SR ARQ Delay Statistics on N-State Markov Channels with Non-Instantaneous Feedback

Michele Rossi, Leonardo Badia, and Michele Zorzi

***Abstract*— In this paper the packet delay statistics of a fully reliable Selective Repeat ARQ (SR–ARQ) scheme is investigated. An N-State Discrete Time Markov Channel model is used to describe the packet error process and the channel round trip delay is considered to be non zero, i.e., ACK/NACK messages are received at the transmitter m channel slots after the packet transmission started. The ARQ packet delay statistics is evaluated by means of an exact analysis by jointly tracking packet errors and channel state evolution. Furthermore, procedures to derive a Markov Channel description of a Rayleigh fading process are discussed and the delay statistics obtained from the Markov analysis is compared with that estimated by simulation of the SR–ARQ protocol over the actual fading process. The accuracy of the delay statistics obtained from the Markov Channel representation of the actual fading process is investigated by explicitly addressing the effect of the number of states considered in the Markov channel model and the impact of the Doppler frequency. Finally, besides giving a new analysis to obtain link layer statistics over N-State Markov channels, the paper provides important considerations on the adequacy of the widely used Markov modeling approach for the characterization of higher layer performance.**

***Index Terms*— Automatic repeat request, data communication, Markov processes, error analysis, delay estimation, modeling.**

I. INTRODUCTION AND MOTIVATION

MULTIMEDIA applications in modern communication systems are highly sensitive to channel impairments and require effective error control techniques. Such techniques often rely on Forward Error Correction (FEC), Automatic Retransmission reQuest (ARQ), or a combination of the two. In such scenarios, a trade-off exists between data reliability, latency, and efficient bandwidth usage, and a good understanding of the impact of error control strategies becomes pivotal to provide adequate application level performance depending on the underlying channel behavior. In the past, great effort has been devoted to the characterization and the understanding of these error recovery systems. This paper is a natural follow-up of this research activity with the aim of understanding the impact of the channel error statistics on the performance of ARQ error control schemes.

In this work, we focus on a fully reliable Selective Repeat (SR) ARQ [1]; our aim is to derive delay statistics so as to understand how the retransmission process introduced by the ARQ error recovery algorithm affects the packet delay experienced at higher layers. In SR–ARQ, the transmitter sends packets (PDUs) in order of their arrival time at the link layer buffer, while the receiver replies to each received PDU with ACK/NACK messages by sending them over a feedback channel. The sender retransmits only negatively acknowledged (NACK) packets and then resumes the transmission process from the last packet sent so far. In this paper, we consider the statistics of the *delivery delay*, defined in the literature [1] as the sum of *transmission delay* and *re-sequencing delay*. These quantities are the delay between the first transmission and the correct reception of the PDU and the time spent in the receiver re-sequencing buffer for the packet to be released in-sequence at higher layers, respectively.

This problem has already been studied in the literature, but only partial solutions have been provided. Basically, the complexity of the analysis has lead to the introduction of several approximations [1]–[3] or to considering an indirect approach, i.e., to study the transmitter/receiver buffer occupancy [4]–[6], or to account for a zero round trip delay [7] so as to limit the ARQ system memory which needs to be tracked in the analysis.

A first analysis dealing with a round trip time larger than 0 has been proposed in [8], where the independent channel error model has been considered, by providing exact analysis and simple heuristics for the approximation of the delay statistics over channels characterized by a large round trip time. However, the accuracy of the *iid* model heavily depends on the specific radio technology that is considered at the physical layer of the wireless system under analysis, as well as on the wireless channel behavior. A further study, dealing with a Two-State Markov channel, has been proposed in [9], which presents an analytical framework to obtain the delivery delay statistics in the time-varying channel case. However, this work was developed under the assumption of a Two-State Markov packet error model, which may somewhat limit the validity of the analysis. The first contribution of the present paper is to relax this assumption by allowing the packet error process to be described by means of a Markov chain with an arbitrary number of states, which widely generalizes the problem.

Note that the analysis reported in this paper is not only an extension of the study presented for the Two-State Markov channel model [9], that by itself represents a new theoretical contribution. In this paper, we also indicate how accurately the Markov modeling approach can be applied to model the delay behavior of a retransmission protocol layer. To the best of our knowledge, this is the first contribution that provides explicit

Manuscript received July 19, 2004; revised January 20, 2005 and April 1, 2005; accepted April 13, 2005. The associate editor coordinating the review of this paper and approving it for publication was Z. Zhang. Part of this work has been presented at IEEE Globecom 2004.

M. Rossi and L. Badia are with the Department of Engineering, University of Ferrara, via Saragat 1, 44100 Ferrara, Italy (e-mail: {mrossi, lbadia}@ing.unife.it).

M. Zorzi is with the Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy (e-mail: zorzi@ing.unife.it).

Digital Object Identifier 10.1109/TWC.2006.xxxxx.

and quantitative indications in this sense, i.e., on the statistical impact at the protocol layer when the N-State Markov channel is used to capture the Rayleigh fading physical error process. Many past studies (see [10]–[12], among others) dealt with the Markov modeling of the actual fading process, but mainly focused on physical layer aspects rather than considering the impact on the higher layer performance. To this respect, other contributions of the current paper are to devise a new method to derive the Markov chain from the underlying fading statistics and to investigate the appropriate number of states to be considered to obtain the best fitting of protocol layer performance. Regarding the first point, we introduce here a simple but effective strategy that considers the shape of the protocol packet error function directly in the partitioning of the SNR, i.e., to decide how the Signal to Noise Ratio axis has to be subdivided in the Markov channel state assignment. As will be shown in the sequel this method, in spite of its simplicity, gives good results. An important finding of the paper is that the Markov model approach can reproduce with sufficient accuracy the statistics at the protocol layer. This is an important conclusion since, as highlighted in past studies [10], this model often fails at the physical layer, where it is unable to fully capture the complexity of the underlying fading channel. In this case, relevant discrepancies can be observed, for example, in terms of autocorrelation function. However, the same Markov model seems accurate enough to capture higher layer protocol statistics.

A last and minor contribution of the current paper is to extend the analysis above by accounting for an unreliable feedback channel. This aspect has often been neglected in previous studies due to the expected lower error rates over such channels, thanks to the higher degree of FEC that can be used to protect feedback messages and to their smaller size that also implies lower error rates. As will be shown in the sequel, the analysis in this case can be kept unchanged through an expansion of the number of states of the channel transition matrix.

The remaining part of the paper is organized as follows. In Section II, the SR–ARQ transmission process is described. The Markov Channel model is introduced in Section III, where the case including erroneous feedback is also reported. In Section IV, we present the exact analysis for the evaluation of the ARQ delivery delay statistics over an N-State Markov model, whereas in Section V we introduce a new method to obtain an N-State Markov model to represent a quantized Rayleigh fading channel. The results about these two contributions are then reported and compared in Section VI. More specifically, Subsection VI-A presents the analytical results of the delay statistics compared with others obtained by simulation, whereas Subsection VI-B qualitatively and quantitatively discuss the accuracy that can be achieved in terms of protocol layer performance when an N-State Markov model is used to represent the actual error process of a channel characterized by Rayleigh fading. Finally, Section VII concludes the paper.

II. Model for ARQ Queueing and Transmission Processes

Consider a transmitter and a receiver, exchanging packets through a noisy and fading wireless link. For the analysis, we assume that time is slotted, where the slot duration corresponds to the (constant) transmission time for a single packet. We consider a non-zero round-trip time, equal to m slots.

In the SR–ARQ scheme (see [13] for further details), data packets (ARQ PDUs) are transmitted continuously by the sender, whereas the receiver informs the transmitter about packet receptions with acknowledgment (ACK) or negative acknowledgement (NACK) messages, so that as long as ACKs are received, the sender transmits packets in increasing numerical order. We assume that the transmitter adopts a stringent time-out, so that receiving an undecodable feedback packet or not receiving a feedback packet at all within m slots from the transmission of a data PDU is implicitly equivalent to a NACK message for that specific PDU. In this way, the outcome of a transmitted data packet is always known after a full round-trip time. When a negative feedback for a transmission (NACK or timeout) is received, a pre-emptive retransmission is selectively triggered. We assume that a fully reliable Link Layer protocol, i.e., with unlimited retransmission attempts, is used to counteract channel impairments. It is also assumed that both nodes have unlimited buffer size.

We will consider both cases of error-free ACK/NACK messages, which is the simplest possibility, and erroneous feedback channel. In Section III it will be shown that this latter case can be accounted for through an extension of the number of states of the channel transition matrix.

The round-trip delay m is commonly referred to in the literature [1] as the *ARQ window size*. The ACK/NACK message for a packet transmitted in the generic slot t is received after the transmission of up to $m-1$ PDUs (new or retransmitted), i.e., at the end of slot $t+m-1$. This means that in case of NACK the sender always retransmits in slot $t+m$ the PDU transmitted m slots earlier, and at each instant there are exactly m transmissions for which the feedback message is still pending.

Our analysis focuses on the delivery delay; indeed, other delay terms could be considered, such as the queueing delay at the transmitter buffer. The motivation of our choice lies in the fact that the delivery delay is adequate to study the impact of Markov modeling approaches, which is the main focus of the present paper. In addition, the analysis of the delivery delay justifies simpler assumptions for the arrival process, since this delay component is not so sensitive to the traffic intensity, as shown in [1], [5], [9], [14], whereas it is strongly impacted by channel characteristics.

This reasoning allows us to consider a simple model for the arrival process, although our analysis could be extended, if necessary, at the price of additional complexity. Hence, we consider that once a PDU is correctly transmitted, a new one is always present in the source buffer (*Heavy Traffic* [1]) which describes exactly a continuous packet source.

This assumption is often reasonable to model situation of practical interest such as a TCP file transfer (FTP-like session) or video/audio continuous data streaming transmitted using UDP. To motivate the first example, consider the case of a server placed within the fixed Internet and transmitting a data flow to a mobile terminal (MT) connected through a wireless channel to an Access Point (AP). In this case, fully reliable

ARQ can be exploited to promptly recover the errors over the wireless link. If the packet error rate on the wired portion of the network is reasonably low and the buffer at the AP properly dimensioned, TCP can be transmitted from the server to the MT without significant degradation and, in most cases, by filling the wireless channel pipe [15]. If these conditions are verified, the wireless channel is filled by TCP packets and the ARQ buffer is never emptied. However, note that the hypothesis of Heavy Traffic can be relaxed and hence the analysis generalized by following an approach as in [5], if needed.

III. Channel Model

Consider an N-State Discrete Time Markov Chain (DTMC), where the slot duration corresponds to the ARQ packet transmission time. We account here for a general Markov model where states $0, 1, \ldots, \nu-1$ correspond to error–free packet transmission, whereas the remaining states $\nu, \nu+1, \ldots, N-1$ mean erroneous transmission. Formally, each state $n \in \{0, 1, \ldots, N-1\}$ is associated with a packet error probability $P_e[n] = u[n-\nu]$, where $u[\cdot]$ is the unit step, i.e., $u[n] = 1$ if n is greater than or equal to 0, and $u[n] = 0$ otherwise. The model is fully described by the $N \times N$ transition probability matrix $\mathbf{P} = \{p\}_{ij}$, where p_{ij} is the probability that the state in the next slot is j given that the state in the current slot is i.

This formulation can include the case of erroneous feedback channel as well. For instance, assume that the forward and reverse channel are characterized by means of two independent DTMCs. Let $\mathbf{F} = \{f\}_{ij}$ and $\mathbf{R} = \{r\}_{ij}$ be the related channel transition matrices with $G = g+1$ and $S = s+1$ states, respectively. Formally:

$$\mathbf{F} = \begin{pmatrix} f_{00} & \cdots & f_{0g} \\ \vdots & \ddots & \vdots \\ f_{g0} & \cdots & f_{gg} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} r_{00} & \cdots & r_{0s} \\ \vdots & \ddots & \vdots \\ r_{s0} & \cdots & r_{ss} \end{pmatrix} \tag{1}$$

Hence, the whole channel state can be represented by joining the state of both forward and reverse channel. By considering each possible combination of states, we can define a $N \times N$ matrix $\mathbf{P}$, with $N = n+1 = GS$ states:

$$\mathbf{P} = \begin{pmatrix} p_{00} & \cdots & p_{0n} \\ \vdots & \ddots & \vdots \\ p_{n0} & \cdots & p_{nn} \end{pmatrix} = \mathbf{F} \bigotimes \mathbf{R} =$$

$$= \begin{pmatrix} f_{00}r_{00} & \cdots & f_{00}r_{0s} & & f_{0g}r_{00} & \cdots & f_{0g}r_{0s} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ f_{00}r_{s0} & \cdots & f_{00}r_{ss} & & f_{0g}r_{s0} & \cdots & f_{0g}r_{ss} \\ & \vdots & & \ddots & & \vdots & \\ f_{g0}r_{00} & \cdots & f_{g0}r_{0s} & & f_{gg}r_{00} & \cdots & f_{gg}r_{0s} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ f_{g0}r_{s0} & \cdots & f_{g0}r_{ss} & & f_{gg}r_{s0} & \cdots & f_{gg}r_{ss} \end{pmatrix} \tag{2}$$

where $\bigotimes$ is the Kronecker product between matrices [16]. The channel can then be modeled by a DTMC, whose transition matrix is $\mathbf{P}$. The number of states has been expanded to $N = GS$. If the forward and backward channel have γ and σ error-free states over G and S respectively, it is straightforward to prove that the number of good states of the resulting DTMC is $\gamma\sigma$. In Subsection VI-A we will show that in practical cases, instead of following the exact approach presented above (i.e., by means of the Kronecker product between matrices), it is possible to account for the feedback channel by just increasing the forward channel error rate.

The channel model presented above (including the extension for erroneous feedback) comprises, as a particular case, the widely used [10], [17] K-state model where $\mathcal{S} = \{0, 1, ..., K-1\}$ is the set of the states and for $i, j \in \mathcal{S}$, $\epsilon_i \in [0, 1]$ is the PDU error rate in state i, i.e., a PDU transmitted when the channel is in state i is erroneous with probability ϵ_i, whereas t_{ij} is the transition probability from state i to state j. This last Markov Chain is completely specified by the pair $(\mathbf{T}, \mathcal{E})$, where $\mathbf{T} = \{t\}_{ij}$ is the transition probability matrix and $\mathcal{E} = (\epsilon_i)$ is the state error probability vector. The model is equivalent to considering $N = 2K$ states (with $\nu = K$) in $\mathcal{S}' = \{0, 1, \ldots, \nu-1 = K-1, \nu = K, \ldots, N-1\}$, where the transition probabilities p_{ij}, $0 \leq i, j \leq N-1$ are derived as follows:

$$p_{ij} = \begin{cases} t_{xy}(1-\epsilon_y) & j \in \{0, 1, \ldots, K-1\} \\ t_{xy}\epsilon_y & j \in \{K, K+1, \ldots, N-1\} \end{cases} \tag{3}$$

where $x = i - Ku[i-K]$, $y = j - Ku[j-K]$, $0 \leq x, y \leq K-1$. Moreover, note that in the extended model states $\{0, 1, \ldots, K-1\}$ are error free, i.e., a PDU is always transmitted correctly in these states, whereas in states K through $N-1$ PDUs are always transmitted erroneously.

This procedure will be extensively used throughout the paper, in particular we will derive a Markov representation of a Rayleigh fading channel with K states, each of them with error probability ϵ_i, and this will be extended to an N-State Markov chain as explained above. This latter Markov chain will be used to track the ARQ layer packet delivery process. Hence, by jointly considering the two contributions of this paper, i.e., Markov models of fading channels and analysis of the SR–ARQ delivery delay for Markov channels, we will be able to obtain the delivery delay statistics directly from the physical channels parameters. This will also lead in Subsection VI-B to discuss the appropriate number of states of the involved Markov chains.

IV. Computation of the Delivery Delay Statistics in an N-State Markov Channel

Computing the delivery delay statistics for a single reference PDU, called in the sequel *tagged PDU*, transmitted using Selective Repeat ARQ, can be done by tracking the successful reception of the tagged PDU, as well as all PDUs with lower identifier. In fact, PDUs are always released *in-order* to the upper layers.

First of all, suppose the tagged PDU is transmitted for the first time in slot $t = m$. It can be proven that this implies that all previous PDUs (i.e., those with smaller identifier), excluding at most the $m-1$ PDUs transmitted in slots 1 through $m-1$, have been successfully received, and that in slot

0 a successful transmission occurred (otherwise in slot m we would have a retransmission). Hence, the delivery delay of the tagged PDU equals the time required for the correct reception of all packets in the m-sized window comprising slots from 1 to m, which will be called *fundamental window*, due to its key role in the analysis. The formal proof of this statement can be found in [9], where an algorithm to evaluate the resolution time of the entire fundamental window is presented for the Two-State Markov model. Here, we extend this method to the more general N-State case, as outlined below.

Before proceeding with the analysis, note that due to the finite round-trip time, there is a constant time gap between the transmission of a PDU and its arrival at receiver's side. Let us call this value t_c. It is straightforward to see that t_c, being just a constant term approximately equal to $m/2$, can be neglected for the analysis in order to simplify the notation. Hence, in the following we will consider the statistics as the probability $P_d[k]$ that the delivery delay equals k slots plus the constant t_c. We will often omit this by speaking, for the sake of simplicity, of a delay equal to k slots, even though the full delay must always include also the additional constant term equal to t_c.

Consider now the first transmission of the fundamental window. Some of the PDUs transmitted during slots $1, 2, \ldots, m$ are correctly received. In this case, we denote the corresponding slots as *resolved*, to indicate that such PDUs do not need to be retransmitted. In other words, a resolved slot contains a PDU which does not block the release of the tagged one. If a slot i is resolved, all slots $i + \kappa m$, with κ a positive integer, can be marked as resolved. In fact, they correspond to the transmission of a PDU with higher id than the tagged one, and therefore can be ignored for the delivery analysis. On the other hand, if the transmission in the fundamental window is erroneous, we denote the corresponding slot j as *unresolved*. Such a label means that this slot prevents the tagged PDU from being released, and that m slots later, i.e., in slot $j + m$, a retransmission will take place. If the retransmission is successful, the corresponding slot, and also every slot $j + \kappa m$, with κ positive integer, will be marked as resolved. Otherwise, $j + m$ is also marked as unresolved and the procedure is repeated. Therefore, the tagged PDU is released after the mth slot of a sequence of consecutive slots marked as resolved, i.e., when the last unresolved slot becomes resolved.

To show the behavior of the algorithm, consider this simple example. If $m = 3$ and the channel starts from a good state and then is alternatively bad or good, the algorithm ends in slot 5 after the following outcome: 1=resolved, 2=unresolved, 3=resolved, 4=resolved (despite the channel error, as it was previously marked), 5=resolved, and every further slot is also resolved.

Since this algorithm requires to check whether a sequence of m resolved slots occurs, it can be applied by tracking the resolved/unresolved status of the $m-1$ most recent past slots, which at time t are slots $t-m+1, t-m+2, \ldots, t-1$. We carry this information with a vector of binary variables b_j, for $j = 0, 1, \ldots, m-2$, so that $b_j = 1$ if slot $t - m + 1 + j$ is still unresolved, and $b_j = 0$ otherwise. In this way we obtain a string of bits denoted by $\mathbf{b}$ that keeps memory of which slots are yet to be resolved. To simplify the notation, in the sequel we represent the bitmap $\mathbf{b}$ as the integer $i = \sum_{j=0}^{m-2} b_j 2^j$. Instead, the status of the current slot, i.e., slot t, is not depicted by a simple binary variable, since in this case also the channel state has to be tracked. This is the only information corresponding to a channel state required in the analysis, as the Markovian nature of the channel allows to ignore the channel state in slots $t-m+1, t-m+2, \ldots, t-1$ once it is known in slot t. To represent this last information, we associate to the last PDU a variable ω, which has $2N - \nu$ possible values. In fact, there are three main cases:

i) the channel is good, which implies that the slot is resolved (it does not matter in this case if the slot was already resolved, or it is resolved exactly now). This possibility comprises ν states, where ν is the number of error-free states of the Markov channel.
ii) the channel is bad, but the slot was resolved in a previous transmission. There are $N - \nu$ possibilities to be in this states, one for each erroneous state of the channel.
iii) the channel is bad and the slot remains unresolved as was before. As the previous one, this comprises $N - \nu$ cases.

Therefore, ω has a value in the ranges $\{0, 1, \ldots, \nu-1\}$ in case i), $\{\nu, \nu+1, \ldots, N-1\}$ in case ii), $\{N, N+1, \ldots, 2N-\nu-1\}$ in case iii), so that ω is equal to the channel state in cases i) and ii), whereas for case iii) the value of ω equals the channel state augmented by $N - \nu$.

Consider now the random process $X(t) = (i(t), \omega(t))$ which jointly tracks slot-by-slot the Markov channel evolution and the status of the m latest slots. According to the above definitions and discussion, this process is a Markov chain. In general, each state has $\mathcal{M}$ possible values, where $\mathcal{M}$ depends on the structure of $X(t) = (i(t), \omega(t))$. Since $i(t)$ can assume 2^{m-1} possible values (remember that $i(t)$ has a binary representation with $m-1$ digits) and $\omega(t)$ belongs to $\{0, 1, \ldots, 2N-\nu-1\}$, we have that $\mathcal{M} = (2N-\nu)\cdot 2^{m-1}$. According to this description, the resolution of the fundamental window corresponds to the first transition of the Markov chain $X(t)$ through one of the states $(0,0), (0,1), (0,2), \ldots (0, N-1)$. In fact, the $m-1$ past PDUs are resolved if $i = 0$ and the current one is resolved if $\omega < N$. In other words, we must account for every combination of the fully resolved window with any channel state.

In order to determine the possible transitions and the corresponding transition probabilities, assume the values of $\mathbf{b} = (b_0, b_1, \ldots, b_{m-2})$ and ω at time t are known. At time $t + 1$ the new values $\mathbf{b}'$ and ω' of these variables depend on $\mathbf{b}$, ω and, due to the Markov nature of the channel, on the channel state at time $t + 1$, called y, $0 \le y \le N - 1$. In particular, $\mathbf{b}'$ is a clocked version of $\mathbf{b}$ into the past, i.e., $(b'_0, b'_1, \ldots, b'_{m-3}, b'_{m-2}) = (b_1, b_2, \ldots, b_{m-2}, f(\omega))$. The last bit of $\mathbf{b}'$, which depends also on ω, is $f(\omega) = 1$ if $\omega \ge N$ (current slot at time t was still unresolved), and $f(\omega) = 0$ if $\omega < N$. More compactly, $b'_{m-2} = f(\omega) = u[\omega - N]$. For what concerns ω', if $b_0 = 0$ the corresponding slot has already been resolved, and therefore $\omega' = y$, i.e., $0 \le \omega' \le N - 1$ according to the channel state y at time $t + 1$. On the other hand, if $b_0 = 1$, the slot is still unresolved at time t, and therefore we have $0 \le \omega' \le \nu - 1$ if the channel at time $t+1$ is good (slot is resolved at this time) and $N \le \omega' \le 2N-\nu-1$

otherwise (slot remains unresolved). In the former case, it is again $\omega' = y$, whereas in the latter $\omega' = y + N - \nu$. Note that given $X(t)$ there are only N possible destinations for $X(t+1)$, since the shift of the bitmap is deterministic and the only random variable is y which can assume N values. Formally, the transition probabilities from (i,ω) to (i',ω') are:

- if i is even (i.e., $b_0 = 0$), then:

$$\begin{aligned} P[X(t+1) &= (i',\omega')|X(t)=(i,\omega)] \\ &= \begin{cases} p_{xy} & \text{if } i' = \lfloor \frac{i}{2} \rfloor + u[\omega - N]2^{m-2}, \\ & x = \omega - (N-\nu)u[\omega - N], \\ & \omega' = y, y = 0,1,\ldots,N-1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (4)$$

- if i is odd (i.e., $b_0 = 1$), then:

$$\begin{aligned} P[X(t+1) &= (i',\omega')|X(t)=(i,\omega)] \\ &= \begin{cases} p_{xy} & \text{if } i' = \lfloor \frac{i}{2} \rfloor + u[\omega - N]2^{m-2}, \\ & x = \omega - (N-\nu)u[\omega - N], \\ & \omega' = y + (N-\nu)u[y-\nu], \\ & \quad y = 0,1,\ldots,N-1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

where the use of $\omega' = y + (N-\nu)u[\omega - N]$ in the latter case means that a good channel $0 \le y \le \nu - 1$ leads to $\omega' = y$ whereas a bad channel $\nu \le y \le N-1$ leads to $\omega' = y+N-\nu$, $N \le \omega' \le 2N - \nu - 1$, i.e., a situation of bad channel and unresolved slot. According to the above rules, the transition probability matrix can be built, which will have only N non-zero entries per row.

In order to find the delay statistics, we proceed as follows. First of all, let us define an appropriate function τ:

$$\begin{aligned} \tau : \mathcal{I}_N^{m-1} &\rightarrow \{0,1\}^{m-1} \\ \tau(\boldsymbol{\beta}) = \tau(\beta_0,\beta_1,\ldots,\beta_{m-2}) &= (d_0,d_1,\ldots,d_{m-2}) \\ \text{s.t. } d_j &= u[\beta_j - \nu] \\ \forall j &= 0,1\ldots,m-2 \end{aligned} \quad (6)$$

where $\mathcal{I}_N = \{0,1,2,\ldots,N-1\}$. The meaning of $\tau(\cdot)$ is to transform vectors ($\boldsymbol{\beta}$) of base-N digits into binary digits so that the output digit is 0 if the input digit is less than ν, and 1 otherwise. That is, if $\boldsymbol{\beta} = (\beta_0,\beta_1,\ldots,\beta_{m-2})$ contains the Markov channel states ($\beta_j \in \{0,1,2,\ldots,N-1\}$) in $m-1$ consecutive slots, each element of $\mathbf{d} = (d_0,d_1,\ldots,d_{m-2})$ is a binary digit equal to 0 for a slot where a successful transmission has occurred (good channel), whereas $d_j = 1$ corresponds to an erroneous transmission in slot j (bad channel state).

Let $\mathbf{\Pi} = [\Pi_0 \ \Pi_1 \ \cdots \ \Pi_{\mathcal{M}}]^T$ be a column vector whose $\mathcal{M} = (2N-\nu)\cdot 2^{m-1}$ scalar entries represent the probabilities that the system starts in a given state. This *starting state* is defined as the system state at time $t = m$, when the tagged packet is assumed to be transmitted, and can be decomposed in the evaluation of $i(m)$ and $\omega(m)$.

It is easy to see that the former, which corresponds to the resolved/unresolved status for all the slots 1 through $m-1$, only depends on the channel evolution. In fact, every slot of the fundamental window is marked according to the channel state only. Thus, the function $\tau(\cdot)$ can be applied to the vector of channel states during the fundamental window, giving 0 or 1 according to the channel state being good or bad. If we consider the binary-wise representation of $i(t)$, which we have already named $\mathbf{b}$, and assume that $\boldsymbol{\beta}$ is a sequence of values in $\{0,1,\ldots,N-1\}$ representing the evolution of the channel from slot 1 to slot $m-1$, it is easy to recognize that $\mathbf{b} = \tau(\boldsymbol{\beta})$. In other words, the function $\tau(\cdot)$ is used to translate the information regarding the channel state in slots 1 through $m-1$ (N possible values) into the corresponding resolved/unresolved status (two possible values).

To evaluate the value of $\omega(m)$ instead, we have to keep in mind that the channel in slot 0 is bound to be error-free. Hence, $\mathbf{\Pi}$ is computed as follows:

- if $\omega \in \{0,1,\ldots,\nu-1\} \cup \{N,N+1,\ldots,2N-\nu-1\}$:

$$\mathbf{\Pi}_{(i,\omega)} = \sum_{z=0}^{\nu-1} \frac{\pi_z}{\mathcal{S}_\pi} \sum_{\boldsymbol{\beta}\in\mathcal{G}_{\mathbf{b}}} p_{z\beta_0} \left[\prod_{j=0}^{m-2} p_{\beta_{j-1}\beta_j} \right] p_{\beta_{m-2}g} \quad (7)$$

- if $\omega \in \{\nu,\nu+1,\ldots,N-1\}$:

$$\mathbf{\Pi}_{(i,\omega)} = 0 \quad (8)$$

where $\mathcal{S}_\pi = \sum_{j=0}^{\nu-1}\pi_j$, $g = \omega - (N-\nu)u[\omega - N]$ and $\mathcal{G}_{\mathbf{b}} = \{\boldsymbol{\beta} \in \mathcal{I}_N^{m-1} : \tau(\boldsymbol{\beta}) = \mathbf{b}\}$ and π_z, $z \in \{0,1,\ldots,N-1\}$ is the Markov Channel steady state probability of state z. These equations simply track all possible combinations of initial state at time 0 (which is constrained to be good, hence between 0 and $\nu-1$) and evolution of the channel during the fundamental window, represented by the vector $\boldsymbol{\beta}$ (which must satisfy $\tau(\boldsymbol{\beta}) = \mathbf{b}$). Also, note that it is impossible to have $\omega(m) \in \{\nu,\nu+1,\ldots,N-1\}$ since it corresponds to have a resolved slot but an erroneous channel at time m. Since slot m corresponds to the *first* instance of transmission for the tagged PDU, the only possibilities are that the slot is either resolved or unresolved according to the channel state.

In this way, the starting state is determined. For a slot $t > m$ instead the evolution is more complicated, since it depends on previous states also. For example, a slot can be resolved even if the channel is in an erroneous state, hence $i(t)$ no longer corresponds to the channel state only. However, by exploiting Markovian behavior of the system, since $(i(m),\omega(m))$ is known we can evaluate $(i(t),\omega(t))$ also for $t > m$ by recursively applying Eqs. (4) and (5). This can be done in a compact way as follows.

Let $\boldsymbol{e}_0 = [(i_0,\omega_0) \ (i_1,\omega_1) \ \cdots \ (i_{\mathcal{M}},\omega_{\mathcal{M}})]^T$ be a column $\mathcal{M}$-sized vector of all zeros except for the entries corresponding to states $(i,\omega) \in \{(0,0),(0,1),\ldots,(0,N-1)\}$, that are equal to 1. According to the previous reasoning, these are the only states where the fundamental window is resolved. If $\mathbf{\Phi}$ is the transition matrix of the Markov chain $X(t)$, we determine:

$$\mathcal{P}_c[k] = \mathbf{\Pi}\mathbf{\Phi}^k \boldsymbol{e}_0, \quad k \ge 0\,. \quad (9)$$

$\mathcal{P}_c[k]$ is the probability that the delivery delay is less than or equal to k slots plus the propagation delay t_c (which however, being just a constant term, can be neglected as previously discussed). Finally, the delivery delay statistics $P_d[k]$ is determined as:

$$P_d[0] = \mathcal{P}_c[0]\,, \quad P_d[k] = \mathcal{P}_c[k] - \mathcal{P}_c[k-1] \quad \forall k > 0\,. \quad (10)$$

V. Derivation of the N-State Markov Model

In this section we discuss possible procedures to derive an N-State Markov model of a Rayleigh fading channel, including both known and original proposals. In fact, several techniques to deal with this problem have been presented in previous research (see [10]–[12], among others). Here, the simple method of [10] is considered first. Later on, in order to gain some understanding on the impact of the selected method, we propose a new approach which takes into account the shape of the packet error probability function. This is done with the aim of using a given number of Markov channel states (N) in an efficient way, i.e., in order to better describe the packet error behavior at the ARQ layer. Since the scope of this section it mainly to validate the analysis presented above, we limit here our investigation to these two techniques. In spite of their simplicity, these methods are effective and lead to an accurate description of the ARQ delay statistics.

Consider the transmission system introduced in Section II and let Γ denote the received signal to noise ratio (SNR). The pdf of Γ is exponential as follows [10]:

$$p_\Gamma(\gamma) = \frac{1}{\gamma_0} e^{-\gamma/\gamma_0}, \;\; \gamma \geq 0 \tag{11}$$

where $\gamma_0 = E[\Gamma]$. Let $0 = \Gamma_0 < \Gamma_1 < \cdots < \Gamma_{K-1} < \Gamma_K = +\infty$ be $K+1$ thresholds for the SNR. The Rayleigh channel is said to be in state $x = 0, 1, \ldots, K-1$ if the received SNR is in the interval $[\Gamma_x, \Gamma_{x+1})$. Moreover, associated with each state x there is an error probability ϵ_x that is the PDU error rate experienced in state x. We define $\mathcal{F}(\gamma)$ as the function mapping the instantaneous SNR level γ into the conditional PDU error probability. Once the threshold levels are chosen for every state, the PDU error rate in the generic state x is found as $\epsilon_x = (\int_{\Gamma_x}^{\Gamma_{x+1}} \mathcal{F}(\gamma) p_\Gamma(\gamma) \mathrm{d}\gamma)/\theta_x$, where θ_x is the steady state probability to be in state x. In this work we assume a $\pi/4$-DQPSK modulation scheme [11], i.e., the bit error probability can be approximated as $\varepsilon(\gamma) \approx (4/3)\mathrm{erfc}(\sqrt{\gamma})$. $\mathcal{F}(\gamma)$ is then derived as $1-(1-\varepsilon(\gamma))^L$, where L is the ARQ packet length expressed in bits.[1] The steady state probability θ_x is computed as:

$$\theta_x = \int_{\Gamma_x}^{\Gamma_{x+1}} p_\Gamma(\gamma)\, \mathrm{d}\gamma = e^{-\Gamma_x/\gamma_0} - e^{-\Gamma_{x+1}/\gamma_0} \tag{12}$$

The simplest approach for choosing the SNR thresholds [10] is to consider $\theta_x = 1/K$, $\forall\, x = 0, \ldots, K-1$. In this case, the threshold levels can be easily estimated by recursively applying Eq. (12), given that Γ_0 is known. However, this procedure leads to a rough estimation of the underlying fading process (see [11], [12]). For this reason, we consider here an improved threshold selection criterion. We first choose two numbers, ℓ_1 and ℓ_{K-1} so that ℓ_1 is close to 1 and ℓ_{K-1} is close to zero. Then we choose the first (Γ_1) and the last (Γ_{K-1}) unknown thresholds such that $\Gamma_1 = \mathcal{F}^{-1}(\ell_1)$ and $\Gamma_{K-1} = \mathcal{F}^{-1}(\ell_{K-1})$. Once Γ_1 and Γ_{K-1} are known, θ_0 and θ_{K-1} can be evaluated by Eq. (12). In this procedure we assign first the states 0 and $K-1$ to the SNR levels corresponding to a PDU error rate that is larger than ℓ_1 and smaller than ℓ_{K-1}, respectively. At this point, we use the remaining $K-2$ states to characterize the SNR interval between Γ_1 and Γ_{K-1}, i.e., where the PDU error rate is in the range $[\ell_1, \ell_{K-1}]$. The remaining $K-2$ thresholds are chosen to satisfy $\theta_x = (1-\theta_0-\theta_{K-1})/(K-2) = e^{-\Gamma_x/\gamma_0} - e^{-\Gamma_{x+1}/\gamma_0}$. A graphical representation of this procedure is reported in Fig. 1. In practice, with our novel method, thresholds Γ_1 and Γ_{K-1} are moved according to the shape of the error probability function, thereby reducing the space between them. The remaining $K-2$ states are used to map the range $[\Gamma_1, \Gamma_{K-1})$ with a finer partitioning, so that we have a higher number of equivalent states tracking this interval with respect to previous techniques (see Fig. 2). In other words, we concentrate the SNR quantization on the most critical part in terms of error probability. This is the reason of the good performance that will be shown later on in Subsection VI-B.

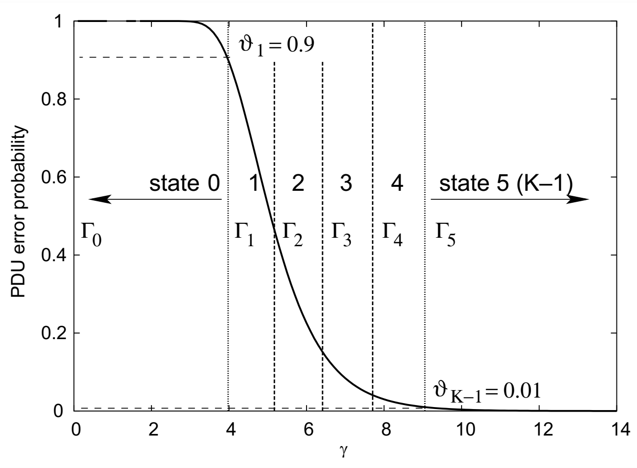


Fig. 1. SNR partition method: example for $K = 6$.

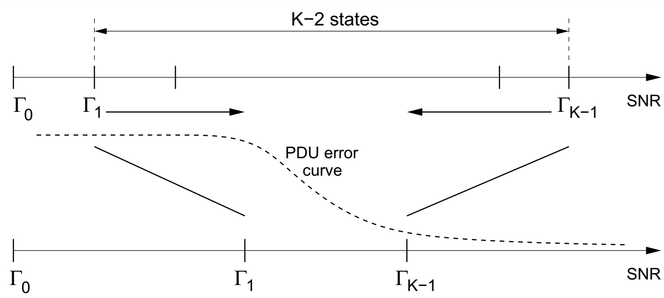


Fig. 2. Graphical representation of the novel SNR partition method.

Once the thresholds have been computed, the transition probabilities are derived as in [12] according to:

$$\begin{aligned} t_{ij} &= \frac{\int_{\zeta_i}^{\zeta_{i+1}} \int_{\zeta_j}^{\zeta_{j+1}} f_{R_1R_2}(r_1, r_2, \rho) dr_1 dr_2}{\theta_i} \\ f_{R_1R_2}(r_1, r_2, \rho) &= \frac{4 r_1 r_2}{\lambda} e^{-(r_1^2+r_2^2)/\lambda} I_0(2\rho r_1 r_2/\lambda) \end{aligned} \tag{13}$$

where $\zeta_i = \sqrt{\Gamma_i/\gamma_0}$, $f_{R_1R_2}(r_1, r_2, \rho)$ is the bivariate Rayleigh joint pdf [12], $\lambda = 1-\rho$, $\rho = J_0(2\pi f_d T_p)$ is the correlation of two samples of the underlying Gaussian process that are spaced by T_p seconds, f_d is the Doppler frequency, T_p is the ARQ PDU transmission duration, $J_0(\cdot)$ and $I_0(\cdot)$ are the Bessel function and the modified Bessel function of the first

[1] More complicated expressions could be used to account for the use of error correction codes.

kind and order zero.

By using this method, a Markov Chain specified by $(\mathbf{T}, \mathcal{E})$ can be derived, where $\mathbf{T} = \{t\}_{ij}$, $\mathcal{E} = (\epsilon_i)$ with $i, j = 0, 1, \ldots, K-1$. To evaluate the delivery delay with the theoretical procedure described in Section IV, it is necessary to transform this Markov chain into a modified version with N states. This can be done by following the approach explained in Section III, in particular by considering Eq. (3) which relates $(\mathbf{T}, \mathcal{E})$ to a $N \times N$ matrix $\mathbf{P}$, where each of the K intervals is mapped into 2 states, one error-free and one always erroneous. Thus, the number of error-free states ν is equal to K (and so is the number of erroneous states), i.e., $N = 2K$. The matrix $\mathbf{P}$ obtained in this way is finally used to derive the ARQ delay statistics as in Section IV.

VI. RESULTS AND DISCUSSIONS

In this section, we report some examples for the delivery delay statistics and we discuss the goodness of a Markov channel model in the approximation of the ARQ packet delay statistics in a Rayleigh fading channel. In the following Subsection VI-A, some examples for the link layer delivery delay statistics with erroneous feedback and over a fading channel are reported first. Later on, in Subsection VI-B, the accuracy of the Markov modeling approach will be discussed considering the effect of the Doppler frequency f_d and of the number of Markov channel states N.

A. Results for the Delay Statistics over an N-State Markov Channel

First of all, we present some results showing the impact of erroneous feedback channel, introduced in Section III. We consider two independent DTMCs for forward and reverse channel. For the sake of simplicity, we account for two states only (good and bad channel) in both DTMCs ($G = S = 2$). Hence, the number of states of the DTMC for the whole channel, described by the matrix $\mathbf{P}$ is four, according to the analysis in Section III. We refer to ε_f and ε_r as the steady-state error probabilities of forward and reverse channel, respectively. Formally, $\varepsilon_f = f_{01}/(f_{01} + f_{10})$ and $\varepsilon_r = r_{01}/(r_{01} + r_{10})$. Now, to completely specify the channel matrices it is sufficient to define the average error burst length, that can be computed as $1/f_{10}$ and $1/r_{10}$ for the forward and reverse channel, respectively. The resulting matrix $\mathbf{P}$ has one good state ($\nu = 1$, i.e., both forward and feedback channels are error-free) and three bad states (either forward or feedback channel, or both, are bad). The analysis in Section IV can therefore be applied to this channel matrix to obtain the delay statistics in the erroneous feedback case. In the following, we will consider two cases, an *iid* channel, where the burst length is derived as $1/(1-\varepsilon_f)$ and $1/(1-\varepsilon_r)$, for the forward and reverse channel, respectively, and a bursty channel, with a given burst length b (equal to 15 in the reported results) which is the same for both channels.

In Figs. 3 and 4, we consider different values of ε_f and plot the mean value and variance of the delivery delay, as a function of $\rho = \varepsilon_r/\varepsilon_f$. In these figures, both mean and variance have been normalized to the case of no feedback errors and the round trip time is $m = 6$. It is emphasized that for low values of the ratio $\varepsilon_r/\varepsilon_f$ the mean delay in the presence of feedback errors (Fig. 3) is approximately increased by a factor $1 + \rho$ (i.e., the normalized delay is approximately linear in ρ), which means that the effect of the reverse channel impairments can be translated on the forward channel by considering an equivalent steady state error probability $\varepsilon = \varepsilon_f + \varepsilon_r$. The lower ε_f, the better the approximation. A similar phenomenon holds also for the variance metric (Fig. 4), even though there is a slightly larger discrepancy between the curves and the linear behavior in ρ.

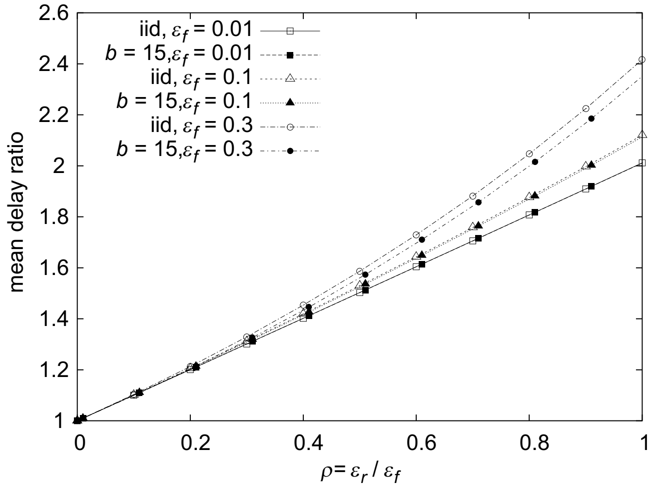


Fig. 3. Mean delivery delay ratio vs. $\varepsilon_r/\varepsilon_f$ $(m = 6)$.

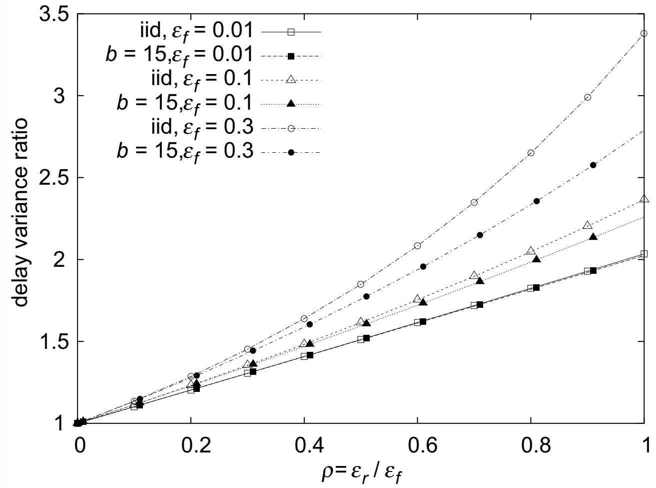


Fig. 4. Delivery delay variance ratio Vs. $\varepsilon_r/\varepsilon_f$ $(m = 6)$.

A general conclusion which holds for many cases of interest is that the effect of an erroneous reverse channel can be simply accounted for by increasing the forward channel error probability. This approximation is good when $\varepsilon_r \ll \varepsilon_f$, which is reasonable in realistic cases as usually acknowledgement packets are significantly shorter than data packets, resulting in a lower packet error rate. Moreover, in force of their smaller size, they can be more easily protected by means of FEC techniques. For these reasons, the results discussed in the following will consider always error-free ACK/NACKs, even though they can be promptly extended as discussed in Section III to the erroneous feedback channel case.

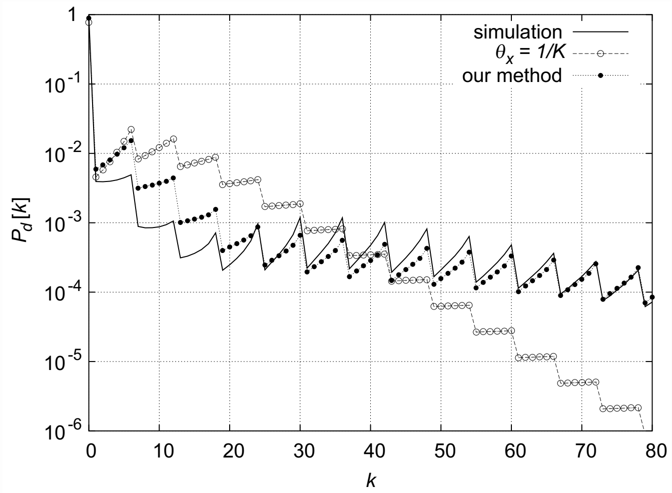


Fig. 5. Delivery delay statistics: comparison between Markov Channel analysis and Rayleigh channel simulation with $L = 360$ bits, bit rate 1024 Kbps, $f_d = 10$ Hz, $f_d T_p = 0.00343$, $m = 6$, $N = 6$.

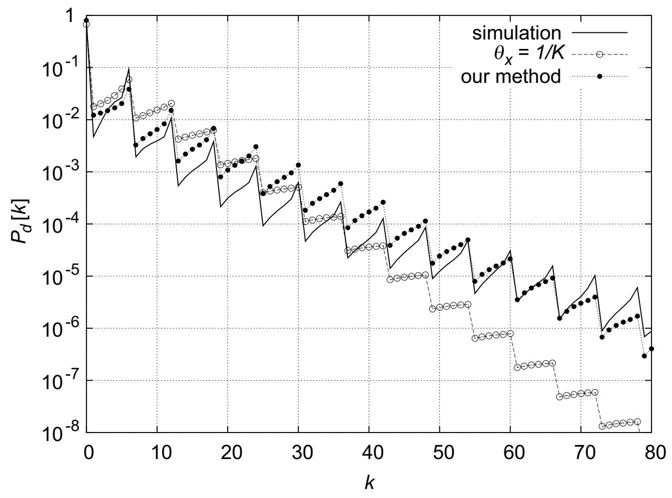


Fig. 6. Delivery delay statistics: comparison between Markov Channel analysis and Rayleigh channel simulation with $L = 360$ bits, bit rate 1024 Kbps, $f_d = 100$ Hz, $f_d T_p = 0.03433$, $m = 6$, $N = 6$.

We focus now on the second contribution of the paper, i.e., Markov modeling of Rayleigh fading channels. In Figs. 5 and 6, we report the delivery delay statistics considering $f_d = 10$ Hz and $f_d = 100$ Hz, respectively. In both graphs, the statistics obtained by simulation is compared against the two threshold selection methods, i.e., the equal probability method ($\theta_x = 1/K$) presented in [10] and the novel procedure presented in the previous Section. It shall be observed that our model better succeeds in approximating the real behavior. However, a Markov approximation of the actual channel error process is, in general, not able to perfectly match the real statistics. This is, indeed, a limitation of the Markov model that, even when a large number of states is considered, does not perfectly fit the actual fading process statistics. However, it is worth noting that fading is a complex process that we are trying to approximate using a relatively simple model. In this view, our approach leads to statistics that are reasonably close to the real behavior. We also shall observe that the differences in the delay performance between Markov modeling and actual

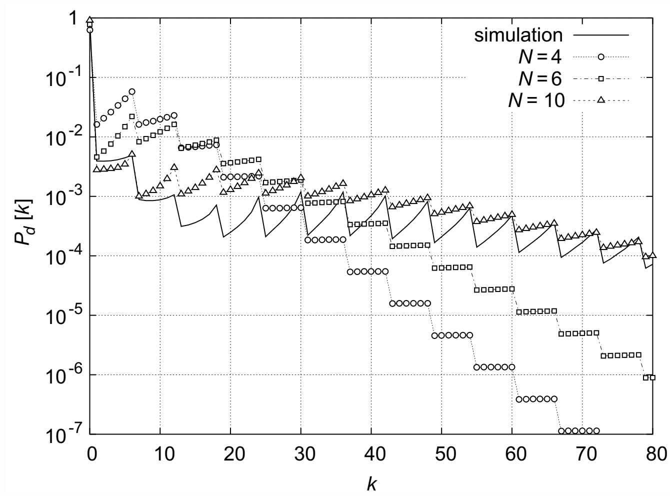


Fig. 7. Delivery delay statistics: comparison between Markov Channel analysis derived from the uniform SNR partitioning method ($\theta_x = 1/K$) and Rayleigh channel simulation as a function of k by varying N and considering $L = 360$ bits, bit rate 1024 Kbps, $f_d = 10$ Hz, $m = 6$.

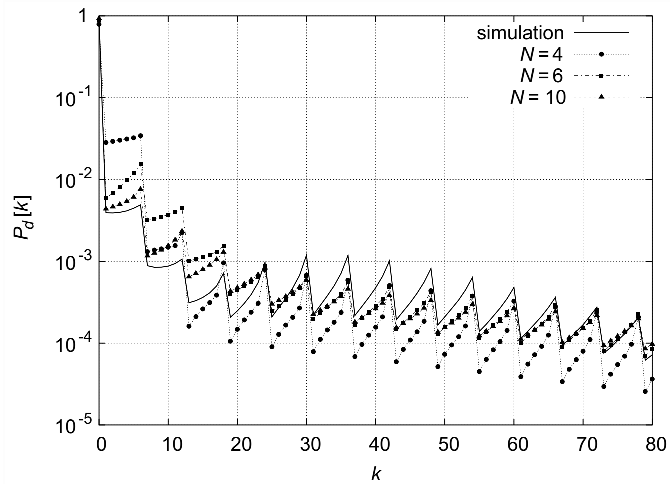


Fig. 8. Delivery delay statistics: comparison between Markov Channel analysis derived from our SNR partitioning method and Rayleigh channel simulation as a function of k by varying N and considering $L = 360$ bits, bit rate 1024 Kbps, $f_d = 10$ Hz, $m = 6$.

Rayleigh channel are very small in spite of the substantial gaps observed in the autocorrelation function [12]. This supports the validity of the Markov approach and also allows us to infer that not all the physical properties of the underlying channel have to be taken into account in order to accurately model ARQ protocol performance.

In Figs. 7 and 8 we report some curves to qualitatively discuss[2] the dependence on the number of states of the Markov model, N. In more detail, Fig. 7 shows the uniform SNR partitioning, whereas in Fig. 8 our novel approach is reported. For low values of N, the fit between simulation and analysis substantially improves by increasing the number of states. However, this trend does not hold indefinitely. In the uniform method, $N = 10$ is emphasized as a better choice than lower values, but in our approach when $N \geq 6$ no further significant improvements are observed, since the modeling becomes better in the first few rounds but slightly worse

[2]A quantitative comparison is reported in Section VI-B.

elsewhere. In general, our method exhibits good performance with a lower value of N than uniform partitioning. However, as N increases both methods do not improve that much and also the advantage in using our approach vanishes. Thus, our proposal appears to be suitable to derive good results with low values of N (low-complexity models).

To have a large number of states is inefficient from the computational point-of-view and also seems to fail to accurately model the channel beyond a certain limit. In order to obtain better results within a Markov model it would be interesting to investigate how the statistics improves considering a radically different approach to derive the Markov chain. For instance, in [18] the authors considered the fading derivative as an additional dimension for this purpose. In that paper, they proved that this method can better reproduce the oscillatory behavior of the autocorrelation function. In [19] some results are reported concerning the mean throughput value of the SR–ARQ protocol. Further investigations on how these techniques can improve the ARQ delivery statistics are left for future research.

B. On the Accuracy of Link Layer Statistics Obtained by means of the Markov Modeling Approach

In this subsection, we present some results on the accuracy of the delay statistics derived in Section V. We refer here to $P_d^{an}[k]$ as the statistics derived analytically, i.e., to the distribution derived by means of the Markov modeling approach, whereas we refer to $P_d^{sim}[k]$ as the delay distribution, which has been directly measured by simulating the SR–ARQ algorithm discussed in Section II over a Rayleigh fading channel. To introduce Rayleigh fading behavior we use the well-known Jakes model [20].

In order to weigh the difference between these two statistics, we consider here the Kullback Leibler distance (see [21], p.18). The Kullback Leibler distance $D(p \parallel q)$ between two generic distributions $p[n]$ and $q[n]$ is a measure of the inefficiency of assuming that the exact distribution is $q[n]$ when the true distribution is $p[n]$. For example, if we knew the real distribution of the random variable ($p[n]$), then we could construct a code with average length $H(p)$ (where H is the entropy associated to the distribution p). If, instead, we used the code to describe the approximate distribution $q[n]$, we would need $H(p) + D(p \parallel q)$ bits on average to describe the random variable. The Kullback Leibler distance arises as an expected logarithm of the likelihood ratio of the two distributions:

$$D(p \parallel q) = \sum_{n \in \mathcal{N}} p[n] \log \frac{p[n]}{q[n]} \tag{14}$$

where $\mathcal{N}$ is the (common) domain set of the distribution functions.

Fig. 9 reports the distance between the two distributions $P_d^{an}[k]$ and $P_d^{sim}[k]$ as a function of the number of states N used to build the Markov chain, whereas in Figs. 10 and 11 we plot on the x-axis the Doppler frequency (f_d). In Fig. 9 we consider the i.i.d. case, taken as a reference value, by assuming a constant error probability, regardless of the state of the Markov chain. Our partitioning approach is used with

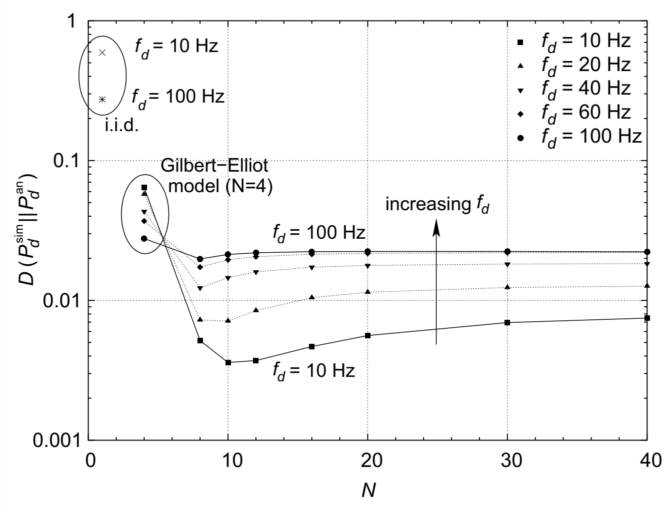


Fig. 9. Kullback Leibler distance between the distributions obtained by simulation ($P_d^{\mathrm{sim}}[k]$) and analysis ($P_d^{\mathrm{an}}[k]$) using our SNR partitioning approach as a function of the number of states N, for different Doppler frequencies f_d.

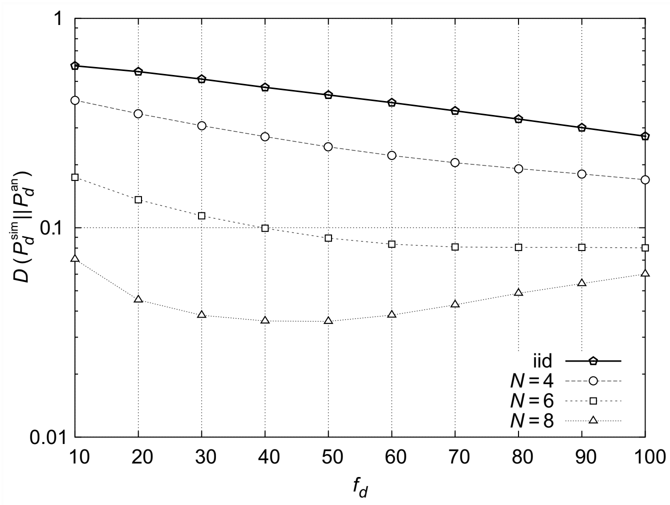


Fig. 10. Kullback Leibler distance between the distributions obtained by simulation ($P_d^{\mathrm{sim}}[k]$) and analysis ($P_d^{\mathrm{an}}[k]$) using the uniform SNR partitioning approach ($\theta_x = 1/K$) as a function of the Doppler frequency f_d, for different values of N.

$\ell_1 = 0.9$ to derive the case $N = 4$, which corresponds to a classic Gilbert-Elliot channel model. Then, $\ell_{K-1} = 0.01$ is added for the cases with $N \geq 6$. Several curves are plotted for all these values of N by considering different Doppler frequencies. It is interesting to observe that the distance metric has a minimum, i.e., an optimal number of states for the construction of the Markov chain exists. Moreover, in all cases we considered, this optimal number of states is upper bounded by 10. Hence, with our method, more than 10 states do not help to increase the accuracy of the estimated delay statistics. It is clear that the independent error assumption provides poor results, even at high Doppler frequencies. The Gilbert Elliot model substantially improves the performance of the i.i.d. model by about one order of magnitude. However, especially for correlated channels ($f_d = 10$ Hz), the accuracy of the 4-state model is still worse than the precision achievable with $N = 6$.

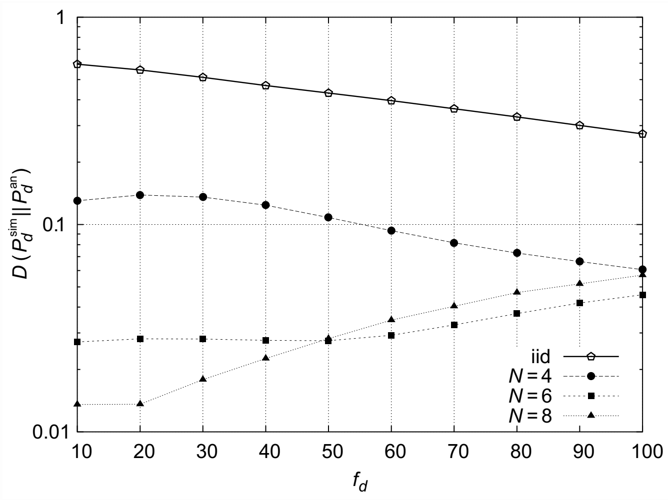


Fig. 11. Kullback Leibler distance between the distributions obtained by simulation ($P_d^{\text{sim}}[k]$) and analysis ($P_d^{\text{an}}[k]$) using our SNR partitioning approach as a function of the Doppler frequency f_d, for different values of N.

In Figs. 10 and 11, the distance metric is plotted as a function of f_d reporting both SNR partitioning methods considered in this paper, i.e., the equal probability method [10] and our novel improved proposal, respectively. Both approaches are compared also to the independent error case. Again, we can observe how the independent error assumption, where a single state is used to model the Rayleigh fading channel, fails to accurately model the underlying channel behavior. From Fig. 10, it is also evident that the equal probability method ($\theta_x = 1/K$) provides a rough approximation for the delivery delay statistics when $N \leq 6$. The problem, in such a case, is that the SNR range is not partitioned taking into account the shape of the PDU error probability function (Fig. 1). Our method, reported in Fig. 11, obtains better results for $N = 4, 6$. Already for $N = 8$, the two solutions are shown to be approximately equivalent. Moreover, the newly proposed method does not further improve that much; henceforth, it is confirmed that it is useful to apply it with a low N. From both figures, two main cases of system operating conditions can be highlighted, namely correlated (i.e., $f_d < 50$) and uncorrelated ($f_d \geq 50$), for which the system performs differently. In more detail, an increase in the number of the states of the Markov model is beneficial only in the former case, whereas it does not lead to substantial improvements in the latter for the uniform partitioning, and performs even worse for our method.

VII. Conclusions

In this work two main contributions are presented. First of all, an exact analysis to derive the delivery delay statistics of SR ARQ packets in an N-State Markov Model is presented. Secondly, this analysis is used to provide some results on the goodness of the Markov approximation of a Rayleigh fading channel in terms of delay statistics. The accuracy of the obtained statistics has been quantitatively evaluated and the impact of the number of states considered for the Markov channel modeling has been discussed.

The obtained results show that the statistics obtained using a Markov channel is reasonably close to the actual ones. However, the match between these distributions can not be made arbitrarily good by increasing the number of states due to inherent limitations of Markov channel modeling. Further, the independent error model has been confirmed to be a poor approximation for the delay statistics over a Rayleigh fading channel, even for large values of the Doppler frequency. Instead, the approximation introduced by the Markov modeling approach is satisfactory when the number of states is larger than 6 and the appropriate Signal-to-Noise Ratio partitioning method is selected. To this end, we discussed and evaluated existing strategies and proposed a new method, which is simpler than other techniques but is in general better capable of modeling the underlying channel, by achieving in some cases more satisfactory performance.

Finally, the delay analysis has been extended to the unreliable feedback case, by giving, also in this case, some useful insights on the impact of ACK/NACK errors.

References

[1] J. G. Kim and M. M. Krunz, "Delay analysis of selective repeat ARQ for a Markovian source over a wireless channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1968–1981, Sept. 2000.

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

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

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

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

[6] R. Fantacci, "Queueing analysis of the selective repeat automatic repeat request protocol for wireless packet networks," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 258–264, May 1996.

[7] M. Zorzi, R. Rao, and L. Milstein, "Error statistics in data transmission over fading channels," *IEEE Trans.Commun.*, vol. 46, no. 11, pp. 1468–1477, Nov. 1998.

[8] M. Rossi and M. Zorzi, "Analysis and heuristics for the characterization of selective repeat ARQ statistics over wireless channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1365–1377, Sept. 2003.

[9] M. Rossi, L. Badia, and M. Zorzi, "Exact statistics of ARQ packet delivery delay over Markov channels with finite round-trip delay," in *Proc. IEEE Globecom 2003*, vol. 6, pp. 3356–3360.

[10] H. S. Wang and N. Moayeri, "Finite-state Markov channel–a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[11] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[12] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2032–2040, Dec. 2000.

[13] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1987.

[14] M. Rossi, L. Badia, and M. Zorzi, "On the delay statistics of an aggregate of SR-ARQ packets over Markov channels with finite round-trip delay," in *Proc. IEEE WCNC 2003*, vol. 3, pp. 1773–1778.

[15] H. M. Chaskar, T. Lakshman, and U. Madhow, "TCP over wireless with link level error control: analysis and design methodology," *IEEE/ACM Trans. Networking*, vol. 7, no. 5, pp. 605–615, Oct. 1999.

[16] W. Turin, *Performance Analysis of Digital Transmission Systems*. Springer, 2004.

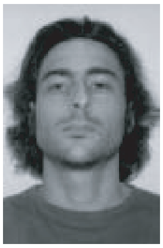
[17] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[18] P. Bergamo, D. Maniezzo, A. Giovanardi, G. Mazzini, and M. Zorzi, "An improved Markov model for Rayleigh fading envelope," *IEE Electron. Lett.*, vol. 38, no. 10, pp. 477–478, May 2002.

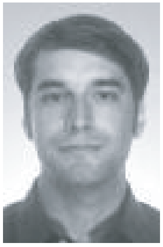
[19] ——, "An improved Markov chain description for fading processes," in *Proc. IEEE ICC 2002*, vol. 3, pp. 1347–1351.

[20] W. C. Jakes, *Microwave Mobile Communications*. Wiley–IEEE Press, May 1994.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.



Michele Rossi (S'02-M'04) was born in Ferrara, Italy on October 30th, 1974. He received the Laurea Degree in Electrical Engineering and the Ph.D. in Information Engineering (both with honors) from the University of Ferrara, Italy, in 2000 and 2004, respectively. Since 2000 he has been a Research Fellow at the Department of Engineering, University of Ferrara. During year 2003 he was on leave at the Center for Wireless Communications (CWC) at the University of California San Diego (UCSD), where he did research on wireless sensor networks. His research interests are on: TCP/IP protocols on wireless networks, TCP/IP header compression, performance analysis of Selective Repeat Link Layer retransmission techniques, efficient multicast data delivery in 3G cellular networks and integrated MAC/routing algorithms for wireless sensor networks.



Leonardo Badia (S'02-M'04) was born in Ferrara, Italy, in 1977. He received the Laurea Degree in Electrical Engineering and the Ph.D. in Information Engineering (both with honors) from the University of Ferrara, Italy, in 2000 and 2004, respectively. Since 2001 he has been with the Department of Engineering of the University of Ferrara. During 2002 and 2003 he was on leave at the Radio System Technology Labs (now Wireless@KTH), Royal Institute of Technology of Stockholm, Sweden. His research interests include energy efficient Ad Hoc Networks, transmission protocol modeling, Admission Control and Scheduling for wireless networks and economic modeling of Radio Resource Management.

Michele Zorzi (S'89–M'95–SM'98) was born in Venice, Italy, in 1966. He received the Laurea degree and the Ph.D. degree in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992/93, he was on leave at 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, Italy. After spending three years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, and in 2003 joined the Department of Information Engineering of the University of Padova, Italy, where he is currently a Professor. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, and energy constrained communications protocols.

Dr. Zorzi is the Editor-In-Chief of the *IEEE Wireless Communications Magazine*, and currently serves on the Editorial Boards of the *IEEE Transactions on Communications*, the *IEEE Transactions on Wireless Communications*, the *IEEE Transactions on Mobile Computing*, the *Wiley Journal of Wireless Communications and Mobile Computing* and the *ACM/URSI/Kluwer Journal of Wireless Networks*. He was 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* (Multi-media Network Radios).