buffer. A constant service rate $c$ (in packets/second) is assigned to the wireless connection, but the actual drain rate observed at the buffer is reduced due to retransmissions and FEC overhead. The way of estimating the actual service rate will be discussed in the following section. In our study, we consider a particular hybrid ARQ/FEC approach in which the cyclic redundancy check (CRC) code is applied first to a packet, followed by FEC. We assume that the CRC code can alone detect almost all bit errors in a packet. In contrast, only a subset of the errors can be corrected by FEC. In addition, we impose a limit on the number of packet transmissions. Imposing such a limit can be used to provide delay guarantees for real-time traffic. Once a packet hits the limit, it will be discarded. We define the packet discard rate (PDR) as the ratio of packets reaching the limit over the total number of transmitted packets (excluding lost packets due to buffer overflow). For simplicity, we ignore the overhead of the medium access control (MAC) layer.

The above model has three control parameters: the service rate (or assigned bandwidth), the FEC code rate, and the limit on the number of transmissions. These parameters can be adjusted during connection setup to satisfy certain QoS requirements. From the network point of view, the selection of these parameters is very crucial and requires thorough understanding of their impact on the packet-level performance. The main theme of this study is to investigate the packet-level performance of a wireless link as a function of the assigned bandwidth, the limit on transmissions, and error control schemes.

### 2.2 Queueing Model

In this section, we introduce the queueing model that we use for analyzing the packet-level performance for a single traffic stream transported over a wireless link. The traffic source is characterized by an on-off fluid model, with peak rate $r$ and with exponentially distributed on and off periods with means $1/\alpha$ and $1/\beta$, respectively. The wireless channel is modeled using a fluid version of Gilbert-Elliott model, which is often used to investigate the performance of wireless links [14]. As explained in Figure 2, Gilbert-Elliott model is Markovian with two alternating states: *Good* and *Bad*. The bit error rates (BER) during the Good and Bad states are given by $P_{eg}$ and $P_{eb}$, respectively, where $P_{eg} \ll P_{eb}$. The durations of the Good and Bad states are exponentially distributed with means $1/\delta$ and $1/\gamma$, respectively.

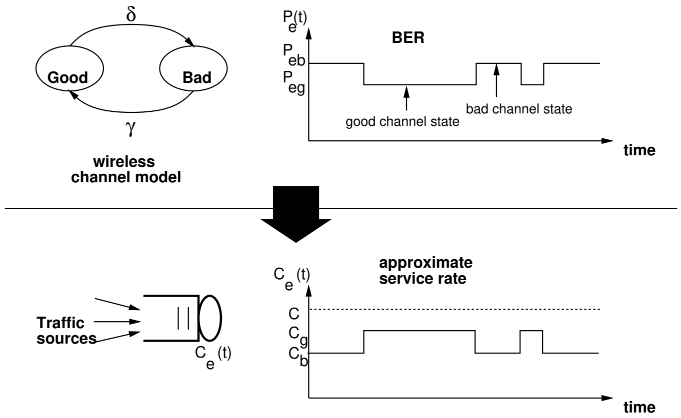


Figure 2: Wireless channel model and corresponding service rate model.

The FEC capability in the underlying hybrid ARQ/FEC mechanism is characterized by three parameters: the number of bits in a code block ($n$), the number of payload bits ($k$), and the maximum number of correctable bits in a code block ($\tau$). Note that $n$ consists of the $k$ payload bits and the extra parity bits. The FEC code rate $e(\tau)$ is defined as

$$e(\tau) = \frac{k}{n(\tau)}.$$

Assuming that a FEC code can correct up to $\tau$ bits and that bit errors during a given channel state are independent, the probability that a packet contains a non-correctable error is given by:

$$P_c(p_b, \tau) = \sum_{j=\tau+1}^{n(\tau)} \binom{n(\tau)}{j} p_b^j (1 - p_b)^{n(\tau)-j} \tag{1}$$

where $p_b$ is the bit error probability; $p_b \in \{P_{eg}, P_{eb}\}$. To account for the FEC overhead, we obtain the actual service rate $c_e$ observed at the output of the buffer:

$$c_e = c \cdot e(\tau) \tag{2}$$

where $c$ is the bandwidth assigned to the connection.

The *exact* behavior of ARQ and FEC in the underlying queueing model is difficult to analyze. To obtain analytically tractable results, we assume that the packet departure process follows a fluid process with a service rate that is modulated by the channel state (see Fig. 2). This approximation implies that there are two deterministic service rates: $c_g$ during Good states and $c_b$ during Bad states. We assume that the feedback delay for sending an acknowledgment from the receiver to the sender is negligible, thus packets are always delivered to the receiver in sequence[1]. A packet is successively retransmitted until it is received correctly at the destination or until the limit on the number of transmissions is reached. In this scenario, the total time needed to successfully transmit a packet conditioned on the channel state follows a truncated geometric distribution. Let $N_{tr}$ denote the number of transmissions until the packet is successfully received or is discarded due to the limit on transmissions. For a given packet error probability $P_c$ and a limit on transmissions $N_l$, the expected value of $N_{tr}$ is given by:

$$E[N_{tr}] = \frac{1 - P_c^{N_l}}{1 - P_c}. \tag{3}$$

Thus, $c_g$ and $c_b$ correspond to the mean transmission rates of the truncated geometric trial with parameters $(P_{c,g}, N_l)$ and $(P_{c,b}, N_l)$, respectively, where $P_{c,g}$ and $P_{c,b}$ are the packet error probabilities in Good and Bad states, respectively, given by (1). Formally,

$$c_g = \frac{c \cdot e(\tau) \cdot (1 - P_{c,g})}{1 - P_{c,g}^{N_l}} \tag{4}$$

$$c_b = \frac{c \cdot e(\tau) \cdot (1 - P_{c,b})}{1 - P_{c,b}^{N_l}}. \tag{5}$$

where $P_{c,g} = P_c(P_{e,g}, \tau)$ and $P_{c,b} = P_c(P_{e,b}, \tau)$.

[1]Consider a time-division multiple access (TDMA) protocol which is providing multiple slots to a MT based on the required service rate. In such a scenario, the service discipline of the constant service rate to MT, which is assumed in our framework, can be implemented by assigning the TDMA slots to the MT periodically during a frame (or across frames). In this case, the period of packet scheduling is longer than the packet transmission time, thus avoiding out-of-sequence delivery of packets.

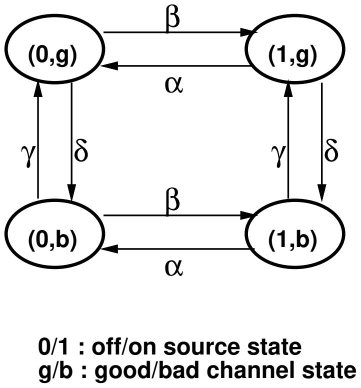


Figure 3: State transition diagram.

# 3 Analysis of Delay Performance

## 3.1 Queue Length Distribution

Following the discussion in the previous section, we construct the Markovian queueing system with four states as shown in Figure 3. Let $S$ denote the state space. Thus,

$$S = \{(0,g),(0,b),(1,g),(1,b)\} \tag{6}$$

where 0 and 1 denote the on and off states of a traffic source, respectively, and $g$ and $b$ denote Good and Bad channel states, respectively.

Following a standard fluid approach (see [15], for example), the evolution of the buffer content can be described by the following differential equation:

$$\frac{d\mathbf{\Pi}(x)}{dx}\boldsymbol{D} = \mathbf{\Pi}(x)\boldsymbol{M} \tag{7}$$

where $\boldsymbol{D} \triangleq \mathrm{diag}[-c_g, -c_b, r-c_g, r-c_b]$, $\mathbf{\Pi}(x) \triangleq [\ \Pi_{0,g}(x) \quad \Pi_{0,b}(x) \quad \Pi_{1,g}(x) \quad \Pi_{1,b}(x)\ ]$, $\Pi_s(x) \triangleq \Pr\{\text{buffer content} \le x \text{ and the system } s \in S\}$, and $\boldsymbol{M}$ is the generator matrix of the underlying Markov chain:

$$\boldsymbol{M} = \begin{bmatrix} -(\beta+\delta) & \delta & \beta & 0 \\ \gamma & -(\beta+\gamma) & 0 & \beta \\ \alpha & 0 & -(\alpha+\delta) & \delta \\ 0 & \alpha & \gamma & -(\alpha+\gamma) \end{bmatrix}.$$

Throughout the paper, matrices and vectors are boldfaced.

The solution of (7) corresponds to the solution of the eigenvalue/eigenvector problem:

$$z\boldsymbol{\phi}\boldsymbol{D} = \boldsymbol{\phi}\boldsymbol{M} \tag{8}$$

which is generally given by

$$\boldsymbol{\Pi}(x) = \sum_{z_i \leq 0} a_i \exp(z_i x)\phi_i \tag{9}$$

where $a_i$'s are constant coefficients and the pairs $(z_i, \boldsymbol{\phi}_i), i = 1, 2, \cdots,$ are the eigenvalues and the right eigenvectors of the matrix $\boldsymbol{M}\boldsymbol{D}^{-1}$ [15, 16]. Let $\boldsymbol{w}$ denote the stationary probability vector of the Markov chain; $\boldsymbol{w}$ satisfies $\boldsymbol{w}\boldsymbol{M} = 0$ and $\boldsymbol{w}\mathbf{1} = 1$, where $\mathbf{1}$ is a column vector of ones. Then $\boldsymbol{w}$ is given by:

$$\boldsymbol{w} = \frac{1}{(\alpha+\beta)(\delta+\gamma)} \left[ \begin{array}{cccc} \alpha\gamma & \alpha\delta & \beta\gamma & \beta\delta \end{array} \right]. \tag{10}$$

In order to solve (8), we follow the approach used in [16]. The four-state Markov process is decomposed into two processes; one describes the on-off source and the other describes the state of the channel. We skip the detailed description on derivation which is available in [13, 17].

After obtaining the eigenvalues, the eigenvectors, and the coefficients, we can construct the stationary buffer content distribution $\boldsymbol{\Pi}(x)$. Consequently, the packet loss rate due to buffer overflow $G(x)$ is given by:

$$G(x) \;=\; 1 - \mathbf{1}\boldsymbol{\Pi}(x). \tag{11}$$

In the following section, $\boldsymbol{\Pi}(x)$ is used to obtain the delay distribution.

### 3.2 Delay Distribution

In fluid queueing models with an error-free channel and constant service rate, e.g., ATM link, the packet delay distribution can be directly obtained from the queue length distribution [18]. However, the scenario we consider in this study includes a time-varying wireless channel that is being approximated by a two-state Markov modulated fluid process. In this case, the packet delay distribution is much more difficult to obtain since one has to take into account the time-varying service rate as well as the queue length.

We assume an infinite-capacity buffer. Let $D$ denote the delay experienced by an arriving packet. Let $C(t)$ denote the accumulative amount of service during a period of length $t$:

$$C(t) \;\triangleq\; \int_0^t c(s)ds$$

where $c(s)$ is the service rate at the time $s$. The channel state at time $t$ is denoted by $h(t) \in \{g, b\}$, where $g$ and $b$ denote Good and Bad states, respectively. The probability that the delay seen by a packet is less than or equal to $t$ is equal to the probability that $C(t)$ is greater than or equal to the queue length $Q_0$ at the instant of the packet arrival. Thus, we have

$$\begin{aligned} \Pr[D \le t] &= \Pr[C(t) \ge Q_0] \\ &= \sum_{i \in S} \int_{0-}^{\infty} \Pr[C(t) \ge x | i, Q_0 = x] \pi_i(x) dx \\ &= \frac{r}{T} \int_{0-}^{\infty} \Pr[C_g(t) \ge x] \pi_{1,g}(x) + \Pr[C_b(t) \ge x] \pi_{1,b}(x) dx \end{aligned} \tag{12}$$

where $T$ is the throughput, $\pi_i$ is the pdf of the queue length in a state $i$, $i \in S$, and

$$C_i(t) \triangleq \int_0^t c(s) ds \quad \text{given } h(0) = i, \text{ for } i \in \{g, b\}.$$

The quantity $r\pi(x)/T$ represents the fraction of carried flow that arrives at the queue when its content is $x$. For an infinite-capacity buffer, the throughput $T$ is given by:

$$T = r(w_{1,g} + w_{1,b}). \tag{13}$$

In order to obtain $\Pr[C_i(t) \le x]$, $i \in \{g, b\}$, we provide two methods: direct calculation using Laplace transform and uniformization. The equivalence of these approaches will be verified using numerical examples.

### Laplace Transform Approach

For numerical convenience, we transform the random variable $C_i(t)$ to $\bar{C}_i(t)$ defined as:

$$\bar{C}_i(t) \triangleq C_i(t) - c_b t, \quad i \in \{g, b\}.$$

By this transformation, $\bar{C}_i(t)$ is the accumulative service resulting from a "normalized" channel with service rates $c_g - c_b$ (during Good states) and 0 (during Bad states). Note that the minimum amount of accumulative service in a period of length $t$ is $c_b t$. Thus,

$$\Pr[C_i(t) \ge x] = \begin{cases} 1, & \text{if } x < c_b t \\ 1 - \Pr[\bar{C}_i(t) \le x - c_b t], & \text{if } x \ge c_b t. \end{cases}$$

The following proposition gives the probabilities $\Pr[C_g(t) \ge x]$ and $\Pr[C_b(t) \ge x]$ by solving the partial differential equations (PDE) for the consumption rate $C(t)$.

**Proposition 3.1** *The probabilities* $\Pr[C_i(t) \geq x], i \in \{g, b\}$ *when* $x \geq c_b t$, *are given by*

$$\Pr[C_g(t) \geq x] = e^{-\delta \hat{x}} \left( e^{-\gamma(t-\hat{x})} J_0(2\sqrt{-\delta\gamma\hat{x}(t-\hat{x})}) + \sum_{n=0}^{\infty} \frac{\delta \hat{x}}{(n!)^2} \Gamma(n+1, \gamma(t-\hat{x})) \right) \quad (14)$$

$$\Pr[C_b(t) \geq x] = e^{-\delta \hat{x}} \sum_{n=0}^{\infty} \frac{\delta \hat{x}}{(n!)^2} \Gamma(n+1, \gamma(t-\hat{x})) \quad (15)$$

*where* $\hat{x} = (x - c_b t)/(c_g - c_b)$, $J_0(z)$ *is the Bessel function of order* 0 *given by*

$$J_0(z) = \sum_{n=0}^{\infty} \frac{(-1)^n (z/2)^{2n}}{(n!)^2},$$

*and* $\Gamma(n, z)$ *is the incomplete gamma function given by:*

$$\Gamma(n, z) = \int_0^z e^{-x} x^{n-1} dx.$$

**Proof.** See Appendix A.

Equation (15) in the Proposition 3.1 is substituted into (12) to evaluate the delay distribution. Numerical complexity of Equation (14) and (15) is associated with the infinite sum in them. We observe that the value of the infinite sum converges fast with moderate iteration, e.g., $n = 20$.

### Uniformization Approach

As a second approach to obtaining $\Pr[C_i(t) \leq x]$, $i \in \{g, b\}$, we use the uniformization approach. In continuous-time Markov chains, uniformization is a technique for uniformizing the transition rates between states by introducing transitions from a state to itself [19]. This technique facilitates the analysis of the problem in this study. Refer to [19] for details.

Let $t_g$ and $t_b$ denote the accumulative sojourn times of Good and Bad channel states during an interval of length $t$, respectively. That is,

$$t_g \triangleq \int_0^t \mathbf{1}_{\{h(s)=g\}} ds$$
$$t_b \triangleq \int_0^t \mathbf{1}_{\{h(s)=b\}} ds.$$

Then, the accumulative service $C(t)$ is given by

$$C(t) = c_g t_g + c_b t_b, \quad 0 \leq t_g, t_b \leq t. \quad (16)$$

Since $t = t_g + t_b$, $C(t)$ can be expressed as

$$C(t) = c_g t_g + c_b(t - t_g) \quad \text{or} \quad C(t) = c_g(t - t_b) + c_b t_b. \tag{17}$$

In Appendix B, we provide the probability distribution of $t_g$ and $t_b$ conditioned on the channel state. Then, the probability $\Pr[C_i(t) \geq x]$, $i \in \{g, b\}$ can be directly obtained from (17).

**Proposition 3.2** *The probabilities* $\Pr[C_g(t) \geq x]$ *and* $\Pr[C_b(t) \geq x]$ *are given by*

$$\Pr[C_g(t) \geq x] = 1 - e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \chi^i (1-\chi)^{n-i} \tag{18}$$

*and*

$$\Pr[C_b(t) \geq x] = e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \chi^{n-i} (1-\chi)^{i} \tag{19}$$

*where* $\chi = \frac{x - c_b t}{(c_g - c_b)t}$.

**Proof.** See Appendix B.

The equations in Proposition 3.2 also have numerical issues due to the presence of multiple sums. The triple sums in the equations cause significant complexity. However, observing the duplicate computation in the last sum for consecutive indexes $k$'s, we can save the computation time. Figure 4 shows numerical examples for the equations in Proposition 3.1 and 3.2. This example is for the service rate $c = 1000$ (packets/sec), $t = 0.5, 1.0$ (sec), and other parameters taken from the default values in Table 2. The exact match is observed over the range of values.

Up to this point, we have discussed the alternative approaches to obtaining the probability $\Pr[C_i(t) \geq x], i \in \{g, b\}$. In the following, the results from Proposition 3.2 are used to obtain the delay distribution. The following Proposition gives the closed-form expression for delay distribution by substituting (18) and (19) into (12).

**Proposition 3.3**

$$\begin{aligned}
\Pr[D \leq t] = {} & \frac{r}{T}(\Pi_{1,g}(c_g t) + \Pi_{1,b}(c_b t)) - \frac{r}{T} e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \\
& \cdot \sum_{i=k}^{n} (c_g - c_b)t \sum_{l} a_l e^{z_l c_b t} \Phi(i+1; n+2; z_l(c_g - c_b)t) \\
& + \frac{r}{T} e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \\
& \cdot \sum_{i=k}^{n} (c_g - c_b)t \sum_{m} a_m e^{z_m c_b t} \Phi(n-i+1; n+2; z_m(c_g - c_b)t)
\end{aligned} \tag{20}$$

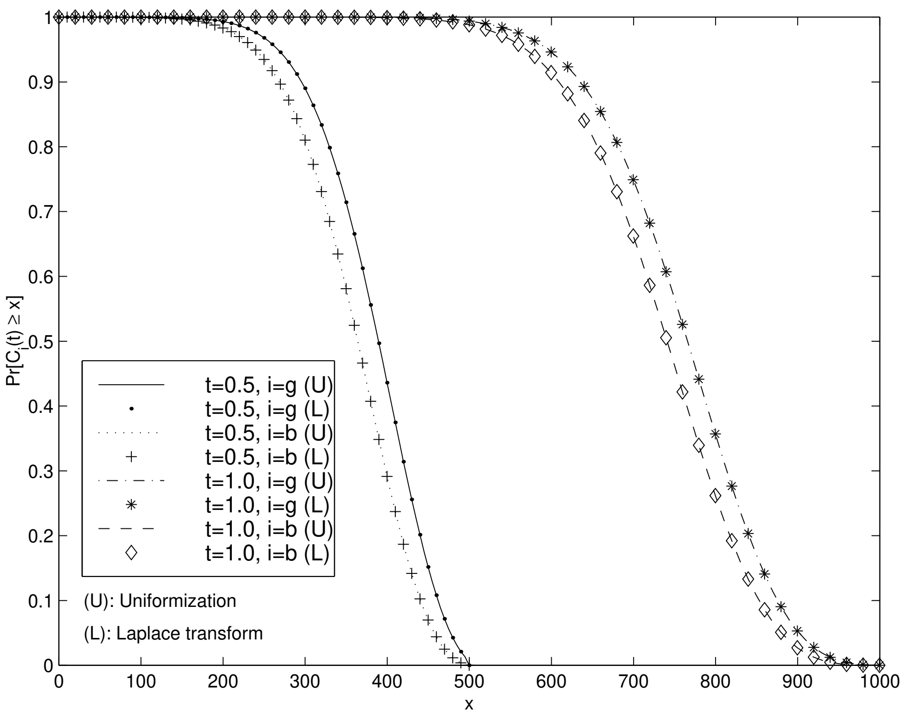


Figure 4: Accumulative service distribution based on uniformization and Laplace transform approaches.

*where* $\pi_{1,g}(x) = \sum_l a_l e^{z_l x}$, $\pi_{1,b}(x) = \sum_m a_m e^{z_m x}$, *the indexes* $l, m$ *are used to index the negative eigenvalues, and*

$$\Phi(x; y; z) \quad \triangleq \quad \sum_{k=0}^{\infty} \frac{(x)_k}{(y)_k} \frac{z^k}{k!}$$

*with* $(a)_n \triangleq a(a+1)\cdots(a+n-1)$.

**Proof.** Refer to Appendix C.

Since Equation (20) has similar numerical structure as the expressions in Proposition 3.2, we can save the computation time by avoiding the duplicate sum as mentioned previously. We also obtained the delay distribution using Proposition 3.1. However, it generated the multiple sums and is not advantageous over the uniformization approach. Thus, we use numerical integration to obtain the delay distribution for Proposition 3.1. Most numerical results in Section 5 are based on Proposition 3.3 whereas the numerical integration using Proposition 3.1 is also used to verify the accuracy.

**Wireless Effective Bandwidth**

The notion of effective bandwidth has been employed to achieve efficient provisioning of QoS guarantees in ATM networks [20]. In this study, we extend the notion for a wireless connection under probabilistic delay constraints.

We defined the *wireless effective bandwidth* $c_{eb}$ under the delay constraint $\Pr[\text{delay} > t] = \varepsilon$ by

$$c_{eb} \overset{\triangle}{=} \min\{c | c \text{ satisfies } \Pr[\text{delay} > t] = \varepsilon\} \tag{21}$$

where $c$ is the service rate. In contrast to wireline effective bandwidth, the wireless effective bandwidth is configured along with the optimal number of correctable bits which minimizes the use of bandwidth while providing the requested reliability at the physical link. In Section 5, we provide the numerical examples on this issue and investigate the characteristic of the pair of QoS parameters $(c_{eb}, \tau)$ in detail.

# 4 Packet Discard Rate

In the scenario we consider, a packet is discarded when the number of transmissions reaches its maximum limit $N_l$. Such a limit is determined based on the due-date (delay requirement) of the transmitted packet. Let $p(n|i)$ denote the probability of $n$ consecutive transmission failures given that the channel state is $i$ at the first attempt of packet transmission, where $i \in \{g, b\}$. Observing the recursive structure between the consecutive transmissions, we obtain the following relation

$$\begin{aligned} p(n|g) &= p(n-1|g)P_{g,g}(\hat{t})P_{c,g} + p(n-1|b)P_{g,b}(\hat{t})P_{c,g} \\ p(n|b) &= p(n-1|g)P_{b,g}(\hat{t})P_{c,b} + p(n-1|b)P_{b,b}(\hat{t})P_{c,b} \end{aligned}$$

where $P_{i,j}(\hat{t}), i, j \in \{g, b\}$ is the probability that the channel state changes from $i$ to $j$ in time $\hat{t}$. The quantity $\hat{t}$ here corresponds to the turnaround time of the packet transmission. The right-hand side of the first equation accounts for the events of the first transmission failure at Good channel state followed by $n-1$ consecutive transmission failures after the channel transition into Good and Bad, respectively. The second equation can be explained in a similar way except that the initial channel state is Bad.

Rearranging the previous equations in a matrix form, we have

$$\begin{bmatrix} p(n|g) \\ p(n|b) \end{bmatrix} = \begin{bmatrix} P_{g,g}(\hat{t})P_{c,g} & P_{g,b}(\hat{t})P_{c,g} \\ P_{b,g}(\hat{t})P_{c,b} & P_{b,b}(\hat{t})P_{c,b} \end{bmatrix} \begin{bmatrix} p(n-1|g) \\ p(n-1|b) \end{bmatrix} \tag{22}$$

with the initial condition

$$\begin{bmatrix} p(1|g) \\ p(1|b) \end{bmatrix} = \begin{bmatrix} P_{c,g} \\ P_{c,b} \end{bmatrix}.$$

The probabilities $P_{c,g}$ and $P_{c,b}$ are given in (1).

Solving the previous equations recursively, we obtain the packet discarding rate ($p_d$) under the limit on the retransmission ($N_l$):

$$p_d = \begin{bmatrix} \hat{w}_g & \hat{w}_b \end{bmatrix} \begin{bmatrix} P_{g,g}(\hat{t})P_{c,g} & P_{g,b}(\hat{t})P_{c,g} \\ P_{b,g}(\hat{t})P_{c,b} & P_{b,b}(\hat{t})P_{c,b} \end{bmatrix}^{N_l-1} \begin{bmatrix} P_{c,g} \\ P_{c,b} \end{bmatrix} \tag{23}$$

where $\hat{w}_g$ and $\hat{w}_b$ are the steady-state probabilities that the channel state is Good and Bad, respectively, as seen by the packet at HOL in its first attempt of transmission:

$$\hat{w}_i \stackrel{\triangle}{=} \Pr[\text{the channel state is } i \text{ seen by a packet at HOL}] \tag{24}$$
$$= \text{the fraction of flow drained at channel state } i. \tag{25}$$

The probability $\hat{w}_i$ is different from $w_i$, and it plays an important role in estimating the packet drop rate as will be discussed later.

Since the flow drains continuously as long as the queue is non-empty, the fraction of flow drained at channel state $i$, $i \in \{g, b\}$ corresponds to the probability that the queue is non-empty multiplied by a fraction of the service rate over the throughput. Thus, we obtain

$$\hat{w}_g = \frac{c_g}{T}(w_{0,g} - \Pi_{0,g}(0) + w_{1,g} - \Pi_{1,g}(0)) \tag{26}$$
$$\hat{w}_b = \frac{c_b}{T}(w_{0,b} - \Pi_{0,b}(0) + w_{1,b} - \Pi_{1,b}(0)). \tag{27}$$

And, $P_{i,j}(t)$, $i, j \in \{g, b\}$ can be obtained by using Kolmogorov's equation [19]. Thus, we have

$$\begin{aligned} P_{g,g}(t) &= \tfrac{\gamma}{\delta+\gamma} + \tfrac{\delta}{\delta+\gamma}e^{-(\delta+\gamma)t} \\ P_{b,b}(t) &= \tfrac{\delta}{\delta+\gamma} + \tfrac{\gamma}{\delta+\gamma}e^{-(\delta+\gamma)t} \\ P_{g,b}(t) &= 1 - P_{g,g}(t) \\ P_{b,g}(t) &= 1 - P_{b,b}(t). \end{aligned} \tag{28}$$

Substituting the expressions in (27) and (28) into (23), we obtain the packet discard rate $p_d$ expressed in a closed form:

**Proposition 4.1**

$$\begin{aligned} p_d &= \hat{w}_g P_{c,g}(\alpha_0 + \alpha_1(P_{g,g}(\hat{t})P_{c,g} + P_{g,b}(\hat{t})P_{c,b})) \\ &\quad + \hat{w}_b P_{c,b}(\alpha_0(1 + P_{b,g}(\hat{t})P_{c,g}) + \alpha_1 P_{b,b}(\hat{t})P_{c,b}) \end{aligned} \tag{29}$$

*where*

$$\alpha_0 = \frac{\lambda_1\lambda_2(\lambda_2^{N_l-2} - \lambda_1^{N_l-2})}{\lambda_1 - \lambda_2} \tag{30}$$

$$\alpha_1 = \frac{\lambda_1^{N_l-1} - \lambda_2^{N_l-1}}{\lambda_1 - \lambda_2} \tag{31}$$

*and* $\lambda_{1,2}$ *is the eigenvalues of the square matrix in (23). That is,*

$$\begin{aligned} \lambda_{1,2} = & \frac{P_{g,g}(\hat{t})P_{c,g} + P_{b,b}(\hat{t})P_{c,b}}{2} \\ & \pm \frac{\sqrt{(P_{g,g}(\hat{t})P_{c,g} + P_{b,b}(\hat{t})P_{c,b})^2 + 4(1 - P_{g,g}(\hat{t}) - P_{b,b}(\hat{t}))P_{c,g}P_{c,b}}}{2}. \end{aligned} \tag{32}$$

**Proof.** By matrix diagonalization, the powers of the square matrix in (23) are calculated. The proof is omitted.

## 5 Numerical Results and Discussion

In this section, we present numerical examples of our analytical results. We verify the adequacy of these results by contrasting them against more realistic simulations.

Similar to the analysis, the simulation results are obtained using on-off traffic sources with exponentially distributed on and off periods. The ARQ retransmission process is simulated in a more realistic manner, whereby a packet is transmitted repeatedly until it is received with no errors or until it reaches the limit on the number of transmissions. The probability of a packet error is computed from (1) for both channel states. Transitions between Good and Bad states are assumed to occur only at the beginning of a packet transmission slot. A packet is retransmitted if it has uncorrectable errors. It is assumed that the propagation delay is small, so that the ACK/NAK message for a packet is received at the sender before the next attempt of transmission. Finally, we use an infinite-capacity buffer in our simulations.

In our experiments, we vary the BER during the Bad state ($P_{eb}$), and we fix the BER during the Good state at $P_{eg} = 10^{-6}$. We set the mean of the off period to ten times that of the on period. In addition, we take the parameters related to the wireless channel from [14]. We adopt Bose-Chaudhuri-Hocquenghem (BCH) code [21] for FEC . We consider fixed packet sizes, e.g., ATM cell. Since we treat the CRC code as part of the payload, the FEC code is applied to 424-bit blocks (i.e., $k = 424$ bits). Table 1 shows the size, the code rate, and the number of correctable bits of the BCH code used in our examples. All simulation results are reported with 95% confidence intervals. For the delay distribution, $10^7$ to $4 \times 10^7$ samples were needed in the simulations. Table 2 summarizes the values of the various parameters in the simulations and numerical examples. For the parameters $c$, $P_{eb}$, $\tau$, and $N_l$, the values in the parenthesis are assumed unless specified otherwise.

| $(n, k)$ | $\tau$ | $e(\tau)$ |
|---|---|---|
| $(442, 424)$ | 2 | 0.959 |
| $(451, 424)$ | 3 | 0.940 |
| $(460, 424)$ | 4 | 0.921 |
| $(469, 424)$ | 5 | 0.904 |
| $(478, 424)$ | 6 | 0.887 |

Table 1: FEC code rate versus number of correctable bits ($\tau$) using BCH scheme.

| Parameter | Symbol | Value |
|---|---|---|
| source peak rate | $r$ | 1 Mbps (or 2604.1667 packets/sec) |
| service rate | $c$ | 100 – 8000 (packets/sec) (1000) |
| mean on period | $1/\alpha$ | 0.02304 sec |
| mean off period | $1/\beta$ | 0.2304 sec |
| mean Good channel period | $1/\delta$ | 0.1 sec |
| mean Bad channel period | $1/\gamma$ | 0.0333 sec |
| BER in Good channel state | $P_{eg}$ | $10^{-6}$ |
| BER in Bad channel state | $P_{eb}$ | $10^{-2} - 10^{-5}$ ($10^{-2}$) |
| number of correctable bits | $\tau$ | 0 – 20 (7) |
| Limit on transmissions | $N_l$ | $1 - \infty$ ($\infty$) |

Table 2: Parameter values used in the simulations and numerical results.

Figure 5 depicts the complementary cumulative distribution for delay Pr[delay $> t$]. We vary the service rates ($c$) from 800 to 1200 (packets/sec) while fixing other parameters, $P_{eb} = 10^{-2}, \tau = 7$, and $N_l = \infty$. The difference between the analytical and simulation results is negligible for all service rates. We observe a slight deviation at the tail part of the distribution. However, it is associated with the number of samples taken from the simulation. For $c = 1200$, we generated $4 \times 10^7$ packets to obtain the shown results. Note that the simulation is based on the realistic scenario in which the packet is transmitted until it is successfully transmitted or until it reaches the limit on the retransmission, whereas the analytical results are based on the fluid analysis.

Figure 6 shows the complementary delay distribution for different values of $\tau$ (number of correctable bits). The error correction capability increases from $\tau = 0$ (ARQ) to $\tau = 7$. For all $\tau$'s, the analytical and simulation results match well. This figure indicates the use of FEC improves the transport performance over wireless links despite its overhead. This issue will be discussed later in detail.

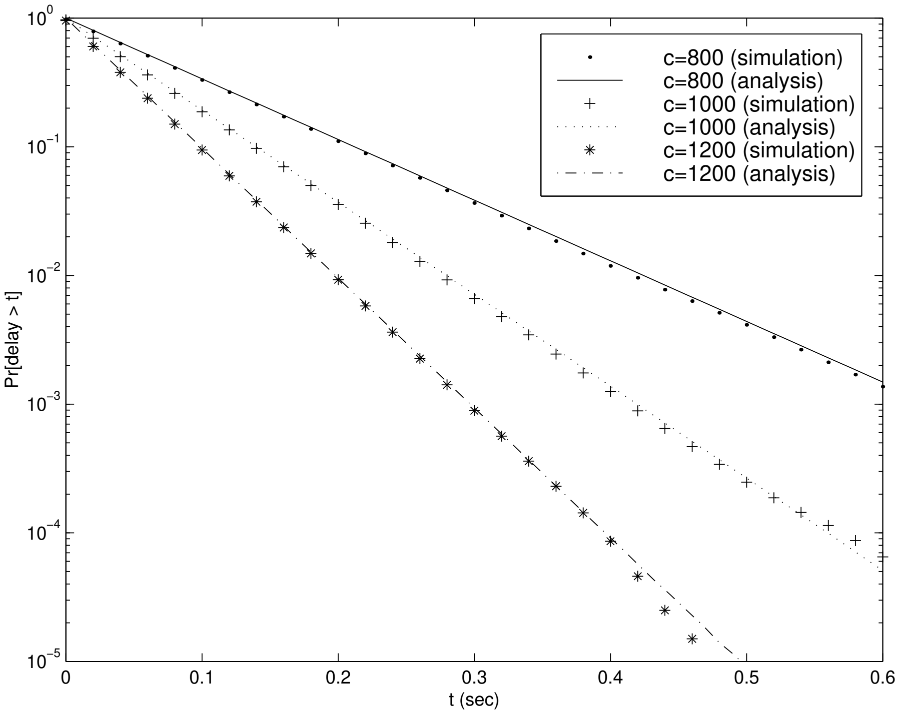


Figure 5: Complementary delay distribution for different service rates.

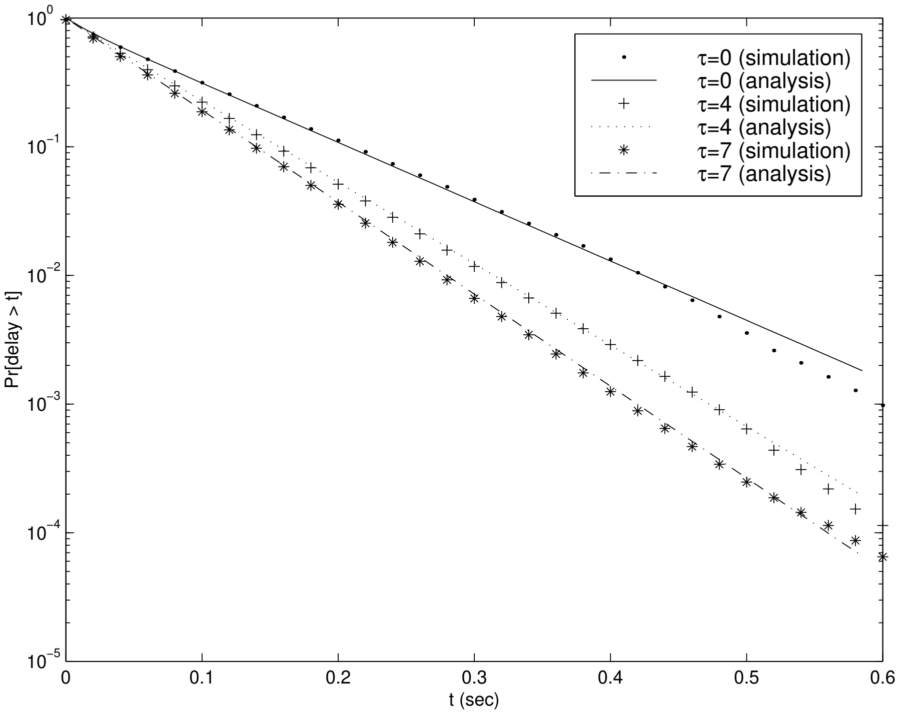


Figure 6: Complementary delay distribution for different values of $\tau$.

Figure 7 shows the complementary delay distribution as a function of the limit on the number of transmissions ($N_l$). As discussed earlier, such a limit is needed to satisfy the due-date of the transmitted packet. For $N_l = 1, 6$, and 60, a highly acceptable agreement is observed between the analytical and simulation results. Note that our delay analysis does not take into account the loss due to the packet discard, assuming a very small discard rate such as much less than 0.01, whereas it is captured in the queue length distribution. However, even the case of the limit $N_l = 1$, corresponding to the packet discard rate $p_d = 0.0298$, shows a good estimation of our analysis. As shown in the figure, the delay distribution varies significantly for different $N_l$'s. This indicates the usefulness of $N_l$ as a control parameter for satisfying the delay and loss requirements simultaneously since it can play a role of a trade-off parameter between them.

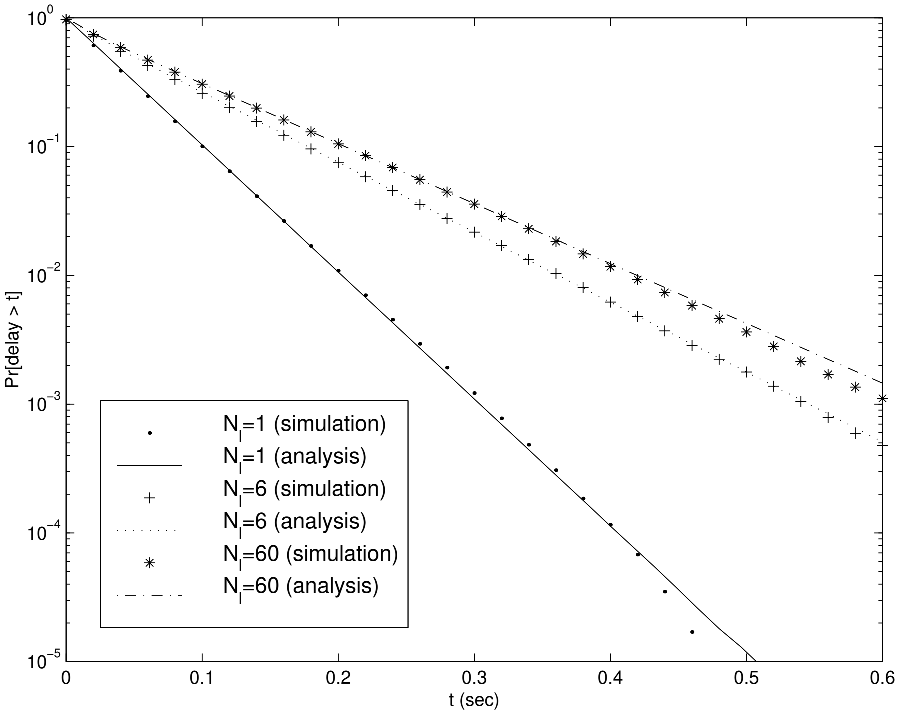


Figure 7: Complementary delay distribution for different values of $N_l$.

Figure 8 shows the effective bandwidth as a function of the number of correctable bits ($\tau$) for the three target delay constraints $\Pr[\text{delay} > 0.01] = 0.01, 0.05, 0.1$. Expectedly, more bandwidth is needed as a higher quality (or less $\Pr[\text{delay} > 0.01]$) is required. In particular, we observe that much higher bandwidth is required when only ARQ ($\tau = 0$) is used as an error control scheme. Thus, the appropriate use of FEC is essential in an efficient use of scarce wireless bandwidth. In addition, this figure clearly indicates that there is an optimal $\tau$ ($\tau = 7$ in this example) for a given BER $P_{eb} = 0.01$, which satisfies a delay QoS constraint while minimizing the use of bandwidth.

The optimal number of correctable bits as a function of the BER in Bad channel state is shown in Fig. 9. The target delay constraint is fixed to $\Pr[\text{delay} > 0.01] = 0.25$. We vary the BER from

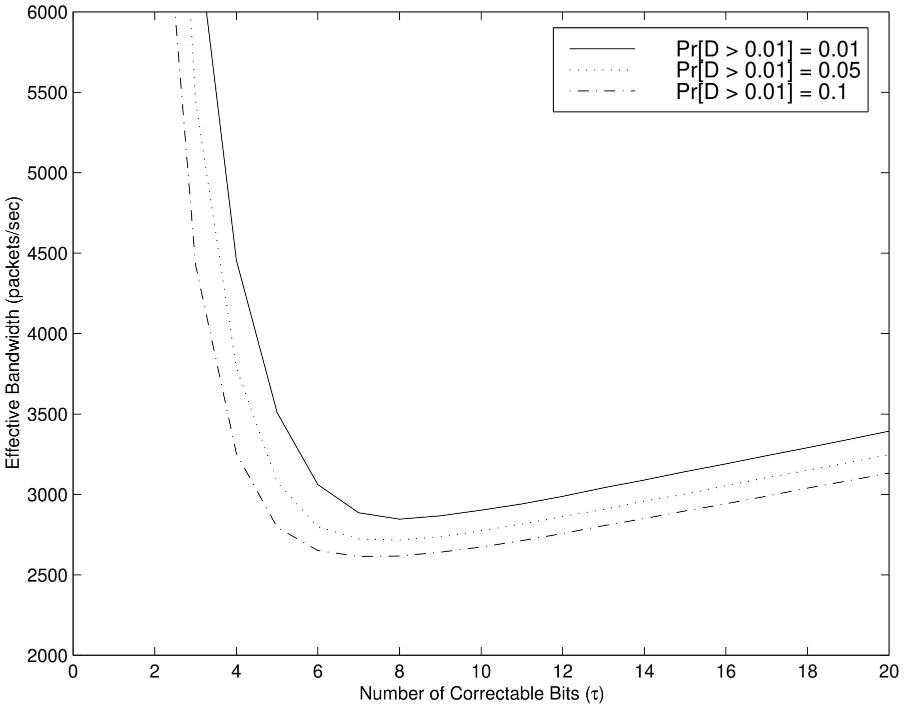


Figure 8: Effective bandwidth versus $\tau$ for target delay constraints Pr[delay $> 0.01$].

$P_{e,b} = 0.001$ to $P_{e,b} = 0.0158$. For each BER, we observe the optimal number of correctable bits exists. To better pictorial view, we provide the effective bandwidth as a function of $\tau$'s and BER's ($P_{e,b}$) altogether in Fig. 10.

Figure 11 shows the effective bandwidth versus the target delay constraints specified in terms of Pr[delay $> t$]. Each curve in the figure corresponds to four values of $t = 0.001, 0.005, 0.01, 0.1$ (sec). As expected, more bandwidth is required for a more stringent quality requirement, i.e., less Pr[delay $> t$] for a fixed $t$. Notice that the minimum and maximum of the effective bandwidth are $c = 282$ and $c = 3250$ (packets/sec), respectively. The minimum effective bandwidth corresponds the minimum service rate satisfying the stability condition of the queue, whereas the maximum is the minimum service rate satisfying the condition $c_b > r$, i.e., $\min\{c|c_b > r\}$. Considering the source peak rate $r = 2604.1667$ (packets/sec), we observe the maximum effective bandwidth is greater than the source peak rate. This is against the results in wireline effective bandwidth indicating the effective bandwidth is bounded by the source peak rate. In the wireless scenario we consider, the assigned service rate is reduced due to the packet error and FEC, the principle in wireline effective bandwidth cannot be thus applied to the wireless case. Another interesting facts observed in this figure is associated with the shape of the curves with decreasing Pr[delay $> t$]. Consider the case of $t = 0.001$. To improve the QoS from Pr[delay $> 0.001] = 1.0$ to Pr[delay $> 0.001] = 0.8$, we need to assign an extra bandwidth 1950.0 (packets/sec), whereas we need 143.0 (packets/sec) for $t = 0.1$. That is, the behavior of the effective bandwidth depends on the selection

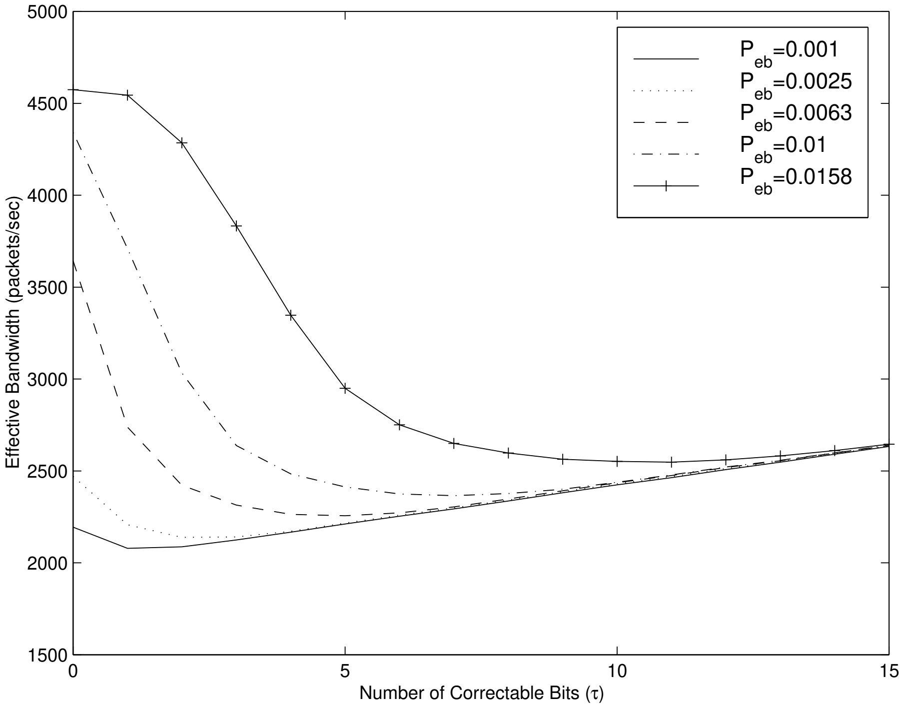


Figure 9: Effective bandwidth versus $\tau$ for different BER's.

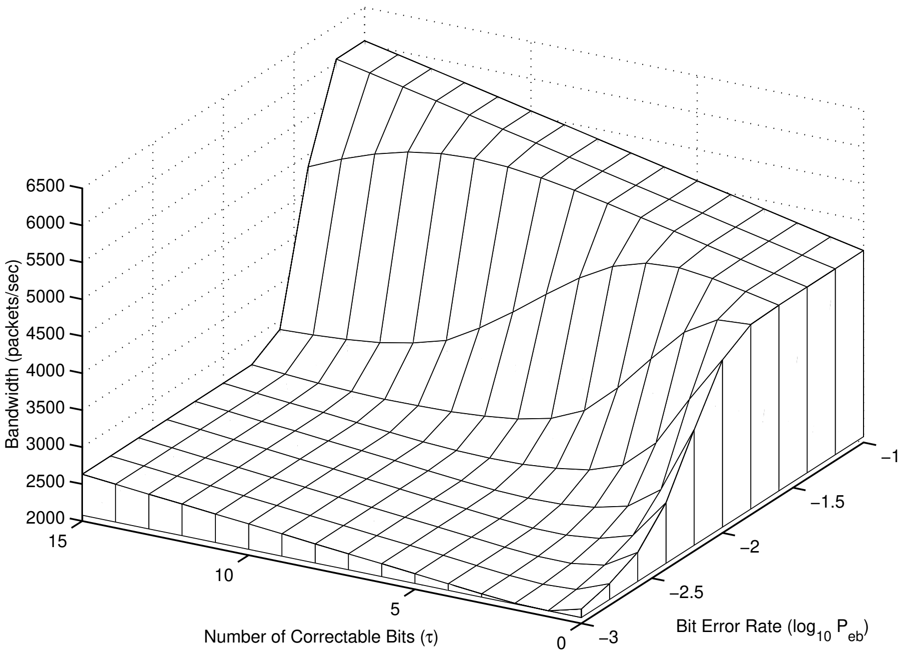


Figure 10: Effective bandwidth versus BER's ($P_{eb}$) and $\tau$'s.

of $t$ as well as the constraint Pr[delay $> t$].

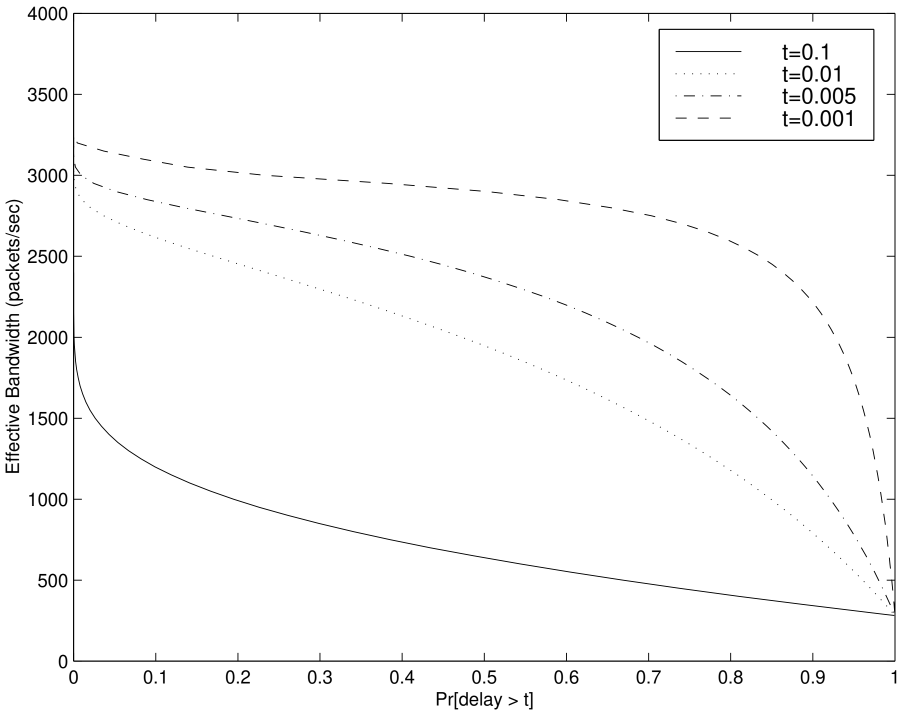


Figure 11: Effective bandwidth versus target delay constraints Pr[delay $> t$].

Figure 12 shows the packet discard rate as a function of the limit on the number of transmissions ($N_l$) for different $\gamma$'s. The value $1/\gamma$ is the mean period of Bad channel state. Thus, $\gamma = 30$ and $\gamma = 100$ corresponds to the period of 0.0333 and 0.01 (sec). Over the whole range of $N_l$, we observe the slight deviation between the analytical and simulation results. The main source of the deviation results from the estimation on $\hat{w}_g$ and $\hat{w}_b$. Our analytical estimation based on fluid analysis tends to overestimate slightly $\hat{w}_g$ when compared to $\hat{w}_g$ based on the packet-based simulation. However, it is quite acceptable in consideration of the requirement on accuracy of the packet loss rate.

# 6 Conclusions

In this paper we investigated the delay and packet discard performance of a wireless link. The traffic source was modeled by an on-off fluid process and the fluctuations of the wireless channel were appropriately captured using a fluid version of Gilbert-Elliot model. Moreover, the impacts of error control schemes (ARQ and/or FEC) were incorporated. We obtained the delay distribution using two alternative approaches: Laplace transform and uniformization. The analytical results were then used to obtain the wireless effective bandwidth which can be used as a valuable tool in resource allocation and admission control in wireless networks. The adequacy of our analytical

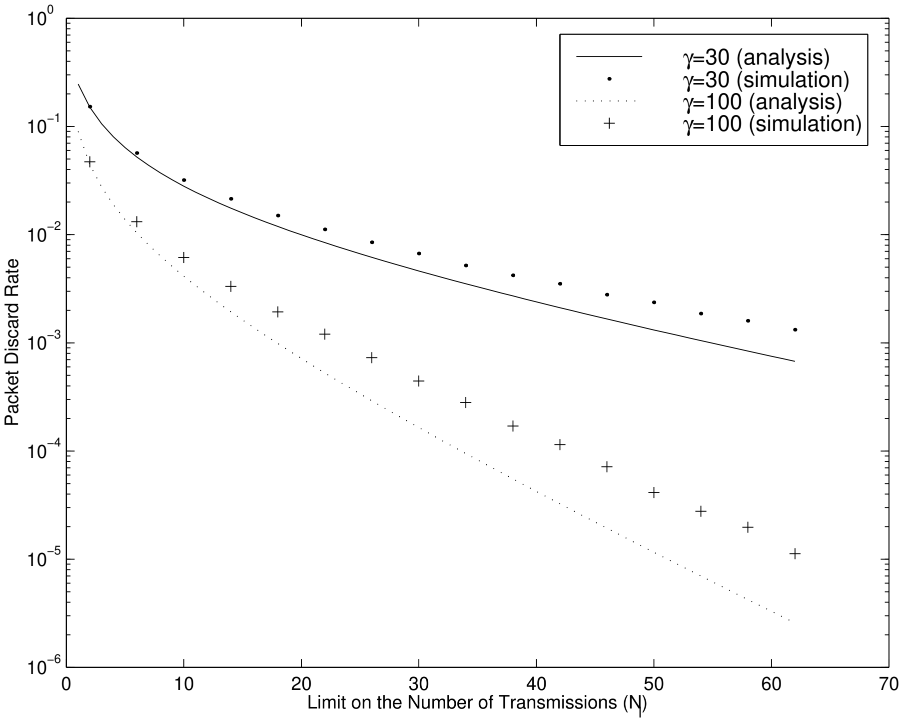


Figure 12: Packet discard rate versus the limit on the number of transmissions ($N_l$).

results were verified by contrasting them with simulations. It is observed that our analytical results consistently provide accurate estimations on delay distribution and packet discard rate. Our focus in this paper was on the packet delay performance as the primary measure of QoS. In the future, we plan to investigate the system capacity of the wireless link, which is constrained by diverse QoS parameters, e.g., delay, loss, and jitter. Other future research directions include the investigation of the impacts of interleaving on the packet-level performance of a wireless link.

# Appendix

## A Proof of Proposition 3.1

We define $P_i(t,x)$ as

$$P_i(t,x) \triangleq \Pr[\bar{C}_i(t) \leq x], \quad i \in \{g,b\}. \tag{33}$$

During a time period $\Delta t$, the system state changes from Good to Bad with probability $\delta\Delta t$. Similarly, it changes from Bad to Good with probability $\gamma\Delta t$. Now, we have

$$\begin{aligned} P_g(t+\Delta t,x) &= (1-\delta\Delta t)P_g(t,x-(c_g-c_b)\Delta t) + \delta\Delta t \cdot P_b(t,x-(c_g-c_b)\Delta t) \quad (34)\\ P_b(t+\Delta t,x) &= \gamma\Delta t \cdot P_g(t,x) + (1-\gamma\Delta t)P_b(t,x). \quad (35)\end{aligned}$$

Taking the limit as $\Delta t \to 0$, we have

$$\begin{aligned} \frac{\partial P_g(t,x)}{\partial t} &= -(c_g-c_b)\frac{\partial P_g(t,x)}{\partial x} - \delta P_g(t,x) + \delta P_b(t,x) \quad (36)\\ \frac{\partial P_b(t,x)}{\partial t} &= \gamma P_g(t,x) - \gamma P_b(t,x). \quad (37)\end{aligned}$$

Arranging the previous equations in the matrix form, we arrive at

$$\frac{\partial \boldsymbol{P}(t,x)}{\partial t} + \boldsymbol{D}\frac{\partial \boldsymbol{P}(t,x)}{\partial x} = \boldsymbol{M}\boldsymbol{P}(t,x) \tag{38}$$

where

$$\boldsymbol{P}(t,x) \triangleq \begin{bmatrix} P_g(t,x) \\ P_b(t,x) \end{bmatrix}, \boldsymbol{M} \triangleq \begin{bmatrix} -\delta & \delta \\ \gamma & -\gamma \end{bmatrix}, \text{ and } \boldsymbol{D} \triangleq \begin{bmatrix} c_g-c_b & 0 \\ 0 & 0 \end{bmatrix}.$$

Applying Laplace transform to (38) with respect to $t$ (notation $\star$), we have

$$s\boldsymbol{P}^\star(s,x) - \boldsymbol{P}(0,x) + \boldsymbol{D}\frac{\partial \boldsymbol{P}^\star(s,x)}{\partial x} = \boldsymbol{M}\boldsymbol{P}^\star(s,x). \tag{39}$$

Applying Laplace transform again with respect to $x$ (notation $\bullet$), we have

$$s\boldsymbol{P}^{\star\bullet}(s,u) - \boldsymbol{P}^\bullet(0,u) + \boldsymbol{D}(u\boldsymbol{P}^{\star\bullet}(s,u) - \boldsymbol{P}^\star(s,0)) = \boldsymbol{M}\boldsymbol{P}^{\star\bullet}(s,u). \tag{40}$$

Arranging the above equation, we have

$$\boldsymbol{P}^{\star\bullet}(s,u) = (u\boldsymbol{D} + s\boldsymbol{I} - \boldsymbol{M})^{-1}(\boldsymbol{P}^\bullet(0,u) + \boldsymbol{D}\boldsymbol{P}^\star(s,0)) \tag{41}$$

where

$$(u\boldsymbol{D} + s\boldsymbol{I} - \boldsymbol{M})^{-1} = \frac{1}{\Delta(s)} \begin{bmatrix} s+\gamma & \delta \\ \gamma & (c_g - c_b)u + s + \delta \end{bmatrix} \tag{42}$$

with $\Delta(s) \stackrel{\triangle}{=} (c_g - c_b)(s+\gamma)u + s(s+\delta+\gamma)$.

The system has the following boundary conditions

$$\boldsymbol{P}(t,0) = \begin{bmatrix} P_g(t,0) \\ P_b(t,0) \end{bmatrix} = \begin{bmatrix} 0 \\ e^{-\gamma t} \end{bmatrix} \text{ and } \boldsymbol{P}(0,x) = \begin{bmatrix} P_g(0,x) \\ P_b(0,x) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{43}$$

With the boundary conditions, we obtain

$$\boldsymbol{P}^{\star\bullet}(s,u) = \frac{1}{u\Delta(s)} \begin{bmatrix} s+\delta+\gamma \\ (c_g - c_b)u + s + \delta + \gamma \end{bmatrix}. \tag{44}$$

Taking inverse Laplace transform with respect to $u$ in (44), we have

$$\boldsymbol{P}^{\star}(s,x) = \begin{bmatrix} \frac{1}{s} - \frac{1}{s}\exp\left(-\frac{1}{c_g - c_b}\left(s+\delta-\frac{\delta\gamma}{s+\gamma}\right)x\right) \\ \frac{1}{s} - \frac{\gamma}{s(s+\gamma)}\exp\left(-\frac{1}{c_g - c_b}\left(s+\delta-\frac{\delta\gamma}{s+\gamma}\right)x\right) \end{bmatrix}. \tag{45}$$

Finally, in order to invert with respect to $t$, we use the convolution property and the following Laplace transform pair:

$$\mathcal{L}[J_0(2\sqrt{at})] = \frac{e^{-as}}{s} \tag{46}$$

where $J_0(z)$ is the Bessel function of order 0 given by

$$J_0(z) = \sum_{n=0}^{\infty} \frac{(-1)^n (z/2)^{2n}}{(n!)^2}.$$

Thus, we have

$$\boldsymbol{P}(t,x) = \begin{bmatrix} 1 - e^{-\delta\hat{x}}(e^{-\gamma(t-\hat{x})}J_0(2\sqrt{-\delta\gamma\hat{x}(t-\hat{x})}) + \sum_{n=0}^{\infty}\frac{\delta\hat{x}}{(n!)^2}\Gamma(n+1,\gamma(t-\hat{x}))) \\ 1 - e^{-\delta\hat{x}}\sum_{n=0}^{\infty}\frac{\delta\hat{x}}{(n!)^2}\Gamma(n+1,\gamma(t-\hat{x})) \end{bmatrix} \tag{47}$$

where $\hat{x} = x/(c_g - c_b)$ and $\Gamma(n,z)$ is the incomplete gamma function given by:

$$\Gamma(n,z) = \int_0^z e^{-x} x^{n-1} dx.$$

This completes the proof of Proposition (3.1).

## B Proof of Proposition 3.2

From (16), we have

$$\Pr[C(t) \geq x] = \begin{cases} \Pr[t_g \geq \frac{x - c_b t}{c_g - c_b}] = \Pr[t_b \leq \frac{c_g t - x}{c_g - c_b}], & \text{if } c_b t < x \leq c_g t \\ 0, & \text{if } x > c_g t \\ 1, & \text{if } x \leq c_b t. \end{cases} \tag{48}$$

To obtain the probabilities $\Pr[t_g \leq s]$ and $\Pr[t_b \leq s]$, we use the uniformization technique in which the transition rates between states are uniformized by introducing transitions from a state to itself (refer to pp. 284 in [19] for details). The probability $\Pr[t_g \leq s | h(0) = g]$, where $s < t$, is given by:

$$\begin{aligned} &\Pr[t_g \leq s | h(0) = g] \\ &= \sum_{n=1}^{\infty} e^{-(\delta+\gamma)t} \frac{((\delta+\gamma)t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta+\gamma}\right)^{k-1} \left(\frac{\delta}{\delta+\gamma}\right)^{n-k+1} \\ &\quad \cdot \sum_{i=k}^{n} \binom{n}{i} \left(\frac{s}{t}\right)^{i} \left(1 - \frac{s}{t}\right)^{n-i} \qquad (49) \\ &= e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \left(\frac{s}{t}\right)^{i} \left(1 - \frac{s}{t}\right)^{n-i}. \qquad (50) \end{aligned}$$

For $s = t$, we have

$$\Pr[t_g = t | h(0) = g] = e^{-\delta t}. \tag{51}$$

In a similar way, the probability $\Pr[t_b \leq s | h(0) = b]$, where $s < t$, is given by:

$$\begin{aligned} &\Pr[t_b \leq s | h(0) = b] \\ &= e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \left(\frac{s}{t}\right)^{i} \left(1 - \frac{s}{t}\right)^{n-i}. \qquad (52) \end{aligned}$$

For $s = t$, we have

$$\Pr[t_b = t | h(0) = b] = e^{-\gamma t}. \tag{53}$$

From (48), we have

$$\Pr[C_g(t) \geq x] = 1 - \Pr\left[t_g \leq \frac{x - c_b t}{c_g - c_b} | h(0) = g\right] \quad (54)$$

$$= 1 - e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \chi^i (1-\chi)^{n-i} \quad (55)$$

where $\chi = \frac{x - c_b t}{(c_g - c_b)t}$.

In a similar way, we obtain $\Pr[C_b(t) \geq x]$

$$\Pr[C_b(t) \geq x] = \Pr\left[t_b \leq \frac{c_g t - x}{c_g - c_b} | h(0) = b\right] \quad (56)$$

$$= e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \sum_{i=k}^{n} \binom{n}{i} \chi^{n-i} (1-\chi)^i. \quad (57)$$

This completes the proof of Proposition 3.2.

# C Proof of Proposition 3.3

From (12), we have

$$\begin{aligned}
&\Pr[D \leq t] \\
&= \frac{r}{T} \int_{0-}^{c_b t} \Pr[C_g(t) \geq x]\pi_{1,g}(x) + \Pr[C_b(t) \geq x]\pi_{1,b}(x)dx \\
&+ \frac{r}{T} \int_{c_b t}^{c_g t} \Pr[C_g(t) \geq x]\pi_{1,g}(x) + \Pr[C_b(t) \geq x]\pi_{1,b}(x)dx \quad (58) \\
&= \frac{r}{T}(\Pi_{1,g}(c_b t) + \Pi_{1,b}(c_b t)) \\
&+ \frac{r}{T} \int_{c_b t}^{c_g t} \Pr[C_g(t) \geq x]\pi_{1,g}(x) + \Pr[C_b(t) \geq x]\pi_{1,b}(x)dx. \quad (59)
\end{aligned}$$

In order to evaluate the integral in the above equation, we substitute (18) into (59):

$$\begin{aligned}
&\int_{c_b t}^{c_g t} \Pr[C_g(t) \geq x]\pi_{1,g}(x)dx \\
&= (\Pi_{1,g}(c_g t) - \Pi_{1,g}(c_b t)) - e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{\delta}{c_g - c_b}\right)^n \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \\
&\cdot \sum_{i=k}^{n} \binom{n}{i} \int_{c_b t}^{c_g t} (x - c_b t)^i (c_g t - x)^{n-i} \pi_{1,g}(x)dx. \quad (60)
\end{aligned}$$

Consider the integral in (60),

$$\int_{c_b t}^{c_g t} (x - c_b t)^i (c_g t - x)^{n-i} \pi_{1,g}(x) dx \tag{61}$$

$$= (c_g t - c_b t)^n \int_{c_b t}^{c_g t} \left( \frac{x - c_b t}{c_g t - c_b t} \right)^i \left( 1 - \frac{x - c_b t}{c_g t - c_b t} \right)^{n-i} \pi_{1,g}(x) dx \tag{62}$$

$$= (c_g t - c_b t)^{n+1} \int_0^1 \chi^i (1-\chi)^{n-i} \pi_{1,g}((c_g - c_b) t \chi + c_b t) d\chi \tag{63}$$

where

$$\chi \triangleq \frac{x - c_b t}{t(c_g - c_b)}. \tag{64}$$

Let $\pi_{1,g}(x) = \sum_l a_l e^{z_l x}$. Then we have

$$\int_0^1 \chi^i (1-\chi)^{n-i} \pi_{1,g}((c_g - c_b) t \chi + c_b t) d\chi = e^{z_l c_b t} \sum_l a_l \int_0^1 \chi^i (1-\chi)^{n-i} e^{z_l (c_g - c_b) t \chi} d\chi. \tag{65}$$

By the formula in [22],

$$\int_0^1 x^i (1-x)^{n-i} e^{\beta x} dx = B(n-i+1, i+1)\, \Phi(i+1; n+2; \beta) \tag{66}$$

where $B(x, y)$ is the beta function given by

$$B(x,y) \triangleq \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1} (1-t)^{y-1} dt \tag{67}$$

and $\Phi(x; y; z)$ is the confluent hypergeometric function given by

$$\Phi(x;y;z) \triangleq \sum_{k=0}^{\infty} \frac{(x)_k}{(y)_k} \frac{z^k}{k!} = \frac{1}{B(x, y-x)} z^{1-y} \int_0^z e^t t^{x-1} (z-t)^{y-x-1} dt \tag{68}$$

with $(a)_n \triangleq a(a+1)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)}$.

Thus, we have

$$\int_0^1 \chi^i (1-\chi)^{n-i} e^{z_l (c_g - c_b) t \chi} d\chi = \frac{(n-i)!i!}{(n+1)!} \Phi(i+1; n+2; z_l (c_g - c_b) t) \tag{69}$$

Now, Equation (60) becomes

$$
\begin{aligned}
&\int_{c_b t}^{c_g t} \Pr[C_g(t) \geq x]\pi_{1,g}(x)dx \\
&= (\Pi_{1,g}(c_g t) - \Pi_{1,g}(c_b t)) - e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \\
&\quad \cdot \sum_{i=k}^{n} \binom{n}{i} (c_g - c_b)t \sum_{l} a_l e^{z_l c_b t} \frac{(n-i)!i!}{(n+1)!} \Phi(i+1; n+2; z_l(c_g - c_b)t) \qquad (70) \\
&= (\Pi_{1,g}(c_g t) - \Pi_{1,g}(c_b t)) - e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \\
&\quad \cdot \sum_{i=k}^{n} (c_g - c_b)t \sum_{l} a_l e^{z_l c_b t} \Phi(i+1; n+2; z_l(c_g - c_b)t). \qquad (71)
\end{aligned}
$$

In a similar way, we obtain

$$
\begin{aligned}
&\int_{c_b t}^{c_g t} \Pr[C_b(t) \geq x]\pi_{1,b}(x)dx = e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \\
&\quad = \sum_{i=k}^{n} (c_g - c_b)t \sum_{m} a_m e^{z_m c_b t} \Phi(n-i+1; n+2; z_m(c_g - c_b)t) \qquad (72)
\end{aligned}
$$

where $\pi_{1,b}(x) = \sum_m a_m e^{z_m x}$.

Substituting (71) and (72) into (59), we have

$$
\begin{aligned}
\Pr[D \leq t] &= \frac{r}{T}(\Pi_{1,g}(c_g t) + \Pi_{1,b}(c_b t)) - \frac{r}{T} e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\delta t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\gamma}{\delta}\right)^{k-1} \\
&\quad \cdot \sum_{i=k}^{n} (c_g - c_b)t \sum_{l} a_l e^{z_l c_b t} \Phi(i+1; n+2; z_l(c_g - c_b)t) \\
&\quad + \frac{r}{T} e^{-(\delta+\gamma)t} \sum_{n=1}^{\infty} \frac{(\gamma t)^n}{(n+1)!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\delta}{\gamma}\right)^{k-1} \\
&\quad \cdot \sum_{i=k}^{n} (c_g - c_b)t \sum_{m} a_m e^{z_m c_b t} \Phi(n-i+1; n+2; z_m(c_g - c_b)t). \qquad (73)
\end{aligned}
$$

This completes the proof of Proposition 3.3.

## References


[1] A. S. Acampora and M. Naghshineh, "An architecture and methodology for mobile–executed handoff in cellular ATM networks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1365–1375, Oct. 1994.

[2] A. Acampora, "Wireless ATM: a perspective on issues and prospects," *IEEE Pers. Commun.*, vol. 3, no. 4, pp. 8–17, Aug. 1996.

[3] D. Raychaudhuri and N. D. Wilson, "ATM-based transport architecture for multiservices wireless personal communication networks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1401–1414, Oct. 1994.

[4] J. B. Cain and D. N. McGregor, "A recommended error control architecture for ATM networks with wireless links," *IEEE J. Select. Areas Commun.*, vol. 15, no. 1, pp. 16–28, Jan. 1997.

[5] M. Schwartz, "Network management and control issues in multimedia wireless networks," *IEEE Pers. Commun.*, vol. 2, no. 3, pp. 8–16, June 1995.

[6] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 756–769, Dec. 1997.

[7] H. Liu and M. El Zarki, "Performance of H.263 video transmission over wireless channels using hybrid ARQ," *IEEE J. Select. Areas Commun.*, vol. 15, no. 9, pp. 1775–1786, Dec. 1997.

[8] 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, Aug. 1996.

[9] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Trans. Networking*, vol. 5, no. 1, pp. 1–12, Feb. 1997.

[10] D. Reininger, R. Izmailov, B. Rajagopalan, M. Ott, and D. Raychaudhuri, "Soft QoS control in the WATMnet broadband wireless system," *IEEE Pers. Commun.*, vol. 6, no. 1, pp. 34–43, Feb. 1999.

[11] J. Capone and I. Stavrakakis, "Achievable QoS and scheduling policies in integrated services wireless networks," *Perform. Eval.*, vol. 27/28, no. 1, pp. 347–365, Oct. 1996.

[12] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," in *SIGCOMM '97*, Sept. 1997.

[13] J. G. Kim and M. Krunz, "Effective bandwidth in wireless ATM networks," in *MobiCom '98*, Oct. 1998, pp. 233–241.

[14] N. Guo and S. D. Morgera, "Frequency-hopped ARQ for wireless network data services," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1324–1336, Sept. 1994.

[15] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data–handling system with multiple sources," *Bell Syst. Tech. J.*, vol. 61, pp. 1871–1894, 1982.

[16] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer," *Adv. Appl. Prob.*, vol. 20, pp. 646–676, 1988.

[17] J. G. Kim and M. Krunz, "Bandwidth allocation in wireless ATM networks with guaranteed cell loss performance," Tech. Rep. CENG-TR-98-115, Department of ECE, University of Arizona, http://www.ece.arizona.edu/~bnlab, Sept. 1998.

[18] A. I. Elwalid and D. Mitra, "Statistical multiplexing with loss priorities in rate-based congestion control of high-speed networks," *IEEE Trans. Commun.*, vol. 42, no. 11, pp. 2989–3002, Nov. 1994.

[19] S. Ross, *Stochastic Processes*, John Wiley & Sons, second edition, 1996.

[20] A. I. Elwalid and D. Mitra, "Effective bandwidth of general markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 329–343, June 1993.

[21] S. Lin and Jr. D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1983.

[22] I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, NY, 1965.