Delay Analysis of Go-Back-N ARQ for Correlated Error Channels

Koen De Turck and Sabine Wittevrongel

SMACS* Research Group

Department of Telecommunications and Information Processing

Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

email: {kdeturck, sw}@telin.UGent.be

Abstract

We investigate the performance of the Go-Back-N ARQ (Automatic Repeat reQuest) protocol over a wireless channel. Data packets are sent from transmitter to receiver over the wireless transmission channel. When a packet is received, the receiver checks whether it has been received correctly or not, and sends a feedback message to notify the transmitter of the condition of that packet. When the transmitter is notified of a transmission error, the incorrectly received packet is sent again, as well as every following packet.

Our modeling assumptions are based on two convictions. On the one hand, a good view of the performance of an ARQ protocol not only requires an analysis of the throughput, but also of the buffer behavior. Therefore, we offer a complete queueing analysis of the transmitter buffer, in addition to a throughput analysis. Secondly, due to the highly variable nature of the error process in wireless networks, we have to take error correlation explicitly into account.

Hence, we model the channel by means of a general Markov chain with M states and a fixed error probability in every state. The transmitter buffer is modeled as a discrete-time queue with infinite storage capacity and independent and identically distributed packet arrivals from slot to slot.

We find concise expressions for the probability generating functions of the unfinished work and the packet delay of the transmitter buffer. Furthermore, we show explicit expressions for the mean and the variance of both system characteristics and we derive some heavy-load approximations. Finally, we provide some numerical examples.

Keywords: wireless telecommunication, queueing theory, performance evaluation, generating functions, automatic repeat request.

1 Introduction

A popular way of protecting against transmission errors is the so-called Automatic Repeat reQuest (ARQ). For this mechanism to work, the transmitter adds a simple error checking code to each packet so the receiver can detect the most common transmission errors. When a packet is received, the receiver checks whether it has been received correctly or not, and sends a feedback message to notify the transmitter of the condition of that packet. That is, a positive acknowledgement (ACK) is sent in case of a correct transmission; a negative acknowledgement (NAK) is sent in case of a transmission error. This organization requires a bi-directional channel between the sender and the receiver. Since not all arriving packets can be sent immediately and moreover the transmitter has to keep a copy of each packet until it is correctly transmitted, a buffer to store packets is required at the transmitter side.

Various types of ARQ protocols have been proposed in the literature [1]. They differ in the way the transmissions and retransmissions of packets are organized. In this paper, we focus on the Go-Back-N ARQ protocol (GBN-ARQ). Its operation is illustrated in Fig. 1. In case of GBN-ARQ, the

---

*SMACS: Stochastic Modeling and Analysis of Communication Systems

transmitter keeps on sending packets to the receiver without interruptions until a NAK is received. Upon reception of a NAK, the incorrectly received packet is sent again, as well as every following packet. We introduce the term feedback delay, which consists of (1) the time during which a packet is travelling through the channel, (2) the processing time of that packet in the receiver, and (3) the time during which the feedback message is travelling back.
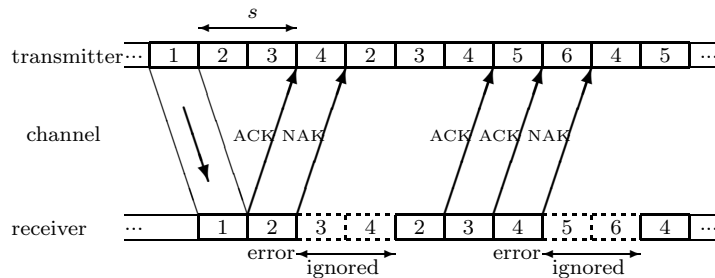

Figure 1: Operation of the Go-Back-N protocol (feedback delay $s = 2$).

The performance of the GBN-ARQ protocol has been investigated before. However, in many existing studies it is assumed that transmission errors occur independently, which leads to a considerably simpler analysis see e.g. [2]–[4]. Such an assumption is not realistic in case of non-stationary wireless transmission channels due to e.g. fading effects, see [5] for details. In particular, the time-varying nature of the channel will result in correlation in the occurrence of transmission errors. In this paper, we therefore model the channel by means of a Markov chain with M states and with a fixed error probability in every state. A special case of our model, where $M = 2$, is known as the Gilbert-Elliott model [6, 7]. Previous work on the behavior of GBN-ARQ for correlated errors is reported in [8], where the throughput and the transmitter buffer content are studied. In our paper, on the other hand, we present a full analysis of another important performance metric, namely the packet delay. Other related work is given in [5] where the performance of Stop-and-Wait ARQ over a dynamic two-state channel is investigated. The selective-repeat protocol with correlated errors is analyzed in [9] under the assumption of so-called ideal ARQ, where the feedback delay is neglected and acknowledgement messages are received instantly.

The paper is organized as follows. The modeling assumptions are given in Section 2. In Section 3, we introduce the *probability generating matrix* (pgm) of the service time, which plays a crucial role in the further analysis. Next, we derive the distribution of the unfinished work at an arbitrary slot boundary in Section 4, as a preparatory step for the analysis of the *probability generating function* (pgf) of the delay in Section 5. The actual computation of the moments of the packet delay is elaborated on in Section 6. We also derive some heavy-loa approximations there. We introduce a generalization of the popular Gilbert-Elliott channel model, which we dub the 'cyclic' channel model in Section 7. In Section 8, we provide numerical examples to show the soundness of our approximations and the impact of the error correlation. Finally, conclusions are drawn in Section 9.

2   Modeling Assumptions

We set out to analyze the behavior of the transmitter buffer for the GBN-ARQ protocol. Throughout the analysis, we assume that the data to be transmitted are divided into fixed-length packets. The time axis is likewise assumed to be divided into fixed-length slots, where a slot corresponds to the time needed to transmit one packet. Synchronous transmission is used, i.e., transmission always starts at the beginning of a slot. Packets are transmitted according to a first-come-first-served discipline. After a constant period of s slots, an acknowledgement message from the receiver indicating whether or not the packet was correctly received, arrives at the transmitter (see Fig. 1). This interval of s slots is referred to as the feedback delay of the channel. It is assumed that no errors occur in
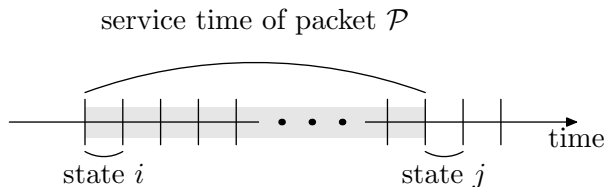
Figure 2: Service time starting in state i and followed by state j.

the acknowledgement messages. This is a reasonable assumption as the information content of an acknowledgement message is low and can thus be heavily protected. Moreover, if we do not make this assumption, then we cannot ensure completely reliable communication, but only minimize the risk, see the so-called Two Generals' Problem [10].

Let the random variable a_k denote the number of packets arriving at the transmitter during slot k. The a_ks are assumed to be independent and identically distributed (iid) variables with mean E[A], variance Var[A] and pgf $A(z)$. The transmitter buffer is assumed to have an infinite storage capacity and packets are stored in this buffer until they (and their predecessors) are successfully transmitted over the channel. Note that due to the synchronous transmission mode, a packet arriving in an empty buffer is not transmitted until the next slot.

The error process in the channel is modeled in this paper by means of a general Markov chain with M states. Let the random value c_k denote the channel state at slot boundary k (i.e. at the beginning of slot k), then the entries of the transition probability matrix $\mathbf{q}$ associated with the Markov chain are given by

$$q_{ij} \doteq \Pr[c_{k+1} = j | c_k = i]. \tag{1}$$

State i has an error probability of e_i. When there is an error, the packet sent during that slot will be incorrectly received, and the transmitter will receive a NAK message s slots later. We define the matrix $\mathbf{e}$ as the diagonal matrix with elements $e_1, \ldots, e_M$. We also introduce the notation $\bar{x} \doteq 1 - x$. Likewise, $\bar{\mathbf{e}}$ is a diagonal matrix with elements $\bar{e}_1, \ldots, \bar{e}_M$.

In the rest of this paper, we will often rely on a trick that was introduced by Towsley and Wolf [2]. The trick is to consider a slightly modified system in which a packet immediately leaves the transmitter buffer after being successfully transmitted, instead of staying another s slots until the positive acknowledgement is received. Note that the delay in the modified system of every packet is exactly s slots less than in the original system, which makes it easy to convert results for the delay obtained for the modified system into results for the original system, as will be elaborated on later.

3 Distribution of the Service Time

By putting the evolution of the service time into a convenient form, we set the stage for the queueing analysis proper. As we will see, the results in this section will greatly simplify the derivations in the subsequent sections.

The service time of a packet is defined as the time interval (expressed in slots) that starts at the slot boundary where the packet is transmitted for the first time and ends with the slot boundary where the packet leaves the system. We consider the modified system, so packets leave the buffer at the end of the slot where they are correctly transmitted. Remark that this also means that in the modified system, service times do not overlap. Service times in the real system are s slots longer, and may overlap.

In view of the above modeling assumptions, the exact distribution of the service time of a packet will depend on the channel state in the slot during which the service of the packet starts. Therefore, consecutive service times in the considered model are not iid, unlike in the case of uncorrelated transmission errors. In order to study the service time, let us introduce a pgm $\mathbf{S}(z)$ of dimension $M \times M$. The element $[\mathbf{S}(z)]_{ij}$ is the partial conditional pgf of the service time that is followed by a

slot with channel state j, given that the service time starts in channel state i (see Fig. 2), that is

$$[\mathbf{S}(z)]_{ij} = \mathrm{E}[z^S \ 1(c_{k'} = j)|c_k = i], \tag{2}$$

where the random variable S denotes a service time, k and k' denote the slot boundary where the service starts and ends respectively, and $1(.)$ is the indicator function of the event between brackets. Note that a service time has some kind of generalized geometric property: when there is no error, the service finishes after exactly one slot; when there is an error, a new transmission starts after $s + 1$ slots, and this can be seen as the start of a completely new service time, but this time starting in state c_{k+s+1}. Indeed, when there is an error, the total service lasts $s + 1$ slots plus a remaining service time. Given the state c_{k+s+1} at the start of this remaining service time, the mechanism that determines the length of the remaining service time is completely equivalent to the one that determines the length of a full service time and therefore this remaining service time also has pgm $\mathbf{S}(z)$, i.e. it has the pgm of a completely new service time. From these observations, we can see that

$$[\mathbf{S}(z)]_{ij} = z \Pr[\text{no error at } k, c_{k'} = j|c_k = i] \tag{3}$$

$$+ z^{s+1} \sum_{i'=1}^{M} \Pr[\text{error at } k, c_{k+s+1} = i'|c_k = i] \, \mathrm{E}[z^S 1(c_{k'} = j)|c_{k+s+1} = i'] \tag{4}$$

$$= z(1 - e_i)[\mathbf{q}]_{ij} + z^{s+1} \sum_{i'=1}^{M} e_i [\mathbf{q}^{s+1}]_{ii'} [\mathbf{S}(z)]_{i'j}, \tag{5}$$

where we have used the fact that the $(s + 1)$-step transition probabilities for the Markov chain of the channel are given by the transition matrix $\mathbf{q}^{s+1}$. Exploiting the matrix notation, we can write pgm $\mathbf{S}(z)$ succinctly as

$$\mathbf{S}(z) = \bar{\mathbf{e}}\mathbf{q}z + \mathbf{e}\mathbf{q}^{s+1}z^{s+1}\mathbf{S}(z). \tag{6}$$

From the above relation, we can derive $\mathbf{S}(z)$ as

$$\mathbf{S}(z) = (\mathbf{I} - \mathbf{e}\mathbf{q}^{s+1}z^{s+1})^{-1}\bar{\mathbf{e}}\mathbf{q}z, \tag{7}$$

where $\mathbf{I}$ denotes the $M \times M$ identity matrix.

The advantage of capturing the distribution of the service time in the pgm $\mathbf{S}(z)$ is that we can express the pgm of the length of n subsequent service times simply as $\mathbf{S}(z)^n$. Equation (7) also allows us to compute the throughput of the protocol. the throughput η is defined as the inverse of the mean service time under assumption that there are always packets available (heavy-traffic assumption). Notice that the matrix $\mathbf{S}(1)$ is the transition matrix that records the channel state transition between the start and the end of a service. The steady-state probability vector $\boldsymbol{\pi}$ is given by

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{S}(1) \quad \text{and} \quad \boldsymbol{\pi}\mathbf{1} = 1, \tag{8}$$

where $\mathbf{1}$ is an $N \times 1$ column vector of ones. The vector $\boldsymbol{\pi}$ records the probabilities of finding the channel in a particular state at the beginning of a service, if new services start uninterruptedly. Now, $[\mathbf{S}'(1)]_{ij}$ equals $\mathrm{E}[S \ 1(c_k' = j)|c_k = i]$, i.e. $[\mathbf{S}'(1)]_{ij}$ denotes the mean length of a service time that is followed by state j, given that it starts in state i. Therefore, we have that

$$\boldsymbol{\pi}\mathbf{S}'(1)\mathbf{1} = \sum_{i=1}^{M}\sum_{j=1}^{M}\Pr[c_k = i] \, \mathrm{E}[S \ 1(c_k' = j)|c_k = i]$$

$$= \mathrm{E}[S]. \tag{9}$$

That is, $\boldsymbol{\pi}\mathbf{S}'(1)\mathbf{1}$ is the expected length of a service time, given that there are no gaps between services. This is equal to the reciproque of the throughput η:

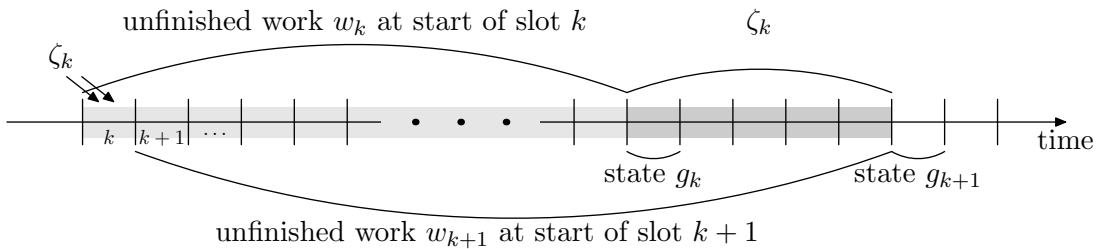$$\eta = \frac{1}{\boldsymbol{\pi}\mathbf{S}'(1)\mathbf{1}}. \tag{10}$$

Figure 3: Definitions of w_k, ζ_k and g_k.

4 Distribution of the Unfinished Work

In this section we derive an expression for the pgf of the unfinished work w_k at the beginning of an arbitrary slot k. This is the time in slots needed to serve the packets that are present in the buffer at the beginning of slot k. We again consider the modified system here.

Let ζ_k denote the amount of work (in slots) that enters the buffer during slot k, i.e. the sum of the service times of all packets arriving in slot k. Then the unfinished work evolves according to the following system equation:

$$w_{k+1} = (w_k - 1)^+ + \zeta_k, \tag{11}$$

where $(\cdots)^+ = \max(0, \cdots)$. The random variables w_k and ζ_k on the right-hand side of (11) are not independent, because both depend on the state of the channel. This means that the set $\{w_k\}$ does not form a Markov chain. We can make it a Markov chain by adding an extra variable. The most convenient choice is the variable g_k, which indicates the state of the channel in the earliest slot where work arriving during slot k could start. So, when $w_k = 0$, g_k corresponds to the channel state during slot $k + 1$. The case where $w_k > 0$ is illustrated in Fig. 3. We have that

$$g_{k+1} = \begin{cases} g_k^{(1)} & \text{when } w_k = \zeta_k = 0, \\ g_k^{(\zeta_k)} & \text{otherwise,} \end{cases} \tag{12}$$

where the notation $g_k^{(n)}$ stands for the state of the channel n slots after the slot with state g_k. From (11) and (12), it is easily seen that the set of random variables $\{w_k, g_k\}$ indeed forms a Markov chain.

Let us now define the partial pgfs

$$W_{j,k}(z) = \sum_{n=0}^{\infty} \Pr[w_k = n, g_k = j] z^n,$$

and introduce the vector notation $\mathbf{W}_k(z) = (W_{1,k}(z), \cdots, W_{M,k}(z))$. Then it is possible to rewrite the above system equations in the z-domain as:

$$\mathbf{W}_{k+1}(z) = \frac{1}{z}(\mathbf{W}_k(z) - \mathbf{W}_k(0))A(\mathbf{S}(z)) + \mathbf{W}_k(0)\left[A(\mathbf{S}(z)) - A(0)\mathbf{I} + A(0)\mathbf{q}\right].$$

Here we have introduced the convenient shorthand notation $A(\mathbf{S}(z))$ to denote a matrix that is a power series expansion in the pgm $\mathbf{S}(z)$ with the same coefficients as the power series expansion in z of $A(z)$. Similar notations will also be used later in this paper. Note that

$$\boldsymbol{\pi} A(\mathbf{S}(1)) = \boldsymbol{\pi} \sum_{k=0}^{\infty} \Pr[a = k]\mathbf{S}(1)^k = \boldsymbol{\pi}, \tag{13}$$

which shows that transition matrices $\mathbf{S}(1)$ and $A(\mathbf{S}(1))$ have the same steady-state probability vector. The system will reach an equilibrium if the average amount of work per slot (commonly called the load ρ) is strictly smaller than 1. The load ρ satisfies

$$\rho = \mathrm{E}[S]\,\mathrm{E}[A] = \boldsymbol{\pi} A'(1)\mathbf{S}'(1)\mathbf{1}. \tag{14}$$

5

In the steady state both $\mathbf{W}_k(z)$ and $\mathbf{W}_{k+1}(z)$ will converge to a common limiting vector $\mathbf{W}(z) = (W_1(z), \cdots, W_M(z))$. Taking limits for $k \to \infty$ and solving the resulting equation for $\mathbf{W}(z)$, we obtain

$$\mathbf{W}(z)[z\mathbf{I} - A(\mathbf{S}(z))] = \mathbf{W}(0)\left[(z-1)A(\mathbf{S}(z)) + zA(0)(\mathbf{q} - \mathbf{I})\right]. \tag{15}$$

Note that this formula still contains the unknown vector $\mathbf{W}(0)$, whose computation is the subject of a huge number of studies [11], [12]. The proposed algorithms fall apart into two broad categories. The first class are the so-called spectral methods, which are based on the property that there are a number of values z_i inside the unit circle, such that $z_i\mathbf{I} - A(\mathbf{S}(z_i))$ is a singular matrix, and such that the corresponding right null spaces $\mathbf{R}_i$ have together a total rank of $M - 1$ [11]. Then we can form a set of $M - 1$ equations

$$\mathbf{0} = \mathbf{W}(0)\left[(z_i - 1)A(\mathbf{S}(z_i)) + z_iA(0)(\mathbf{q} - \mathbf{I})\right]\mathbf{R}_i. \tag{16}$$

The last equation for $\mathbf{W}(0)$ is supplied by the well-known property of single-server queues that utilization equals the load. The utilization equals $1 - \mathbf{W}(0)\mathbf{1}$, while the load is given by Eq. (14), so that we complete the set of equations by stating that

$$\mathbf{W}(0)\mathbf{1} = 1 - \rho. \tag{17}$$

Some authors question the numerical stability of this method, especially if there are z_i which have a corresponding null space of a high dimension, and they give preference to the second method.

This other method is based on the weak canonical Wiener-Hopf factorization of the Laurent series $\mathbf{I} - \frac{A(\mathbf{S}(z))}{z}$, which says that this expression can be factorized under mild conditions into

$$\mathbf{I} - \frac{A(\mathbf{S}(z))}{z} = \mathbf{F}(z)(\mathbf{I} - \frac{1}{z}\mathbf{G}). \tag{18}$$

Here power series $\mathbf{F}(z)$ has no roots inside the unit circle, whereas the roots of expression $\mathbf{I} - \frac{1}{z}\mathbf{G}$ lie inside or on the unit circle. It can be shown [12] that matrix $\mathbf{G}$ is the well-known fundamental matrix as it occurs in M/G/1-type Markov chains, for which many algorithms have been developed. If we substitute the Wiener-Hopf factorization into Eq. (15), and identify the terms in z^{-1}, we find after some manipulations that

$$\mathbf{W}(0) = \mathbf{W}(0)(\mathbf{G} + A(0)(\mathbf{I} - \mathbf{q})). \tag{19}$$

Hence the computation of the boundary probabilities is reduced to a linear system of equations, which must again be supplemented by Eq. (17).

5   Distribution of the Packet Delay

Now that we have found an expression for the pgf of the unfinished work, it is rather straightforward to analyze the delay of an arbitrary packet $\mathcal{P}$ arriving in slot k. Let us first focus again on the modified system. We denote the channel state during the arrival slot of $\mathcal{P}$ as $s(\mathcal{P})$. The channel state in the first slot after the service of $\mathcal{P}$ is denoted by $r(\mathcal{P})$.

The delay $d(\mathcal{P})$ of packet $\mathcal{P}$ consists of two components, as illustrated in Fig. 4: (a) the remaining number of slots from slot $k + 1$ onwards needed to execute the unfinished work present in the system at the beginning of slot k, i.e., $(w_k - 1)^+$, and (b) the service times of the ℓ packets arriving in the buffer in slot k that will be served no later than (but including) $\mathcal{P}$. The distribution of the unfinished work has been derived in the previous section. Owing to the fact that $\mathcal{P}$ is an arbitrary packet, the pgf $L(z)$ of ℓ can be derived as (see e.g. [13]):

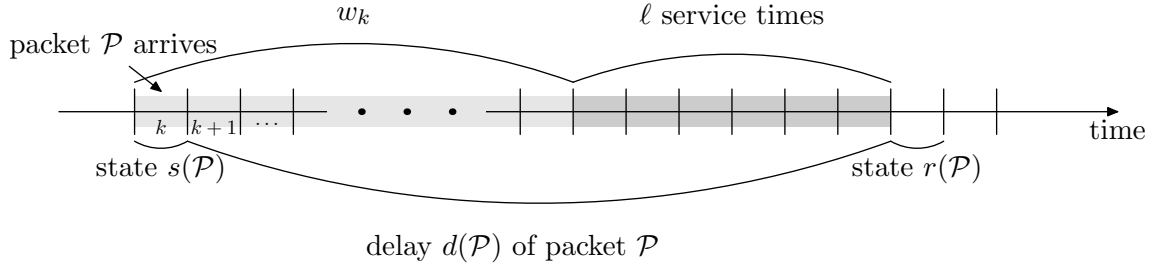$$L(z) = \frac{z(1 - A(z))}{\mathrm{E}[A](1 - z)}. \tag{20}$$

Figure 4: Illustration of the various components of the delay.

Now let $D_j(z)$ denote the partial pgf of the packet delay, provided that $r(\mathcal{P}) = j$. Then we can write, again using vector notation,

$$\mathbf{D}(z) = \frac{1}{z}(\mathbf{W}(z) + (z-1)\mathbf{W}(0))L(\mathbf{S}(z)). \tag{21}$$

The unconditional pgf $\tilde{D}(z)$ of the packet delay in the modified system is given by $\tilde{D}(z) = \mathbf{D}(z)\mathbf{1}$. From (21), it then follows that

$$z\tilde{D}(z) = (\mathbf{W}(z) + (z-1)\mathbf{W}(0))L(\mathbf{S}(z))\mathbf{1}. \tag{22}$$

From this, we can finally obtain the pgf of the delay $D(z)$ in the real system:

$$D(z) = z^s\tilde{D}(z). \tag{23}$$

6 Computing Moments of the Packet Delay

In this section, we explain in detail how to compute the mean and higher-order central moments of the packet delay. Once we have the boundary probabilities in the form of vector $\mathbf{W}(0)$, only a moderate amount of numerical computation is needed to get the first few moments of the delay. We illustrate this by computing the first two central moments, as higher-order moments are obtained by similar means.

Notice that the mean and the variance of the delay in the real system can be trivially obtained in function of the same quantities of the modified system:

$$\mathrm{E}[D] = \mathrm{E}[\tilde{D}] + s \quad \text{and} \quad \mathrm{Var}[D] = \mathrm{Var}[\tilde{D}]. \tag{24}$$

Let us now compute the first two derivatives of $\tilde{D}(z)$ as given in Eq. (22), and evaluate them for $z = 1$:

$$\tilde{D}'(1) + 1 = (\mathbf{W}'(1) + \mathbf{W}(0))\mathbf{1} + \mathbf{W}(1)\boldsymbol{\ell}'(1); \tag{25}$$
$$\tilde{D}''(1) + 2\tilde{D}'(1) = \mathbf{W}''(1)\mathbf{1} + 2(\mathbf{W}'(1) + \mathbf{W}(0))\boldsymbol{\ell}'(1) + \mathbf{W}(1)\boldsymbol{\ell}''(1), \tag{26}$$

where we introduced the column vector $\boldsymbol{\ell}(z) = L(\mathbf{S}(z))\mathbf{1}$. Note that $\boldsymbol{\ell}(1) = \mathbf{1}$. In view of the fact that $\mathrm{Var}[D] = \tilde{D}''(1) + \tilde{D}'(1) - \tilde{D}'(1)^2$, we find after some further manipulations that

$$\mathrm{E}[D] = \mathrm{E}[W] - \rho + \mathbf{W}(1)\boldsymbol{\ell}'(1) + s; \tag{27}$$
$$\mathrm{Var}[D] = \mathrm{Var}[W] + 2(\mathbf{W}'(1) + \mathbf{W}(0))\boldsymbol{\ell}'(1) + \mathbf{W}(1)\boldsymbol{\ell}''(1)$$
$$\qquad - (1 - \rho + \mathbf{W}(1)\boldsymbol{\ell}'(1))(2\,\mathrm{E}[W] - \rho + \mathbf{W}(1)\boldsymbol{\ell}'(1)), \tag{28}$$

7

where we introduced $\mathrm{E}[W] = \mathbf{W}'(1)\mathbf{1}$ and $\mathrm{Var}[W] = \mathbf{W}''(1)\mathbf{1} + \mathbf{W}'(1)\mathbf{1} - (\mathbf{W}'(1)\mathbf{1})^2$. So far, we have reduced the problem of finding the moments of the delay into the problem of finding the row vectors $\mathbf{W}(1), \mathbf{W}'(1)$ and $\mathbf{W}''(1)$ associated with the unfinished work, and the column vectors $\boldsymbol{\ell}'(1)$ and $\boldsymbol{\ell}''(1)$, which are connected to the amount of work that arrives during an arrival slot.

Let us first focus our attention on the unknowns $\mathbf{W}(1), \mathbf{W}'(1)$ and $\mathbf{W}''(1)$. When we evaluate Eq. (15) for $z = 1$ and do the same for its first two derivatives in z, we obtain

$$\mathbf{W}(1)(\mathbf{I} - \mathbf{A}(1)) = \mathbf{W}(0)A(0)(\mathbf{q} - \mathbf{I}); \tag{29}$$

$$\mathbf{W}'(1)(\mathbf{I} - \mathbf{A}(1)) + \mathbf{W}(1)(\mathbf{I} - \mathbf{A}'(1)) = \mathbf{W}(0)(\mathbf{A}(1) + A(0)(\mathbf{q} - \mathbf{I})); \tag{30}$$

$$\mathbf{W}''(1)(\mathbf{I} - \mathbf{A}(1)) + 2\mathbf{W}'(1)(\mathbf{I} - \mathbf{A}'(1)) - \mathbf{W}(1)\mathbf{A}''(1) = 2\mathbf{W}(0)\mathbf{A}'(1), \tag{31}$$

where we have introduced $\mathbf{A}(z) = A(\mathbf{S}(z))$ for reasons of convenience. We see that $\mathbf{W}(1), \mathbf{W}'(1)$ and generally $\mathbf{W}^{(n)}(1)$ satisfy a linear equation of the form $\mathbf{x}(\mathbf{I} - \mathbf{A}(1)) = \mathbf{b}$. However, as $\mathbf{A}(1)$ is a stochastic matrix, the system of equations is singular, and hence we need an extra equation for each $\mathbf{W}^{(n)}(1)$. Note that the solution of an inhomogeneous system of equations can generally be written as the sum of one particular solution of the inhomogeneous system, plus a solution of the homogeneous system, which in this case has the form $c\boldsymbol{\pi}$, for an arbitrary constant c. We define the particular solution of the problem with the help of the group inverse [14], which is the subject of the Appendix. The useful property of the group inverse for our purposes is that the group inverse $(\mathbf{I} - \mathbf{A}(1))^{\#}$ gives us the solution for $\mathbf{x}$ for the equation $\mathbf{x}(\mathbf{I} - \mathbf{A}(1)) = \mathbf{b}$, such that $\mathbf{x}\mathbf{1} = 0$. Let $\tilde{\mathbf{W}}(1), \tilde{\mathbf{W}}'(1)$ and $\tilde{\mathbf{W}}''(1)$ denote the particular solutions of (29)–(31) found by means of the group inverse, i.e.

$$\tilde{\mathbf{W}}(1) = \mathbf{W}(0)A(0)(\mathbf{q} - \mathbf{I})(\mathbf{I} - \mathbf{A}(1))^{\#}; \tag{32}$$

$$\tilde{\mathbf{W}}'(1) = (\mathbf{W}(0)(\mathbf{A}(1) + A(0)(\mathbf{q} - \mathbf{I})) - \mathbf{W}(1)(\mathbf{I} - \mathbf{A}'(1)))(\mathbf{I} - \mathbf{A}(1))^{\#}; \tag{33}$$

$$\tilde{\mathbf{W}}''(1) = (2\mathbf{W}(0)\mathbf{A}'(1) - 2\mathbf{W}'(1)(\mathbf{I} - \mathbf{A}'(1)) + \mathbf{W}(1)\mathbf{A}''(1))(\mathbf{I} - \mathbf{A}(1))^{\#}. \tag{34}$$

Then we can write

$$\mathbf{W}(1) = \tilde{\mathbf{W}}(1) + \mathbf{W}(1)\mathbf{1}\boldsymbol{\pi}; \tag{35}$$

$$\mathbf{W}'(1) = \tilde{\mathbf{W}}'(1) + \mathbf{W}'(1)\mathbf{1}\boldsymbol{\pi}; \tag{36}$$

$$\mathbf{W}''(1) = \tilde{\mathbf{W}}''(1) + \mathbf{W}''(1)\mathbf{1}\boldsymbol{\pi}, \tag{37}$$

where we still have to determine the scalars $\mathbf{W}(1)\mathbf{1}$, $\mathbf{W}'(1)\mathbf{1}$ and $\mathbf{W}''(1)\mathbf{1}$. Note that the normalization condition learns us that $\mathbf{W}(1)\mathbf{1} = 1$. Furthermore, when we substitute Eq. (36) into Eq. (31), and postmultiply by the column vector $\mathbf{1}$, we find

$$2(\tilde{\mathbf{W}}'(1) + \mathbf{W}'(1)\mathbf{1}\boldsymbol{\pi})(\mathbf{I} - \mathbf{A}'(1))\mathbf{1} - \mathbf{W}(1)\mathbf{A}''(1)\mathbf{1} = 2\mathbf{W}(0)\mathbf{A}'(1)\mathbf{1}. \tag{38}$$

Since $\boldsymbol{\pi}(\mathbf{I} - \mathbf{A}'(1))\mathbf{1} = 1 - \rho$, this leads to

$$2(1 - \rho)\mathbf{W}'(1)\mathbf{1} = 2\mathbf{W}(0)\mathbf{A}'(1)\mathbf{1} + \mathbf{W}(1)\mathbf{A}''(1)\mathbf{1} - 2\tilde{\mathbf{W}}'(1)(\mathbf{I} - \mathbf{A}'(1))\mathbf{1}. \tag{39}$$

Similarly, from the evaluation of the third derivative of Eq. (15) for $z = 1$ we can infer the value for $\mathbf{W}''(1)\mathbf{1}$.

Having expounded on the computation of the derivatives of $\mathbf{W}(z)$ in $z = 1$, we now set up a similar scheme for the computation of $\boldsymbol{\ell}'(1)$ and $\boldsymbol{\ell}''(1)$. Note that

$$\mathrm{E}[A](\mathbf{S}(z) - \mathbf{I})\boldsymbol{\ell}(z) = \mathbf{S}(z)(\mathbf{A}(z) - \mathbf{I})\mathbf{1}. \tag{40}$$

The first two derivations of this expression with respect to z, evaluated in $z = 1$, are equal to

$$\mathrm{E}[A](\mathbf{S}'(1)\mathbf{1} + (\mathbf{S}(1) - \mathbf{I})\boldsymbol{\ell}'(1)) = \mathbf{S}(1)\mathbf{A}'(1)\mathbf{1}; \tag{41}$$

$$\mathrm{E}[A](\mathbf{S}''(1)\mathbf{1} + 2\mathbf{S}'(1)\boldsymbol{\ell}'(1) + (\mathbf{S}(1) - \mathbf{I})\boldsymbol{\ell}''(1)) = 2\mathbf{S}'(1)\mathbf{A}'(1)\mathbf{1} + \mathbf{S}(1)\mathbf{A}''(1)\mathbf{1}. \tag{42}$$

Employing again the technique of the group inverse, we establish that

$$\boldsymbol{\ell}'(1) = \tilde{\boldsymbol{\ell}}'(1) + \mathbf{1}\boldsymbol{\pi}\boldsymbol{\ell}'(1), \tag{43}$$

where

$$\tilde{\boldsymbol{\ell}}'(1) = \frac{1}{\mathrm{E}[A]}(\mathbf{I} - \mathbf{S}(1))^{\#}(\mathrm{E}[A]\mathbf{S}'(1)\mathbf{1} - \mathbf{S}(1)\mathbf{A}'(1)\mathbf{1}). \tag{44}$$

We find the scalar $\boldsymbol{\pi}\boldsymbol{\ell}'(1)$ by plugging the previous equation into Eq. (42), and premultiplying by $\boldsymbol{\pi}$:

$$2\rho\boldsymbol{\pi}\boldsymbol{\ell}'(1) = 2\boldsymbol{\pi}\mathbf{S}'(1)\mathbf{A}'(1)\mathbf{1} + \boldsymbol{\pi}\mathbf{A}''(1)\mathbf{1} - \mathrm{E}[A]\boldsymbol{\pi}\mathbf{S}''(1)\mathbf{1} - 2\,\mathrm{E}[A]\boldsymbol{\pi}\mathbf{S}'(1)\tilde{\boldsymbol{\ell}}'(1). \tag{45}$$

We omit the entirely analogous derivation of $\boldsymbol{\ell}''(1)$.

The computational complexity of our method compares favorably to alternatives. After we have computed 2 group inverses of matrices of order $M \times M$, we only have to compute matrix-vector multiplications and vector additions, which have complexities $O(M^2)$ and $O(M)$ respectively. In practice, the real cost is the computation of the boundary probability vector $\mathbf{W}(0)$, which is almost always perforce an iterative procedure. We therefore elaborate on an approximative method for heavy loads (i.e. $\rho \to 1$) which permits to skip the computation of $\mathbf{W}(0)$ altogether.

The idea behind this approximation is that as the load approaches one, $\mathbf{W}(0)$ becomes almost equal to the zero vector, so that the terms in which $\mathbf{W}(0)$ occurs, vanish for loads close to one. Let us mark the system characteristics as they hold for heavy-load conditions by a subscript h, e.g. $\mathrm{E}[D_h], \mathbf{W}_h(1), \mathbf{W}_h'(1), \cdots$. First note that the vector $\mathbf{W}_h(1)$ in that case approaches the vector $\boldsymbol{\pi}$. This can also be intuited from the fact that for heavy loads, empty periods get scarce, and hence services follow each other without gaps, which is exactly the definition of the steady-state vector $\boldsymbol{\pi}$. Similar simplifications in Eq. (39) lead to

$$\mathrm{E}[W_h] = \frac{1}{2(1-\rho)}(\boldsymbol{\pi}\mathbf{A}''(1)\mathbf{1} + 2\boldsymbol{\pi}\mathbf{A}'(1)(\mathbf{I} - \mathbf{A}(1))^{\#}\mathbf{A}'(1)\mathbf{1}). \tag{46}$$

The extra terms in the expression (27) for the mean delay $\mathrm{E}[D]$ can be approximated as well. This leads to the following heavy-load limit for the mean packet delay:

$$\mathrm{E}[D_h] = \mathrm{E}[W_h] - \rho + \boldsymbol{\pi}\boldsymbol{\ell}'(1) + s, \tag{47}$$

where $\boldsymbol{\pi}\boldsymbol{\ell}'(1)$ follows from 45. From the numerical results, it is observed that $\mathrm{E}[D_h]$ is dominated by the contribution of $\mathrm{E}[W_h]$. Note that $\mathrm{E}[W_h]$ quantifies the delay incurred by the buffer, whereas the other terms quantify the delay incurred by other packets arriving during the same slot, and the delay incurred by the feedback channel. It is clear that the delay due to buffering will dwarf the other terms in case of high loads. Note that the computational complexity of the heavy-load mean delay is not so much greater than the complexity of the throughput, yet it offers a lot of additional insight in the performance.

7   Channel Models

A vast amount of papers have been written on the modeling of communication channels in general [6, 7] and wireless communication channels in particular [15, 16, 17]. Many of the proposed models fit into the discrete Markovian framework we have used in this paper.

It is intuitively clear that a larger number of background states can more faithfully model the behavior of real communication channels. However, there are also disadvantages. A model with a large number of background states needs in general a large number of parameters (order M^2). If these have to be estimated from observations, one needs a lot of data to reliably estimate all these parameters. Moreover, it gets difficult to assign an interpretation to the parameters, which is a disadvantage for numerical examples.
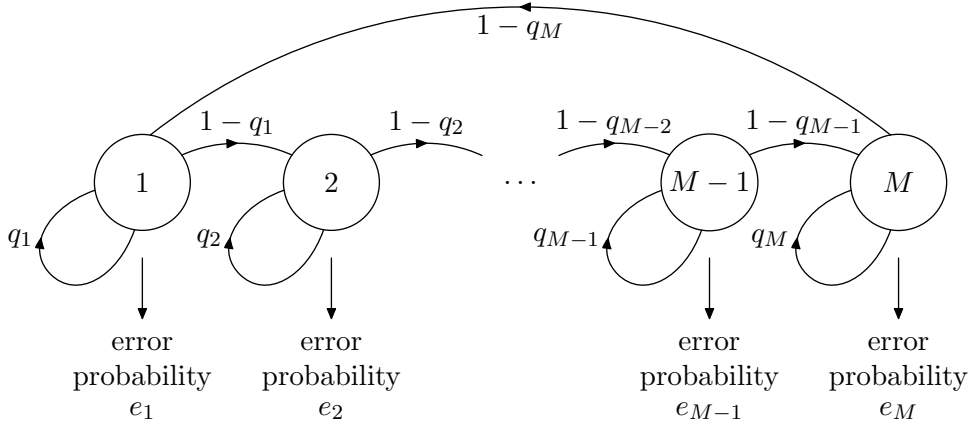
Figure 5: The Markov chain of a cyclic channel model

7.1 Cyclic Channel Model

We propose a channel model that is very versatile yet simple enough to be reasoned about: we allow only cyclic transitions of the channel state, and the sojourn time in each state i, $1 \leq i \leq M$ is (shifted) geometrically distributed with parameter q_i. This chain is illustrated in Fig. 5. We have the following transition matrix:

$$
\mathbf{q} = \begin{pmatrix}
q_1 & \bar{q}_1 & & & \\
& q_2 & \bar{q}_2 & & \\
& & \ddots & \ddots & \\
& & & q_{M-1} & \bar{q}_{M-1} \\
\bar{q}_M & & & & q_M
\end{pmatrix}.
\tag{48}
$$

State i has error probability e_i. Hence, $2M$ parameters define the channel model (M error probabilities and M transition probabilities). As the mean sojourn time in state i equals $\bar{q}_i^{-1}$, the mean length of an entire cycle (a visit to all states 1 to M, in order) equals

$$
L_c = \sum_{j=1}^{M} \frac{1}{1 - q_j}.
\tag{49}
$$

This is an important characteristic of the burstiness of the channel: the higher the mean cycle length, the burstier the channel. The steady-state probabilities σ_i of the channel states are easily obtained as

$$
\sigma_i = \frac{\bar{q}_i^{-1}}{L_c}.
\tag{50}
$$

Instead of using the transition probabilities to describe the channel model, we can also use the steady-state probability vector and the mean cycle length, which may give a somewhat more intuitive description. However, not all cycle lengths are possible for a given probability vector. Considering that $\bar{q}_i \leq 1$, for all i, $1 \leq i \leq M$, we see from Eq. (50) that $L_c \geq \sigma_i^{-1}$ for all i, $1 \leq i \leq M$. Hence the minimal cycle length $L_c^{\min}$ equals

$$
L_c^{\min} = \frac{1}{\min_i \sigma_i}.
\tag{51}
$$

In the numerical examples, we characterize a cyclic channel model by the cycle length L_c, and two vectors $\boldsymbol{\sigma}$ and $\mathbf{e}_c$, which record respectively the steady-state probabilities σ_i and the error probabilities e_i.
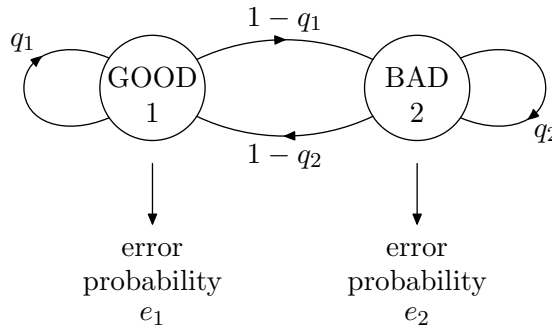
10

Figure 6: The Gilbert-Elliott channel model

7.2 Gilbert-Elliott Channel Model

For $M = 2$, the cyclic channel model reduces to a Gilbert-Elliott channel model [6, 7], which is the staple model in wireless communication modelling.

The two states of the Gilbert-Elliott model are usually labelled 1 and 2, or 'GOOD' and 'BAD' (see Fig. 6). The parameters e_1 and e_2 are the error probabilities of the channel in resp. state 0 and 1. Of course, the designations GOOD and BAD make only sense when $e_1 < e_2$, but this is not a requirement for the analysis.

The transitions in the Gilbert-Elliott model are completely defined by the parameters q_1 and q_2, where

$$q_i = \Pr[\text{state in next slot is } i | \text{ state in current slot is } i].$$

Rather than using q_1 and q_2, we define the parameters

$$\sigma = \frac{1 - q_1}{2 - q_1 - q_2} \text{ and } K = \frac{1}{2 - q_1 - q_2} \tag{52}$$

to be understood as follows: σ is the fraction of the time that the system is in state 2, while the parameter K can be seen as a measure for the mean lengths of 1- and 2-periods. Specifically, the mean length of a 2-period is K/σ, of a 1-period it is $K/\bar{\sigma}$. Therefore, the factor K can be seen as a measure for the *absolute* lengths of the 1-periods and 2-periods, while σ characterizes their *relative* lengths. The parameter K thus characterizes the degree of correlation in the channel state and is therefore referred to further on as the correlation factor, the value $K = 1$ corresponds to an uncorrelated channel state from slot to slot.

Note that in this model, the sojourn times in both states are geometrically distributed, while some measurements have indicated [17] that even though good periods can be modelled more or less faithfully with geometric lengths, bad periods cannot. A cyclic channel with more than one state designated to model the bad periods, can offer a solution to this problem.

8 Numerical results and discussion

In this section, we provide some numerical examples. First, we show some probability mass functions of the delay in Figs. 7 and 8, for a Gilbert-Elliott channel model and geometrically distributed batch arrivals with mean $E[A] = 0.3$. The figures show the influence of the correlation factor K, and the feedback delay s respectively. These curves have been obtained from the pgf $D(z)$ by means of an inversion method explained in [18]. We see that the packet delay over a more correlated channel has a heavier tail, which shows that correlation of the channel has an important influence on the performance of the system. A larger feedback delay also has a negative effect on the delay performance: we see a heavier tail for larger values of s.
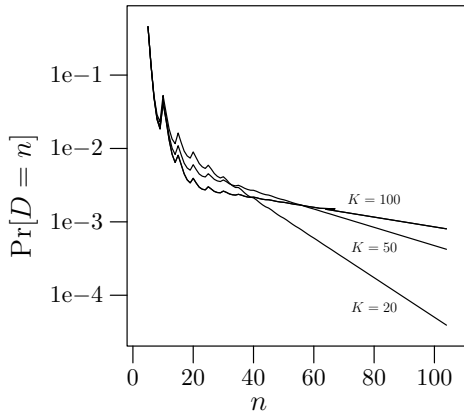
11

Figure 7: Logarithmic plot of $\mathrm{Prob}[D = n]$ for $s = 4, \sigma = 0.2, e_1 = 0.05, e_2 = 0.5$ and $K = 20, 50, 100$.
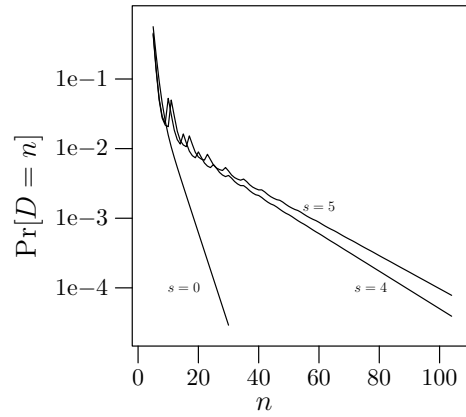

Figure 8: Logarithmic plot of $\mathrm{Prob}[D = n]$ for $\sigma = 0.2, e_1 = 0.05, e_2 = 0.5, K = 20$ and $s = 0, 4, 5$.
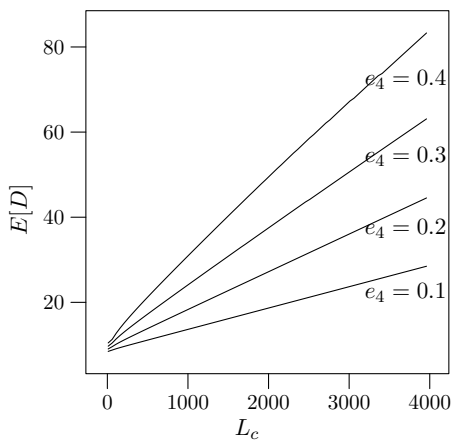

Figure 9: Mean delay versus cycle length L_c, with Poisson arrivals ($\lambda = \frac{1}{4}$), $s = 6$, $\boldsymbol{\sigma} = (0.4, 0.2, 0.2, 0.2)$ and $\mathbf{e}_c = (0.0, 0.1, 0.1, e_4)$, for different values of e_4.
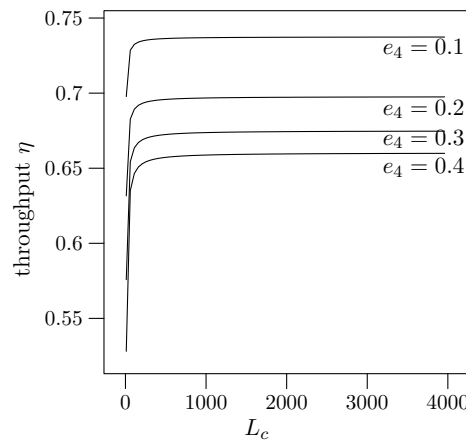

Figure 10: Throughput versus cycle length L_c, with Poisson arrivals ($\lambda = \frac{1}{4}$), $s = 6$, $\boldsymbol{\sigma} = (0.40.20.20.2)$ and $\mathbf{e}_c = (0.0, 0.1, 0.1, e_4)$, for different values of e_4.

   The huge impact of the correlation is also observed in Fig. 9, where we plot the mean delay $\mathrm{E}[D]$ against the cycle length L_c. We used a Poisson arrival process, with pgf $(A(z) = \exp(\lambda(z - 1)))$, with $\lambda = \frac{1}{4}$, feedback delay $s = 6$, and a four-state cyclic channel model with steady-state probability vector $\boldsymbol{\sigma} = (0.4, 0.2, 0.2, 0.2)$, and fixed error probabilities for the first three states (respectively $e_1 = 0, e_2 = 0.1, e_3 = 0.1$). We let the error probability e_4 take on different values, as indicated in the plot. We see an almost linear increase of the mean delay in function of the cycle length. Hence bursty channels deteriorate the performance of the channel tremendously, although the throughput is actually an increasing function for the very same parameters (see Fig. 10). This makes a strong case for the fact that the throughput is not sufficient to evaluate the performance of a retransmission protocol.

   Next, we show that bad periods with non-geometric lengths have indeed a different performance than bad periods with geometric lengths in Fig. 11. We compare channel model A with parameters $\mathbf{e}_c = (0.0, 0.3)$, and $\boldsymbol{\sigma} = (0.2, 0.8)$ with channel model B with parameters $\mathbf{e}_c = (0.0, 0.3 \cdots 0.3)$ and $\boldsymbol{\sigma} = (0.2, 0.1, \cdots, 0.1)$. That is, we compare a Gilbert-Elliott model where the bad period length is geometrically distributed, with a model where the bad period consists of 8 consecutive geometrically distributed subperiods. Note that the error probability in a bad slot in both cases equals 0.3, and moreover, the average length of a bad period is equal to $0.8L_c$ in both cases. In the second model however, the length of a bad period has a significantly lower variance, and (hence) shows a markedly
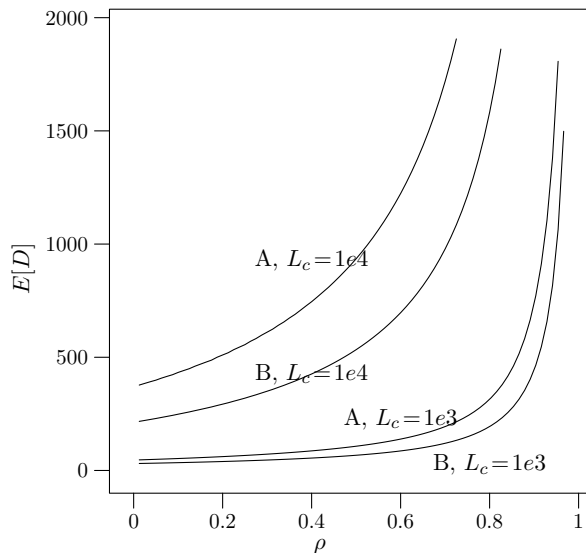
Figure 11: A plot of the mean delay versus the load, with feedback delay $s = 6$ for a two-state channel model A ($\mathbf{e}_c = (0.0, 0.3)$, and $\boldsymbol{\sigma} = (0.2, 0.8)$) and a nine-state channel model B ($\mathbf{e}_c = (0.0, 0.3, \cdots, 0.3)$ and $\boldsymbol{\sigma} = (0.2, 0.1, \cdots, 0.1)$). The load is varied by varying parameter λ of the Poisson arrival process.

better performance than the Gilbert-Elliott model. The difference is particularly large for more correlated channels (longer L_c). This figure shows that more complicated models may indeed be necessary in order to faithfully capture the performance of a protocol over a wireless link.

Finally, we demonstrate the soundness of our heavy-load approximation in Fig. 12. The computationally inexpensive approximations prove to be accurate for loads higher than 0.7 and sometimes even well before that point.

9 Conclusions

We have studied the transmitter buffer behavior for a Go-Back-N ARQ protocol over a time-varying channel. In particular, we have analyzed and found expressions for the steady-state distributions of the unfinished work and of the packet delay, as well as some very efficient heavy-load approximations. Finally, by means of some examples we have discussed the influence of the model parameters on the delay performance.

Appendix: Group Inverse of a Matrix

In this appendix, we elaborate on the technique of the group inverse, which constitutes an alternative for spectral decomposition, although some results [19] hint at a close relationship between the two techniques.

The group inverse of a matrix $\mathbf{M}$ is the unique matrix $\mathbf{M}^{\#}$ which satisfies $\mathbf{MM}^{\#}\mathbf{M} = \mathbf{M}$, $\mathbf{M}^{\#}\mathbf{MM}^{\#} = \mathbf{M}^{\#}$, and $\mathbf{MM}^{\#} = \mathbf{M}^{\#}\mathbf{M}$. We refer to [14] for more details and recall here only that for a transition matrix $\mathbf{A}(1)$ of an irreducible Markov chain the group inverse of $\mathbf{I} - \mathbf{A}(1)$ equals

$$(\mathbf{I} - \mathbf{A}(1))^{\#} = (\mathbf{I} - \mathbf{A}(1) + \mathbf{1}\boldsymbol{\pi})^{-1} - \mathbf{1}\boldsymbol{\pi}, \tag{52}$$

where $\boldsymbol{\pi}$ is the stationary probability vector of $\mathbf{A}(1)$. It can readily be seen that $(\mathbf{I} - \mathbf{A}(1))^{\#}\mathbf{1} = \mathbf{0}$ and $\boldsymbol{\pi}(\mathbf{I} - \mathbf{A}(1))^{\#} = \mathbf{0}$. In general, the solution of $\mathbf{x}(\mathbf{I} - \mathbf{A}(1))\mathbf{b}$ is given by $\mathbf{x} = \mathbf{b}(\mathbf{I} - \mathbf{A}(1))^{\#} + \mathbf{x}\boldsymbol{\pi}$. Note that the group inverse does not necessarily have to be computed via Eq. (52). Indeed, an alternative algorithm can be found in [20], where one does not have to compute $\boldsymbol{\pi}$ first, but rather $\boldsymbol{\pi}$ is obtained as a simple by-product of the group-inverse computation.
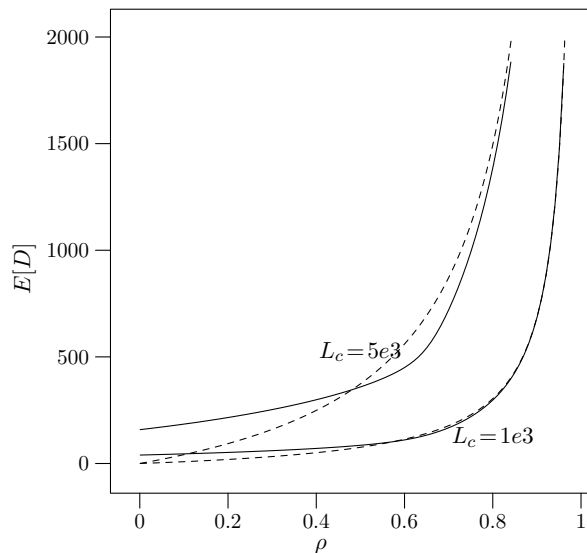
Figure 12: A plot of the mean delay versus the load, with the heavy-load approximation indicated with dashed lines, for a four state model, with $\boldsymbol{\sigma} = (0.4, 0.3, 0.2, 0.1)$ and $\mathbf{e}_c = (0.0, 0.3, 0.3, 0.3)$ for $s = 6$ and $L_c = 1000, 5000$. The load is varied by varying parameter λ of the Poisson arrival process.

It is beyond the scope of this article to show the similarities of and differences between group-inverse and spectral-decomposition techniques. An advantage of the group inverse is that it needs fewer assumptions. It is enough to assume that the matrix $\mathbf{A}(1)$ has a simple eigenvalue in 1 (ergodicity of the corresponding Markov chain is a sufficient condition for that) to ensure that the group inverse is given by Eq. (52), whereas spectral decomposition techniques often need additional assumptions, such as the assumption that $\mathbf{A}(1)$ be diagonalizable. Another tricky aspect that the use of the group inverse manages to circumvent is the normalization of the eigenvectors.

References

[1] Bhunia, C.T.: ARQ – Review and modifications. IETE Technical Review **18** (2001) 381–401

[2] Towsley, D., Wolf, J.K.: On the statistical analysis of queue lengths and waiting times for statistical multiplexers with ARQ retransmission schemes. IEEE Transactions on Communications **25** (1979) 693–703

[3] Konheim, A.G.: A queueing analysis of two ARQ protocols. IEEE Transactions on Communications **28** (1980) 1004–1014

[4] De Munnynck, M., Wittevrongel, S., Lootens, A., Bruneel, H.: Queueing analysis of some continuous ARQ strategies with repeated transmissions. Electronics Letters **38** (2002) 1295 – 1297

[5] De Vuyst, S., Wittevrongel, S., Bruneel, H.: Queueing delay of Stop-and-Wait ARQ over a wireless Markovian channel. Heterogeneous Networks, Vol. III, River Publishers, 2008.

[6] Gilbert, E.N.: Capacity of a burst-noise channel. The Bell System Technical Journal **39** (1960) 1253–1265

[7] Elliott, E.O.: Estimates of error rate for codes on burst-noise channels. Bell System Technical Journal **42** (1963) 1977–1997.

[8] Towsley, D.: A statistical analysis of ARQ protocols operating in a non-independent error environment. IEEE Transactions on Communications **27** (1981) 971–981

[9] Kim, J.G., Krunz, M.: Delay analysis of selective repeat ARQ for transporting Markovian sources over a wireless channel. IEEE Transactions on Vehicular Technology **49** (2000) 1968–1981

[10] E. A. Akkoyunlu , K. Ekanadham , R. V. Huber, Some constraints and tradeoffs in the design of network communications. Proceedings of the fifth ACM symposium on Operating systems principles, p.67-74, November 19-21, 1975, Austin, Texas, United States

[11] Gail, H. R., Hantler, S. L., Taylor, B. A.: Spectral analysis of $M/G/1$ and $G/M/1$ type Markov chains. Adv. Appl. Prob. **28** (1996) 114–165

[12] Bini, D. A., Latouche, G., Meini, B.: Numerical methods for structured Markov chains. Oxford University Press (2005)

[13] Bruneel, H.: Buffers with stochastic output interruptions. Electronics Letters **19** (1983) 735–737

[14] Campbell, S.L., Meyer, C. D.: Generalized Inverses of Linear Transformations. Dover Publications (1991)

[15] Kim, Y.Y., Li, S.Q.: Capturing important statistics of a fading/shadowing channel for network performance analysis. IEEE Journal on Selected Areas in Communications, **17**, (1999) 888–901

[16] Fading channel modeling via variable-length Markov chain technique. IEEE Transactions on Vehicular Technology **57** (2008) 1338–1358

[17] McDougall, J., Miller, S.: Sensitivity of wireless network simulations to a two-state Markov model channel approximation. Proceedings of GLOBECOM 2003, (1-5 December 2003, San Francisco, USA)

[18] Abate, J., Whitt, W.: Numerical inversion of probability generating functions. Operations Research Letters **12** (1992) 245–251

[19] Meyer, C. D., Stewart G. W.: Derivatives and Perturbations of Eigenvectors. SIAM Journal on Numerical Analysis, **25** (1988) 679–691

[20] Meyer, C. D.: The role of the group generalized inverse in the theory of finite Markov chains. SIAM Review, **17** (1975) 443–464