Performance Analysis of Delay-Constrained Communications Over Slow Rayleigh Fading Channels

William Turin, *Senior Member, IEEE,* and Michele Zorzi, *Senior Member, IEEE*

***Abstract*—In a typical wireless system, communications between a transmitter–receiver pair is subject to impairments such as intracell and intercell interference, as well as multipath fading. For practical wireless data transmission systems, e.g., UMTS or GPRS, it has been demonstrated by simulation and supported by real measurements that these impairments can be adequately modeled by hidden Markov models (HMMs). It has also been demonstrated that various types of wireless data arrival process can be modeled by a batch Markov arrival process (BMAP). This paper presents analytical methods for evaluating the packet queue length and packet delay probability distributions assuming that packet arrivals are modeled by a BMAP and packets are transmitted over channels with bursts of errors which are modeled by HMMs. In contrast with simulations, the analytical approach allows a system designer to find and test proper diversity, source and channel coding schemes, and communication protocols more efficiently. Analytical and simulation results are compared to determine the accuracy of the presented methods.**

***Index Terms*—Batch Markov arrival process (BMAP), bursty errors, delay performance, fading, hidden Markov models, queueing analysis.**

I. INTRODUCTION

IN RECENT years, the emergence of data communications services in wireless systems has received a lot of attention from the research community. It is envisioned that, even though not a major source of revenue at present, in the near future data will represent most of the traffic carried by wireless systems, eventually surpassing voice. This transition will be fueled by the fact that popular wireline applications, such as web browsing, will eventually find their way into the wireless world. More specialized applications, such as teleconferencing, may also play some role.

In this context of data communications, classic performance metrics which have been traditionally applied to voice systems no longer apply and more elaborate ways to define quality-of-service (QoS) are being considered. Among the many parameters used for this purpose are throughput, packet loss, delay, and delay jitter.

Emerging applications which are envisioned in recent or soon-to-be-developed wireless systems (e.g., GPRS, UMTS, and cdma2000) include bursty data, file transfer, audio, and video streaming. For the success of these systems, it is extremely important to understand how these applications can be properly supported over the standard air interfaces and how they are affected by the encountered impairments. In a typical wireless system, communication between pairs of users is subject to a variety of impairments, the most important being intracell and intercell interference and propagation effects due to multipath. These impairments exhibit memory, in the sense that impairments experienced by consecutive blocks of data are correlated and so are the errors. It has been recently shown that being able to capture this feature is very important in order to accurately assess the performance of a wireless data communications systems [1]. It is especially important in compressed voice and image transmissions which are extremely sensitive to bursty errors [2].

It has been demonstrated theoretically and confirmed experimentally [3], [4] that hidden Markov models (HMMs) are capable to model accurately channels with memory, due to the generality of this class of models. There are several special cases of HMMs that are popular in applications: the Elliott–Gilbert model with a good state with small error probability and a bad state with larger error probability [5], [6]; the models in which only the transitions between good and bad states are allowed and errors do not occur in good states and occur with probability one in bad states [1], [7], [8]; models based on multiple Markov chains [8], [9]. The popularity of these models is primarily due to a relative simplicity of fitting these models to experimental data.

Other papers take an HMM for the error process and apply it to the study of the performance of some protocols [10], [11]. In the majority of papers and books it is assumed that the feedback channel is ideal; other papers consider a nonideal feedback channel, but they ignore the message arrival discipline [8] or consider models with two states only [10], [12].

In this paper, we extend the analysis of [10] to HMMs with any number of states. Our analysis provides a closed-form result for the probability that a packet's delay (made up of queueing delay and retransmission delay) exceeds a certain value, for a queueing system where the server is up or down according to an HMM.

We present results for the commonly considered Rayleigh fading channel. Parameters for the HMM are evaluated through one of the algorithms available (see, for example, [1, Ch. 3] and [8, Ch. 3]) and the resulting model is applied to our analysis to find the delay distribution for the retransmission protocol considered. Sensitivity of the model parameters and the model order

Manuscript received December 7, 1999; revised March 22, 2000, January 29, 2001 and June 28, 2001; accepted February 25, 2002. The editor coordinating the review of this paper and approving it for publication is V. C. M. Leung.

W. Turin is with AT&T Labs-Research, Shannon Laboratory, Florham Park, NJ 07932 USA (e-mail: wt@research.att.com).

M. Zorzi is with the Dipartimento di Ingegneria, Università di Ferrara, 44100 Ferrara, Italy (e-mail: zorzi@ing.unife.it).

Digital Object Identifier 10.1109/TWC.2002.804184

to the various propagation and physical layer parameters such as Doppler frequency and signal-to-noise ratio (SNR) is also discussed. The accuracy of the approach is assessed through direct simulation of the fading process and of the protocol operation.

II. Channel Model

In this section, we examine models for the wireless channels with and without the use of diversity. We focus on a single communication channel whose signals are affected by fading, interference and noise which leads to bursts of errors. The time axis is slotted and transmission occurs in blocks of N bits. In each slot, packets are generated at the transmitter according to some arrival process and queued in a buffer awaiting transmission. By properly selecting the parameters of the arrival process, a variety of data traffic classes can be considered, including web browsing, audio/video streaming, and ftp [16]. This corresponds to important real-world scenarios such as data transmission in narrowband systems (e.g., based on time-division multiple-access (TDMA) and/or frequency-division multiple-access (FDMA), as in GSM/GPRS), as well as in code-division multiple-access (CDMA)-based systems (e.g., UMTS and cdma2000). Specifically, in TDMA-based systems, the main impairment is due to random fading fluctuations, whose time evolution can be tracked via Markov models [13], [14]. On the other hand, in CDMA systems multiple access interference plays a key role and, also due to the presence of complex dynamic mechanisms such as power control, direct modeling of the channel behavior is not adequate. However, recent studies based on fitting simulation and experimental data [3], [4] have shown that HMMs successfully capture the essential behavior of the data block error process in these more complex scenarios as well.

Consider a transmission channel which can be described by the error process, i.e., a binary process which identifies the correct and erroneous transmissions. We assume that the error process can be modeled by a stationary HMM which is defined by two matrices $\boldsymbol{P}_e$, $e = 0, 1$ whose elements $p_{ij}(e) = \Pr(e, j|i)$ are the conditional probabilities of transferring from channel state i to channel state j and producing $e = 0$ for a correct reception and $e = 1$ for an error. We call the matrices $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ matrix probabilities of a correct reception and error, respectively. The sequence of states is a Markov chain with the state transition probability matrix $\boldsymbol{P}_0 + \boldsymbol{P}_1$. Since we assume that the process is stationary, the state initial probability vector can be obtained from the system

$$\boldsymbol{\pi}(\boldsymbol{P}_0 + \boldsymbol{P}_1) = \boldsymbol{\pi}, \quad \boldsymbol{\pi}\mathbf{1} = 1 \tag{1}$$

This model is general enough to describe various types of channel impairments in the form of fading and interference which causes bursty nature of errors. Common feature of these impairments is that they have finite memory which means that there is a finite number k such that the probability

$$\begin{aligned}\Pr(e_t|e_{t-1}, e_{t-2}, \ldots, e_{t-k}, e_{t-k-1}, \ldots)\\ = \Pr(e_t|e_{t-1}, e_{t-2}, \ldots, e_{t-k}).\end{aligned}$$

That is the probability of the error e_t depends only on a finite number of previous errors. This is a special case of an HMM whose states are defined by the sequences $i = (e_{t-1}, e_{t-2}, \ldots, e_{t-k})$. Thus, a sequence with finite memory can be described as an HMM (in this case, a Markov chain). However, since real wireless channel memory can be large, this description may require a huge state space. Some reduction of the state space can be achieved by considering variable length states [9]. An HMM, on the other hand, being a more general model allows us to reduce the state space even more. Note, that since encoders, modulators, scramblers, etc in addition to the physical layer have finite memory, the combined communication system can be described as an HMM (see [8, Ch. 4]). The model parameters can be estimated from experimental data [8, Ch. 3] or the results of simulation of a wireless channel including Rayleigh and multipath fading and various types of interference [9], [13].

The model just described can be used for error processes at various layers. For example, it can track bit-level errors, as in the original work by Gilbert [6], or it can track byte-level or packet-level errors as illustrated in [15]. In all cases, the HMM is fully characterized by the matrix probabilities of correct decoding, incorrect decoding, detected and undetected errors [8].

In the case of block- or packet-level errors, the matrix probabilities may be expressed in terms of the bit-level matrix probabilities. For example, in the absence of coding, the matrix probability of correct decoding of a block of length n has the form $\boldsymbol{P}(0) = \boldsymbol{P}_0^n$. The matrix probability of receiving a block with errors is $\boldsymbol{P}(1) = (\boldsymbol{P}_0 + \boldsymbol{P}_1)^n - \boldsymbol{P}_0^n$. In the presence of coding, these matrices have a more complex form [8]. In this paper, we will consider an HMM at the block (packet) level. We will derive it from the data obtained by simulating the error process arising in a Rayleigh fading channel [17]. It should be clear, however, that our analysis applies to all cases in which an HMM for the block error model is available.

This Markovian model is easily extended to account for diversity. In the presence of fading, diversity techniques improve performance by using two (or more) suitably spaced antennas. If the antennas receive independently attenuated replicas of the same signal, the probability of failure is reduced. If the subchannels are described by HMMs, the combined channel can be also described as an HMM. The HMM structure and parameters depend on the method of combining.

For example, if an *ideal switched diversity* is used, where the system is able to recognize and select (on a packet-by-packet basis) the antenna with the larger SNR, we can assume that a packet is received successfully in the combined channel if it is received over at least one of the subchannels. If M subchannels fade independently, the matrix probability of the transmission failure can be expressed as a Kronecker product of the corresponding matrix probabilities of the component subchannels [8]

$$\boldsymbol{P}(1) = \bigotimes_{k=1}^{M} \boldsymbol{P}_k(1) \tag{2}$$

and the matrix probability of successful transmission is

$$\boldsymbol{P}(0) = \bigotimes_{k=1}^{M} \boldsymbol{P}_k - \boldsymbol{P}(1). \tag{3}$$

Space–time coding may be used to achieve maximum diversity [18]. In this case, more complex techniques can be used for computing $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$ [8]. We assume in the sequel that

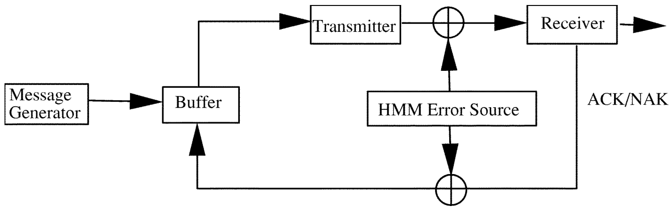

Fig. 1. Block diagram of the system.

the matrices $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$ are known (obtained analytically, through simulation, or estimated from experimental data).

A. Two-Way Channel Model

In two-way systems, information packets are sent over the direct channel and acknowledgments are sent over the feedback channel (see Fig. 1). There are many ways to encode the acknowledgments. They should be encoded in such a way that the probability of their mutual transformation is small. If the feedback channel is used solely for transmitting the acknowledgments and the minimum Hamming distance decoding is used, the positive acknowledgment (ACK) and negative acknowledgment (NAK) should be the bitwise complements of each other. For example, we can choose the all-zero sequence as the ACK and the all-one sequence as the NAK. The feedback channel decoder decodes an ACK, if the number of received ones is less than some number h, otherwise it decodes a NAK.

Usually, the feedback channel is used for transmitting some other information and the acknowledgments occupy a portion of the feedback packets. In this case, the previous scheme can be applied only to the portion occupied by the acknowledgments. Alternatively, an ACK is encoded by a code combination and a NAK is encoded by a noncode combination (for example, by inserting some errors in it). We assume that if the feedback message cannot be decoded, it is assumed to be a NAK. We assume also that the channels are perfectly synchronized. More information on encoding the acknowledgments and computing the corresponding matrix probabilities can be found in [8, Sec. 5.2].

To characterize these systems, we need to consider four different outcomes ℓ, as follows.

- $\ell = (0,0)$: transmission of a message without detected errors and decoding an ACK;
- $\ell = (0,1)$: transmission of a message without detected errors and decoding a NAK;
- $\ell = (1,0)$: detecting errors in a message and decoding an ACK;
- $\ell = (1,1)$: detecting errors in a message and decoding a NAK.

If both channels can be described by an HMM, we need to know four matrix probabilities $\boldsymbol{P}(\ell)$ to describe a two-way communication system. These probabilities can be obtained using the methods described in the previous section.

For example, if we assume that errors in the direct and feedback channels are independent, these matrix probabilities are equal to the Kronecker products of the corresponding matrix probabilities

$$\boldsymbol{P}(e_d, e_f) = \boldsymbol{P}_d(e_d) \otimes \boldsymbol{P}_f(e_f|e_d) \tag{4}$$

where $e_d = 0, 1$ and $e_f = 0, 1$ represent errors in the direct and feedback channels, respectively. Symbol $\boldsymbol{P}(e_f|e_d)$ denotes the conditional matrix probability of receiving a positive ACK or a NAK. For example, $\boldsymbol{P}(1|0)$ denotes the matrix probability of receiving an NAK when an ACK has been sent.

Errors in the feedback channel can cause a message loss (if errors in the direct channel have been detected, but the NAK in the return channel has been decoded as an ACK) or reception of the same message more than once (if $e_d = 0$, but $e_f = 1$).

For the channels described by HMMs, propagation delays can be easily incorporated into computing matrix probabilities. Indeed, let $\boldsymbol{P}$ be the state transition probability matrix of the underlying bit-level Markov chain. Then the matrix probability of e_d in the system with the propagation delay of Δ bits can be computed as $\boldsymbol{P}_d(e_d)\boldsymbol{P}^{\Delta}$ where $\boldsymbol{P}_d(e_d)$ is the corresponding probability for the system without the delay [8]. Notice that in this case more details of the **<Author: Please define "ARQ">** ARQ scheme must be provided.

In two-way wireless communications, the probability $p_d(D)$ that a packet is not transmitted within D slots of its arrival is an important characteristic of the QoS. To compute this probability, we need to know the matrix probabilities $\boldsymbol{P}(A)$ and $\boldsymbol{P}(N)$ of decoding an ACK and NAK, respectively. These matrix probabilities can be expressed as

$$\boldsymbol{P}(A) = \boldsymbol{P}(0,0) + \boldsymbol{P}(1,1) \tag{5}$$

$$\boldsymbol{P}(N) = \boldsymbol{P}(0,1) + \boldsymbol{P}(1,0). \tag{6}$$

We assume in the sequel that the matrices $\boldsymbol{P}(A)$ and $\boldsymbol{P}(N)$ are known (obtained analytically, through simulation, or estimated from experimental data).

III. Queue Length Distribution

Packets arriving at the transmitter enter the transmit buffer and are transmitted in the first-in–first-out (FIFO) fashion. Messages are removed from the buffer after receiving an ACK. A time interval between reception of two consecutive acknowledgments (ACKs or NAKs) we call a slot. Thus, the slot size depends on propagation delays and transmission times in the direct and feedback channels.

We assume that packet arrival is modeled by an HMM whose states may depend on the channel states. This model is general enough to describe large varieties of multimedia traffic: some states may correspond to voice some other states may correspond to data and so on.

Since we assume that the channel is also modeled by an HMM, the system can be described as an input–output HMM (IOHMM) [8] with matrix probabilities $\boldsymbol{P}(X, Y)$ where X is the number of arrivals and $Y \in (A, N)$ is the acknowledgment for the first packet in the queue. We assume that the propagation delays are included into these probabilities. (Note that the case in which the arrival and channel HMMs are independent may be considered as a special case of dependent HMMs with $\boldsymbol{P}(X, Y) = \boldsymbol{P}(X) \otimes \boldsymbol{P}(Y)$.)

If we assume that arrivals depend on the channel state and $a_{\ell X}$ is the probability that X arrivals occur in a slot given that the channel is in state ℓ, then

$$\boldsymbol{P}(X, Y) = \boldsymbol{A}_X \boldsymbol{P}(Y) \tag{7}$$

where $\boldsymbol{A}_X = diag\{a_{\ell X}\}$ is the diagonal matrix of $a_{\ell X}$.

Let us calculate the probability distribution of the number m of the packets in the queue waiting for transmission. According to our model, the number of packets in the queue can be described by a Markov chain whose state transition probability matrix has a special form called the structured stochastic matrix of the M/G/1 type [19]. The blocks of this matrix are defined by the transition matrix probabilities which are given by

$$\boldsymbol{P}(0|0) = \boldsymbol{P}(1, A) + \boldsymbol{P}(0, \Omega) \tag{8}$$

$$\boldsymbol{P}(0, \Omega) = \boldsymbol{P}(0, A) + \boldsymbol{P}(0, N)$$

$$\Omega = A \cup N \tag{9}$$

$$\boldsymbol{P}(m-1|m) = \boldsymbol{P}(0, A) \tag{10}$$

$$\boldsymbol{P}(m+i|m) = \boldsymbol{P}(i, N) + \boldsymbol{P}(i+1, A)$$

$$i = 0, 1, \ldots. \tag{11}$$

The stationary distribution of this Markov chain can be obtained by solving the following system of equations:

$$\boldsymbol{x}_m = \sum_{i=0}^{m+1} \boldsymbol{x}_i \boldsymbol{P}(m|i) \tag{12}$$

where in each vector $\boldsymbol{x}_i$ the subscript refers to the number of packets in the queue, whereas the various entries correspond to different states of the HMM. To solve this system, we denote

$$\boldsymbol{\Phi}(z) = \sum_{i=0}^{\infty} \boldsymbol{x}_i z^i \tag{13}$$

the solution matrix generating function. Using previous equations, we can express this function in the form

$$\boldsymbol{\Phi}(z) = \boldsymbol{x}_0 \boldsymbol{P}(0, A)(z-1)\left[\boldsymbol{I}z - \boldsymbol{\psi}_A(z) - z\boldsymbol{\psi}_N(z)\right]^{-1} \tag{14}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{\psi}_Y(z) = \sum_{X=0}^{\infty} \boldsymbol{P}(X, Y) z^X. \tag{15}$$

The unknown $\boldsymbol{x}_0$ can be obtained from the normalization condition

$$\lim_{z \to 1} \boldsymbol{\Phi}(z) = \boldsymbol{\pi} \tag{16}$$

where $\boldsymbol{\pi}$ is the steady-state probability distribution of the underlying HMM, which can be found from the following system of equations:

$$\boldsymbol{\pi P} = \boldsymbol{\pi}, \quad \boldsymbol{\pi 1} = 1, \quad \boldsymbol{P} = \boldsymbol{\psi}_A(1) + \boldsymbol{\psi}_N(1). \tag{17}$$

Alternatively, the solution can be found using the following equations [20, p. 142]

$$\boldsymbol{x}_0 \boldsymbol{K} = \boldsymbol{x}_0 \tag{18}$$

$$\boldsymbol{K} = \boldsymbol{P}(0|0) + \sum_{i=1}^{\infty} \boldsymbol{P}(0|i)\boldsymbol{G} \tag{19}$$

$$\boldsymbol{G} = \sum_{i=0}^{\infty} \boldsymbol{P}(1+i|1)\boldsymbol{G}^i \tag{20}$$

$$\boldsymbol{x}_i = \left[\boldsymbol{x}_0 \boldsymbol{B}_i + \sum_{j=0}^{i-1} \boldsymbol{x}_j \boldsymbol{C}_{i+1-j}\right](\boldsymbol{I} - \boldsymbol{C}_1)^{-1} \tag{21}$$

$$\boldsymbol{B}_i = \sum_{j=i}^{\infty} \boldsymbol{P}(i|0)\boldsymbol{G}^{j-i}$$

$$\boldsymbol{C}_i = \sum_{j=i}^{\infty} \boldsymbol{P}(1+i|1)\boldsymbol{G}^{j-i}. \tag{22}$$

If the number of arriving packets is bounded ($\boldsymbol{A}_k = 0$ for $k > M$), then $\boldsymbol{\Phi}(z)$ is a rational function and its expansion into a power series can be obtained using its decomposition into partial fractions. In a system in which only one packet can arrive in a slot ($\boldsymbol{A}_k = 0$ for $k > 1$), it is possible to obtain a matrix-geometric solution [19]

$$\boldsymbol{x}_i = \boldsymbol{x}_0 \boldsymbol{R}^i \quad i = 1, 2, \ldots \tag{23}$$

where $\boldsymbol{x}_0$ can be found from the system

$$\boldsymbol{x}_0 = \boldsymbol{x}_0[\boldsymbol{P}(0|0) + \boldsymbol{RP}(0|1)], \quad \boldsymbol{x}_0(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{1} = 1 \tag{24}$$

and $\boldsymbol{R}$ satisfies the following matrix quadratic equation:

$$\boldsymbol{R} = \boldsymbol{P}(1|0) + \boldsymbol{RP}(1|1) + \boldsymbol{R}^2\boldsymbol{P}(1|2). \tag{25}$$

It is not difficult to show that $\boldsymbol{x}_0 = \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{R})$.

As an example, consider a two-state system [10] with the matrix probabilities given by (7) where

$$\boldsymbol{P}(A) = \begin{pmatrix} p_{00} & p_{01} \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{P}(N) = \begin{pmatrix} 0 & 0 \\ p_{10} & p_{11} \end{pmatrix}. \tag{26}$$

In this case

$$\boldsymbol{\psi}_Y(z) = \boldsymbol{A}(z)\boldsymbol{P}(Y), \qquad \boldsymbol{A}(z) = \begin{pmatrix} A_0(z) & 0 \\ 0 & A_1(z) \end{pmatrix} \tag{27}$$

and (14) takes the form

$$\boldsymbol{\Phi}(z) = \frac{(z-1)a_{00}x_{00}}{\Delta(z)}\left[p_{00} + (p_{01} - p_{11})\, A_0(z), p_{01}\right] \tag{28}$$

where

$$\Delta(z) = (z - p_{00}A_0(z))\,(1 - p_{11}A_1(z)) - p_{10}p_{01}A_0(z)A_1(z). \tag{29}$$

By applying the L'Hospital's rule, we obtain from (16)

$$\boldsymbol{x}_0 = \frac{\pi_0\,(1 - A_0'(1)) - \pi_1 A_1'(1)}{a_{00}\,[p_{00} + (p_{01} - p_{11})\, a_{10}]}\,(p_{00} + (p_{01} - p_{11})\, a_{10}, p_{01}) \tag{30}$$

where

$$\boldsymbol{\pi} = (\pi_0, \pi_1) = \left(\frac{p_{10}}{p_{01} + p_{10}}, \frac{p_{01}}{p_{01} + p_{10}}\right) \tag{31}$$

is obtained from (17). This coincides with the result of [10].

If packets arrive according to an independent Bernoulli arrival model with probability λ, then

$$\boldsymbol{A}(z) = (1 - \lambda + \lambda z)\boldsymbol{I} \tag{32}$$

and $\boldsymbol{\Phi}(z)$ is a rational function. Its denominator has two roots $z_1 = 1$ and $z_2 = 1/\beta$ where

$$\beta = \frac{(1-\lambda)p_{11} + \lambda p_{01}}{(1-\lambda)p_{10} + \lambda p_{00}} \cdot \delta, \quad \delta = \frac{\lambda}{1-\lambda} \tag{33}$$

so that, we can write

$$\mathbf{\Phi}(z) = \boldsymbol{x}_0 + (\boldsymbol{\pi} - \boldsymbol{x}_0)\,\frac{(1-\beta)z}{(1-\beta z)} \tag{34}$$

where, according to (30)

$$\boldsymbol{x}_0 = \left[\pi_0 - \frac{\pi_1\lambda}{(1-\lambda)}, \pi_1(1-\beta)\right]. \tag{35}$$

Therefore, the queue length is geometrically distributed

$$\boldsymbol{x}_k = (\boldsymbol{\pi} - \boldsymbol{x}_0)\,(1-\beta)\beta^{k-1}, \quad k = 1, 2, \ldots. \tag{36}$$

This agrees with the results of [10]. The stationary distribution exists if $\beta < 1$.

The same result can also be obtained using the matrix-geometric approach

$$\boldsymbol{x}_i = \boldsymbol{x}_0\boldsymbol{R}^i, \quad \boldsymbol{R} = \begin{pmatrix} 0 & 0 \\ \delta & \beta \end{pmatrix}. \tag{37}$$

It is easy to see that $\boldsymbol{x}_0 = \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{R})$ and $(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{R}(1-\beta) = \boldsymbol{R}$.

IV. Computation of the Lateness Probability

Let us consider now the probability that a packet is not transmitted within D slots of its arrival. In a delay constrained system, this will be the probability that a packet exceeds the maximum tolerable delay.

Let $p_d(D)$ be the probability that a packet is not successfully delivered within D slots from its arrival. This probability can be evaluated as

$$p_d(D) = \sum_{i=0}^{\infty} \boldsymbol{x}_i \boldsymbol{P}_D(g \le i)\mathbf{1} \tag{38}$$

where

$$\boldsymbol{P}_D(g \le i) = \begin{cases} \sum_{g=0}^{i} \boldsymbol{P}_D(g), & \text{for } i \le D \\ [\boldsymbol{P}(N) + \boldsymbol{P}(A)]^D, & \text{otherwise} \end{cases} \tag{39}$$

is the matrix probability that there are no more than i successful transmissions in D slots. The matrix probability $\boldsymbol{P}_D(g)$ of exactly g successes can be obtained using the following recursive equations:

$$\boldsymbol{P}_1(0) = \boldsymbol{P}(N), \quad \boldsymbol{P}_1(1) = \boldsymbol{P}(A) \tag{40}$$

$$\boldsymbol{P}_D(g) = \boldsymbol{P}_{D-1}(g)\boldsymbol{P}(N) + \boldsymbol{P}_{D-1}(g-1)\boldsymbol{P}(A) \tag{41}$$

where $\boldsymbol{P}(A) = \boldsymbol{\psi}_A(1)$ and $\boldsymbol{P}(N) = \boldsymbol{\psi}_N(1)$. These probabilities have the binomial matrix generating function [8]

$$\sum_{g=0}^{D} \boldsymbol{P}_D(g) z^g = [\boldsymbol{P}(N) + \boldsymbol{P}(A)z]^D. \tag{42}$$

By the summation theorem, $\boldsymbol{P}_D(g \le i)$ has the following generating function:

$$\boldsymbol{\phi}_g(z) = \sum_{i=0}^{D} \boldsymbol{P}_D(g \le i) z^i = (1-z)^{-1}[\boldsymbol{P}(N) + \boldsymbol{P}(A)z]^D. \tag{43}$$

Therefore, $p_d(D)$ can be obtained using the well-known identity [21, Sec. 2.3.10]

$$p_d(D) = \frac{1}{2\pi j}\oint_C \mathbf{\Phi}\left(\frac{1}{z}\right)\left(z - z^2\right)^{-1}[\boldsymbol{P}(N) + \boldsymbol{P}(A)z]^D dz\mathbf{1}. \tag{44}$$

The probability $p_d(D)$ can be obtained by the residue theorem.

Alternatively, for the binary arrival process, we can use the matrix-geometric solution which allows us to write

$$p_d(D) = \boldsymbol{x}_0 \sum_{i=0}^{\infty} \boldsymbol{R}^i \boldsymbol{P}_D(g \le i)\mathbf{1}. \tag{45}$$

The sum in this equation can be evaluated using the matrix $\boldsymbol{R}^m$ spectral decomposition [8, Appendix 6]

$$\boldsymbol{R}^m = \sum_{j=1}^{r}\sum_{i=1}^{m_j} \boldsymbol{B}_{ij}\begin{pmatrix} m \\ i-1 \end{pmatrix}\lambda_j^{m-i+1} \tag{46}$$

where λ_j, $j = 1, \ldots, r$, are the matrix $\boldsymbol{R}$ eigenvalues and the m_js are the corresponding multiplicities. Using this representation in (45) and recalling (43), we can write

$$p_d(D) = \boldsymbol{x}_0 \sum_{j=1}^{r}\sum_{i=1}^{m_j} \frac{\boldsymbol{B}_{ij}\boldsymbol{\phi}_g^{(i-1)}(\lambda_j)\mathbf{1}}{(i-1)!} \tag{47}$$

where $\phi_g^{(i)}(z) = \partial^i \phi_g(z)/\partial z^i$. In particular, if the matrix $\boldsymbol{R}$ has a simple structure ($m_j = 1$, $j = 1, 2, \ldots, r$), this formula becomes

$$p_d(D) = \boldsymbol{x}_0 \sum_{j=1}^{r} \boldsymbol{B}_{1j}\boldsymbol{\phi}_g(\lambda_j)\mathbf{1} \tag{48}$$

which after substituting $\boldsymbol{x}_0 = \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{R})$ and using (43) and (46) takes the form

$$p_d(D) = \boldsymbol{\pi}\sum_{j=1}^{r} \boldsymbol{B}_{1j}\left[\boldsymbol{P}(N) + \boldsymbol{P}(A)\lambda_j\right]^D\mathbf{1}. \tag{49}$$

Let us illustrate the calculations for the Gilbert's model and the Bernoulli arrival model. According to (37), matrix $\boldsymbol{R}$ has different eigenvalues: zero and β and, therefore, has a simple structure. In this case, we have the following spectral decomposition:

$$\boldsymbol{R}^m = \boldsymbol{B}\beta^m, \quad \boldsymbol{B} = \beta^{-1}\boldsymbol{R} = \begin{pmatrix} 0 & 0 \\ \frac{\delta}{\beta} & 1 \end{pmatrix}. \tag{50}$$

Thus

$$p_d(D) = \boldsymbol{\pi}\boldsymbol{B}[\boldsymbol{P}(N) + \boldsymbol{P}(A)\beta]^D\mathbf{1} = \pi_1\left[1 + \left(\frac{\delta}{\beta}\right)\right]\rho^D \tag{51}$$

where

$$\rho = p_{11} + \delta p_{01} \tag{52}$$

is the eigenvalue of $\boldsymbol{P}(N) + \boldsymbol{P}(A)\beta$ corresponding to the eigenvector $\boldsymbol{\pi}\boldsymbol{B}$. This agrees with the results of [10].

V. Numerical Example

In this section, we present some numerical results obtained on the basis of a three-state HMM of the packet error process induced by the Rayleigh fading[1]. The packet successes and failures are determined based on a threshold model for the instantaneous fading envelope. More specifically, let F be the fading margin of the channel, i.e., the maximum tolerable attenuation which does not cause a packet to be in error. Packets for which the fading attenuation exceeds F, are declared in error, whereas all others are assumed to be correctly received. The HMM is obtained by numerical fit of the burst and gap length distributions. We remark that this is just an example to illustrate the application of our analysis. Discussion about appropriate ways to model fading channels is not the focus here. We refer to the vast literature on the topic for more details [1]–[14].

Consider the case $F = 20$ dB and let $f_D T = 0.01$ be the value of the Doppler frequency normalized to the packet transmission rate, which determines the process memory span. The fading process is approximated by the HMM with the following matrix probabilities:

$$\boldsymbol{P}(A) = \begin{pmatrix} 0.997\,079 & 0 & 0 \\ 0.353\,477 & 0 & 0 \\ 0.121\,662 & 0 & 0 \end{pmatrix} \tag{53}$$

$$\boldsymbol{P}(N) = \begin{pmatrix} 0 & 0.002\,708 & 0.000\,213 \\ 0 & 0.646\,523 & 0 \\ 0 & 0 & 0.878\,338 \end{pmatrix}. \tag{54}$$

Using (17), we obtained

$$\boldsymbol{\pi} = (\,0.990\,678 \quad 0.007\,590 \quad 0.001\,732\,). \tag{55}$$

For Bernoulli arrivals with probability $\lambda = 0.5$, the solution of (25) is

$$\boldsymbol{R} = \begin{pmatrix} 0.002\,195 & 0.002\,005 & 0.000\,190 \\ 0.478\,724 & 0.478\,633 & 0.000\,091 \\ 0.784\,788 & 0.001\,570 & 0.783\,218 \end{pmatrix}. \tag{56}$$

Equation (46) takes the form

$$\boldsymbol{R}^m = \boldsymbol{B}_{11}\lambda_1^m + \boldsymbol{B}_{12}\lambda_2^m + \boldsymbol{B}_{13}\lambda_3^m \tag{57}$$

where $\lambda_1 = 0$, $\lambda_2 = 0.480\,635$, $\lambda_3 = 0.783\,411$

$$\boldsymbol{B}_{11} = \begin{pmatrix} 0.995\,589 & -0.004\,170 & -0.000\,241 \\ -0.995\,589 & 0.004\,170 & 0.000\,241 \\ -0.995\,589 & 0.004\,170 & 0.000\,241 \end{pmatrix}$$

$$\boldsymbol{B}_{12} = \begin{pmatrix} 0.004\,163 & 0.004\,167 & -0.000\,004 \\ 0.994\,898 & 0.995\,822 & -0.000\,924 \\ -0.015\,961 & -0.015\,975 & 0.000\,015 \end{pmatrix}$$

$$\boldsymbol{B}_{13} = \begin{pmatrix} 0.000\,248 & 0.000\,003 & 0.000\,245 \\ 0.000\,691 & 0.000\,008 & 0.000\,683 \\ 1.011\,550 & 0.011\,805 & 0.999\,744 \end{pmatrix}.$$

The results obtained using (23), (24), and (49) are compared with the protocol simulation in Figs. 2 and 3, where very good agreement can be observed. Fig. 2 shows the results for the queue size distribution, whereas Fig. 3 shows the results for the lateness probability. Note that the same results could be obtained by using a recursive approach similar to the one in [10]. However, that approach is inherently limited by numerical errors as probabilities become small, whereas on the other hand our approach here does not suffer from this limitation.

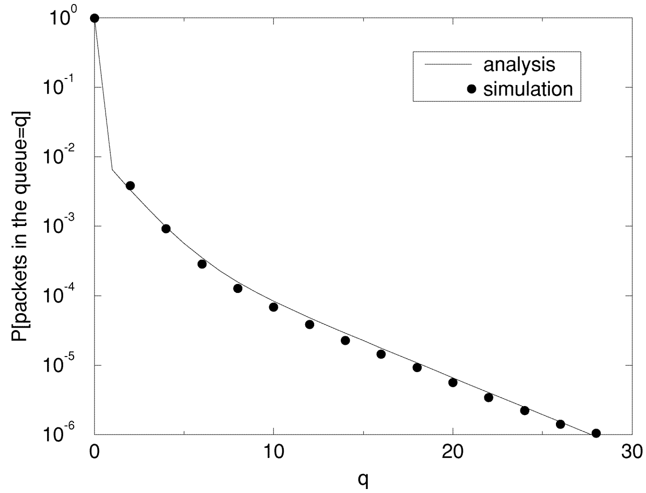

Fig. 2. Probability distribution of the queue size. Fading margin $F = 20$ dB, normalized Doppler frequency $f_D T = 0.01$, Bernoulli arrivals with rate $\lambda = 0.5$. Analytical results are for three-state HMM.

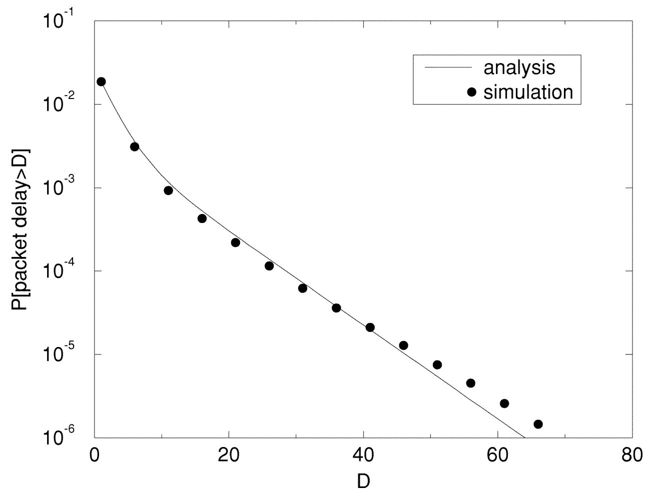

Fig. 3. Complementary distribution of the packet delay. Fading margin $F = 20$ dB, normalized Doppler frequency $f_D T = 0.01$, Bernoulli arrivals with rate $\lambda = 0.5$. Analytical results are for three-state HMM.

VI. Conclusion

In this paper, we have presented an analytical technique for the computation of the queue length distribution and for the probability that a packet suffers a delay larger than a given number of slots. Arrivals are modeled as a general batch Markov arrival process and a packet can be successfully transmitted in a slot according to a probabilistic function of a Markov chain (HMM). Explicit matrix expressions are found, which allow computation of small probabilities, for which other proposed methods may fail due to numerical problems.

The presented analysis can be applied to studying the performance of applications such as audio/video streaming, ftp or web browsing in any environment where communications impairments can be modeled via HMMs, including real-world scenarios such as GPRS, UMTS, and cdma2000.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

[1] M. Zorzi and R. R. Rao, “On channel modeling for delay analysis of packet communications over wireless links,” in *Proc. 36th Annual Allerton Conf.*, **Location**, Sept. 1998.
[2] J. Lu, K. B. Letaief, and M. l. Liou, “Robust video transmission over correlated mobile fading channels,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 737–751, Aug. 1999.

[3] W. Turin, R. Jana, C. Martin, and J. Winters, "Modeling wireless channel fading," in *Proc. Vechicular Technology Conf. VTC'2001*, vol. 3, Oct. 2001, pp. 1740–1744.

[4] M. Zorzi, G. Mazzini, V. Tralli, and A. Giovanardi, "Some results on the error statistics over a WCDMA air interface," presented at the Proc. MMT2000, Florida, Dec. 2000.

[5] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sept. 1963.

[6] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.

[7] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 221–227, Apr. 1967.

[8] W. Turin, *Digital Transmission Systems: Performance Analysis and Modeling*. New York: McGraw-Hill, 1998.

[9] F. Babich, O. E. Kelly, and G. Lombardi, "A variable-order discrete model for the mobile propagation channel," in *Broadband Wireless Channels*. London, U.K.: Springer, 1998, pp. 259–266.

[10] M. Zorzi and R. R. Rao, "Lateness probability of a retransmission scheme for error control on a two-state Markov channel," *IEEE Trans. Commun.*, vol. 47, pp. 1537–1548, Oct. 1999.

[11] D. Towsley, "The analysis of a statistical multiplexer with nonindependent arrivals and errors," *IEEE Trans. Commun.*, vol. COMM-28, pp. 65–72, Jan. 1980.

[12] ——, "A statistical analysis of ARQ protocols operating in a nonindependent error environment," *IEEE Trans. Commun.*, vol. COMM-29, pp. 971–981, July 1981.

[13] W. Turin and R. van Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1809–1817, Dec. 1998.

[14] M. Zorzi, R. R. Rao, and L. B. Milstein, "Error statistics in data transmission over fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1468–1477, Nov. 1998.

[15] J. R. Yee and E. J. Weldon, "Evaluation of the performance of error-correcting codes on a Gilbert channel," *IEEE Trans. Commun.*, vol. 43, pp. 2316–2323, Aug. 1995.

[16] J.-F. Frigon, V. C. M. Leung, and H. C. B. Chan, "Dynamic reservation TDMA protocol for wireless ATM networks," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 370–383, Feb. 2001.

[17] W. C. Jakes Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.

[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance analysis and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[19] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. Baltimore, MD: Johns Hopkins, 1981.

[20] ——, *Structured Stochastic Matrices M/G/1 Type and Their Applications*. New York: Marcel Dekker, 1989.

[21] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

William Turin (SM'83) received the M.S. degree in mathematics from Odessa State University, Odessa, U.S.S.R., in 1958 and the Ph.D. degree in mathematics from the Institute for Problems of Mechanics of the Academy of Sciences, U.S.S.R., Moscow, in 1966.

From 1960 to 1979, he was with the Moscow Institute of Electrical Engineering and Telecommunications, first as Assistant Professor and later as an Associate Professor. From 1980 to 1981, he was a Senior Research Scientist in the Department of Psychology, New York University, NY. Since 1981, he has been a Member of Technical Staff at AT&T Bell Laboratories, Holmdel and Murray Hill, NJ. Currently, he is a Technology Consultant at AT&T Labs-Research, Shannon Laboratory, Florham Park, NJ. His research interests include modeling channels with memory, digital signal processing, handwriting and speech recognition and compression, computer simulation, error correcting codes, microwave radio, and Markov processes. He is the author of five books and numerous papers.

Michele Zorzi (S'89–M'95–SM'98) was born in Venice, Italy, in 1966. He received the Laurea Degree and the Ph.D. in electrical engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992–1993, he was on leave from the University of California, San Diego (UCSD), attending graduate courses and doing research on multiple access in mobile radio networks.

In 1993, he joined the Faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy. After spending three years with the Center for Wireless Communications at UCSD, he joined the School of Engineering of the Università di Ferrara, Italy, in 1998, where he is currently a Professor. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, and energy constrained communications protocols. He serves on the Editorial Boards of the *ACM/URSI/Baltzer Journal of Wireless Networks* and *WILEY Journal of Wireless Communications and Mobile Computing*.

Dr. Zorzi currently serves on the Editorial Boards of the *IEEE Wireless Communications Magazine*, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the *IEEE Transactions on Mobile Computing*. He is also Guest Editor for Special Issues in the *IEEE Personal Communications Magazine* (energy management in personal communications systems) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (multimedia network radios).