Mean value and tail distribution of the message delay in statistical multiplexers with correlated train arrivals

S. De Vuyst*, S. Wittevrongel, H. Bruneel
SMACS (Stochastic Modeling and Analysis of Communication Systems) Research Group, Vakgroep Telecommunicatie en Informatieverwerking, Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

Abstract

In this paper, we study a statistical multiplexer which is modelled as a discrete-time single-server infinite-capacity queueing system. This multiplexer is fed by messages generated by an unbounded population of users. Each message consists of a generally distributed number of fixed-length packets.

We assume the packet arrival process to exhibit simultaneously the following two types of correlation. First, the messages arrive to the multiplexer at the rate of one packet per slot, which results in what we call a primary correlation in the packet arrival process. Also, on a higher level, the arrival process contains an additional secondary correlation, resulting from the fact that the behaviour of the user population is governed by a two-state Markovian environment. Specifically, the state of this user environment in a particular slot determines the distribution of the number of newly generated messages in that slot.

In previous work on this model, we provided analytical results for the moments and the tail distribution of the system contents. Using these results, we now concentrate on the message delay performance of this system, under the important assumption of a first-come-first-served queueing discipline for packets, whereby packets that arrive during the same slot are stored in random order. Closed-form expressions are derived for the mean value of both the total delay and the transmission time of an arbitrary message. Additionally, we provide a reasonably tight upper and lower bound for the tail probabilities of the message delay. By means of some numerical examples, we discuss the influence of the environment parameters on the delay performance. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Discrete-time queueing model; Train arrivals; Variable-length messages; Message delay; Markovian environment

1. Introduction

At the edge of an ATM (asynchronous transfer mode) network, large external data frames, e.g., IP (Internet Protocol) frames, are segmented into fixed-length ATM cells, which are then transported through the network [1,2]. In order to model the correlation in traffic streams that results from this segmentation, several researchers have analysed discrete-time buffer systems where the customers are messages consisting of a random number of fixed-length packets; see, e.g., [3–12]. Here the term message corresponds to the external data format, whereas the term packet is used to denote the internal format. Time is assumed to

* Corresponding author. Tel.: +32-092648902; fax: +32-092644295.
E-mail address: sdv@telin.rug.ac.be (S. De Vuyst).

0166-5316/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0166-5316(02)00033-0
be divided into fixed-length intervals, referred to as slots, such that one slot suffices to transmit exactly one packet from the buffer. A message enters the buffer like a train at the rate of one packet per slot and packets are assumed to leave the buffer at the end of a slot. Various distributions for the lengths of the messages have been considered: a geometric distribution [3–6], constant-length messages [7,8] and an arbitrary distribution [9,10]. The analyses [3–12] however, all assume that the numbers of newly generated messages during the consecutive slots constitute a sequence of independent and identically distributed (i.i.d.) random variables. A somewhat related queueing model is considered in [11], where the delay of messages consisting of a fixed number of packets is treated under the assumption of an uncorrelated packet arrival process. In [12], a queueing model with multiple classes of messages is studied. Depending on its class, the packets of a message arrive either as a train or as a batch, with the one having priority over the other.

An important difference of our model with that of most of the existing studies about multiplexer models (see, e.g., [3,4,13,14]) is that we do not consider the arrival stream to be the superposition of traffic coming from a finite number of sources. Instead of characterising individual sources, we propose a global description of the aggregate packet arrival stream, as explained in the next section. Our model envisions a ‘multiplexer’ in a more generic sense than the classical device with a fixed number of limited-bandwidth input lines. In the classical sense, each message is sent to the queue by a ‘user’ claiming an input line with a link capacity of one packet per slot. The number of concurring messages (or active users) is then bounded by the number of input lines. Instead as an aggregate of individual sources, the arrival stream in our model comes from one centralised source that can be seen as an unbounded pool of ‘users’, capable of sending one message at a time, also at one packet per slot. There is no upper limit on the number of messages (users) that can simultaneously be active, so there is no bandwidth limitation other than the stability condition of the queue. Neither does the generation of new messages depend on how many of them are already active.

Our previous papers [15,16] are the result of an attempt to take into account a possible correlation in the message generation process. In particular, in [15,16], the distribution of the number of newly generated messages or leading packet arrivals in a slot is assumed to depend on the value of some environment variable, which represents the behaviour of the user population. There are two possible environment states, each with geometrically distributed sojourn times. In [15], the mean buffer contents and the mean packet delay were obtained, while [16] also provides results for the variance and tail behaviour of the buffer contents. The present paper is a continuation of [16], in the sense that now results are derived for the message delay and the message transmission time. Note that the considered correlated packet-train arrival process contains two types of correlation: a primary correlation, which results from the fact that one message may cause packet arrivals in the buffer during several consecutive slots, and a secondary correlation, which is due to the non-independent generation of new messages. It is important to note that these types of correlation are strictly Markovian by nature and have an exponential decay over larger time lags. Our arrival model thus qualifies as strictly short range dependent (SRD) and our analysis as it stands now is unable to cope with long range dependent (LRD) traffic.

In Section 2, we give a qualitative description of the system under study, along with some basic assumptions. The stochastic model and the system equations are given in Section 3. Some results of [15], which are necessary for the analysis of the present paper, are summarised in Section 4. The characteristics of the mean message delay and the mean message transmission time are studied in Sections 5 and 6, under the assumption of a first-come-first-served (FCFS) queueing discipline for packets. In Section 7, we discuss the tail behaviour of the message delay and finally, in Section 8, the results are discussed and illustrated with some examples.
2. Model description: non-independent generation of messages

We study a discrete-time single-server queueing system with infinite storage capacity. A user population generates entities, referred to as messages, which consist of a variable number of fixed-length packets. These messages are delivered to the buffer at a rate equal to the output link rate of the queue, i.e. one packet per slot (train arrivals). Note that we do not assume a limit on the number of users simultaneously sending packets to the multiplexer (i.e. being active). Hence, the number of packet arrivals per slot is not necessarily bounded.

In practice, our generic multiplexer model with its particular types of correlation in the arrival stream, is suited to describe queues at places where bandwidth is not an issue (like inside a node or a processor) and where the origin of the burstiness is centralised (like traffic coming mainly from one application). This may for instance be the case in systems where larger blocks of information (messages) are at some point transformed into contiguous streams of smaller fixed-size packets. Suppose the multiplexer operates at the level of a packet-based network layer and its input originates from an application in the layer immediately above. When a message is generated by the application (a user population in the upper layer), it is presented to the lower layer for transmission. However, since the lower layer is packet-based, the message needs to be broken down in packets at the edge of the two layers. As is the case in many realistic systems, the packets originating from one transmission request (one message) do not enter the multiplexer as a batch, but as a train: one by one. It is obvious that, due to this origin, the aggregated packet stream in the lower layer will be heavily correlated (primary correlation): for instance, if a message with a length of two packets is generated during some slot, it is sure that at least one packet will arrive to the multiplexer during the next slot, whereas this does not need to be the case for any arbitrarily chosen slot.

To be able to analyse the multiplexer performance, we still need to specify the message generating behaviour of the user population. Instead of just assuming that this is an uncorrelated process, we want to assess the impact of a possible correlation or burstiness in the generation of messages at the higher layer. Therefore, as in [15,16], we extend the model of [9] with a simple Markovian environment of which the state determines the number of newly generated messages in a slot. Specifically, the user population can be in one of two possible environment states, say ‘1’ and ‘0’. During the ‘1’-slots, the number of new messages is given by a random variable which we assume to have a value of no less than 1. As such, the ‘1’-periods represent bursts of high user activity, in which a lot of new messages (at least one per slot) are initiated. During the ‘0’-slots on the other hand, the activity of the users is rather low and typically, the generation of one or more messages per slot happens only once in a while. That is, during each ‘0’-slot the number of newly initiated messages per slot is given by a random variable which can be 0 with finite probability. Translated to the packet stream, this means that in the ‘1’-periods, at least one packet arrives to the multiplexer per slot and the buffer contents cannot decrease, i.e. the queue is in an overload situation. Of course, if the system is stable, this must be compensated for during the ‘0’-periods, in which slots occur without packet arrivals. In [15,16], we showed that the (secondary) correlation introduced in the packet stream by this environment mechanism has a major impact on the multiplexer performance, at least what the buffer occupancy is concerned.

In the present paper, we investigate the delay and the transmission time experienced by an arbitrary message in the steady state. To this end, we make the following assumptions regarding the single-server queue. Firstly, the queueing discipline is ‘FCFS for packets’. This means that the server is completely
ignorant of the fact that each packet belongs to a certain message and just transmits the packets in the order they were stored in the buffer. Secondly, we note that for the analysis of the buffer contents in terms of packets, the storage order (and hence, the order of transmission) of packets arriving in the same slot was not relevant. However, for the delay analysis, it is crucial to specify the policy used by the multiplexer to decide in which order these simultaneously arriving packets are stored in the buffer. In our model, we assume that this happens in random order.

3. Mathematical model and system equations

Since the environment state process is Markovian, the two states ‘1’ and ‘0’ both have geometric sojourn times:

\[ \Pr[\text{length of a ‘1’-period is } n] = (1 - \alpha)\alpha^{n-1}, \quad n \geq 1. \]
\[ \Pr[\text{length of a ‘0’-period is } n] = (1 - \beta)\beta^{n-1}, \quad n \geq 1. \]

Moreover, because of the memoryless property of the geometric distribution, knowledge of the value of \( t_{k-1} \), the environment state in slot \( k-1 \), is sufficient to determine the probability distribution of \( t_k \). Specifically, we have that

\[ E[z t_k | t_{k-1} = 1] = 1 - \alpha + az \triangleq a(z), \quad E[z t_k | t_{k-1} = 0] = \beta + (1 - \beta)z \triangleq b(z), \]

(1)

where \( E[\cdot] \) denotes the expected value operator. For ease of notation, we also define the function \( R(z) \) as \( R(z) \triangleq a(z)/b(z) \). Instead of by \( \alpha \) and \( \beta \), the user environment can also be characterised by the more comprehensible set of parameters \( \sigma \) and \( K \), which are defined as follows. Suppose the environment state is ‘1’ with probability \( \sigma \) and ‘0’ with probability \( 1 - \sigma \), independently from slot to slot. The mean sojourn times are then given by \( 1/(1 - \sigma) \) and \( 1/\sigma \), respectively. It is clear that the overall fraction of ‘1’-slots remains equal to \( \sigma \) if the mean lengths of the ‘1’-periods and the ‘0’-periods are both multiplied by the same factor \( K \), even if the environment is no longer independent then. So, in general, each \( (\alpha, \beta) \) corresponds to a certain \( (\sigma, K) \) and vice versa, such that

\[ E[\text{length of ‘1’-period}] = \frac{1}{1 - \sigma} = \frac{K}{1 - \sigma}, \quad E[\text{length of ‘0’-period}] = \frac{1}{1 - \beta} = \frac{K}{\sigma}. \]

(2)

The factor \( K \) can be seen as a measure for the absolute lengths of the sojourn times, whereas the parameter \( \sigma \) characterises their relative lengths. Therefore, we shall henceforward call \( K \) the burst-length factor of the environment. Moreover, it turns out that the correlation coefficient \( \phi \) between the environment states in two consecutive slots is determined only by the value of \( K \):

\[ \phi \triangleq \frac{E[t_i t_{i+1}] - E[t_i]E[t_{i+1}]}{\sqrt{\text{Var}[t_i] \text{Var}[t_{i+1}]}} = -1 + \alpha + \beta = 1 - \frac{1}{K}. \]

(3)

The message lengths (the number of composing packets) are i.i.d. random variables with probability mass function \( l_r \) (\( r \geq 1 \)) and probability generating function (pgf) \( L(z) \triangleq \sum_{r=1}^{\infty} l_r z^r \). Since the messages arrive to the buffer in trains, a new message, or a user becoming active, is seen by the buffer as the arrival of a leading packet in the current slot and one packet arrival in each of the following consecutive slots up to a total given by the length of the message. We do not impose an upper bound on the message length although it is necessary to assume that the moments of its distribution
exist, so any heavy-tailed distributions are precluded. Note that the analysis in this paper is a priori valid for finite-length messages, i.e. if \( L(z) \doteq \sum_{r=1}^{N} l_r z^r \) for a certain \( N \geq 1 \), and can be much simplified in this case (see Appendix B). In our further analysis, instead of the probabilities \( l_n \) \( (n \geq 1) \), we shall often use the probabilities \( q(n) \) \( (n \geq 1) \), which are defined as follows: \( q(n) \) denotes the probability that a message that is already \( n \) packets long in the current slot, will still continue in the next slot, i.e.

\[
q(n) \doteq \sum_{r=n+1}^{\infty} l_r \sum_{r=n}^{\infty} l_r, \quad n \geq 1.
\]

Let us now denote by \( m_{n,k} \) the number of users that send the \( n \)th packet of a message during slot \( k \). The number of leading packet arrivals or newly generated messages \( m_{1,k} \) in slot \( k \) is assumed to have a probability distribution that only depends on the environment state \( t_k \) in slot \( k \), and not on the numbers of new messages generated during previous slots. We will use the following notations:

\[
M_0(z) \doteq E[z^{m_{1,k}} | t_k = 0], \quad M_1(z) \doteq E[z^{m_{1,k}} | t_k = 1].
\]

These distributions must also have finite moments. We repeat here the above-mentioned assumption that \( M_1(0) = 0 \), thus interpreting ‘1’ as the ‘active’ environment state in which the users generate at least one new message per slot. Again, for ease of notation, we introduce the function \( \mu(z) \), defined as \( \mu(z) \doteq M_1(z)/M_0(z) \). Now, for \( n \geq 1 \), the random variable \( m_{n+1,k} \) can be expressed as

\[
m_{n+1,k} = \sum_{i=1}^{m_{n,k}} d_{n,i}, \quad n \geq 1.
\]

Indeed, the number of messages that generate their \((n+1)\)th packet in slot \( k \) corresponds to the number of messages that generate their \( n \)th packet during slot \( k-1 \) and continue in slot \( k \) by sending an \((n+1)\)th packet. The \( d_{n,i} \) \( (n \geq 1) \) in (6) are, for given \( n \), i.i.d. random variables with common pgf

\[
D_n(z) = q(n)z + 1 - q(n).
\]

Finally, let \( s_{k+1} \) represent the system contents just after slot \( k \) (i.e. at the beginning of slot \( k+1 \)), where the term system contents indicates the number of packets that are either waiting or being transmitted. Since the server transmits a packet whenever one is available in the buffer, the evolution of the system contents from slot to slot is described by

\[
s_{k+1} = e_k + (s_k - 1)^+,\]

where \((\cdot)^+ = \max(\cdot, 0)\) and \( e_k \) is the total number of packet arrivals in slot \( k \), which can be further expressed as

\[
e_k = \sum_{n=1}^{\infty} m_{n,k}.
\]

It is now clear that the arrival process of our model is determined by the following parameters, listed in Fig. 1: \( \sigma \) and \( K \) for the environment, the pgf \( L(z) \) of the message length and the pgfs \( M_1(z) \) and \( M_0(z) \) of the numbers of new messages in a ‘1’-slot and ‘0’-slot, respectively. This figure also shows an example explaining the meaning of the random variables \( m_{n,k} \) and (9).
In the following sections, we shall always define the delay of either packets or messages to consist of an integer number of slots, i.e. we do not consider a packet to be present in the system before the end of the slot in which it arrived. As such, it suffices to observe the system at slot boundaries only. For the analysis of the buffer contents in terms of packets, the arrival order of the \( e_k \) packets in slot \( k \) is not relevant. However, for the delay analysis, we assume that packets arriving during the same slot are stored in random order.

In Fig. 2, we have used the system equations (1)–(9) to depict the correlation structure of the packet arrival process in the form of a flow chart. First of all, the chart demonstrates the intra-slot dependencies, i.e. how the number of packet arrivals \( e_k \) in slot \( k \) depends on both the environment state \( t_k \) and the numbers of users \( m_{n,k} \) \((n \geq 1)\) that are active in that slot. The chart also shows how these quantities determine the corresponding quantities in the next slot. Specifically, we see that the slot boundary is crossed by two arrows, corresponding to the two types of correlation mentioned earlier. The first arrow passes the environment state \( t_k \) to the next slot and accounts for what we have called the secondary correlation. It is clear from the chart that the environment state process \( \{t_k\} \) forms a one-dimensional Markov chain in its own right. The second arrow passes the value of the variables \( m_{n,k} \) \((n \geq 1)\) to the next slot and accounts for the primary correlation. Indeed, because the distribution of the message length is general, we have to keep track of the numbers of users that became active in each of the previous slots and still have a packet to send during the current slot. The information passed by both arrows together gives a full description of the state of the packet arrival process in slot \( k \). Therefore, the vectors \( \{t_k, m_{n,k} \} \ (n \geq 1) \) form an (infinite-dimensional) Markov chain by which the packet arrival process is characterised. In fact, if we consider the system as a whole (not only the arrival process), we should include a third arrow, which passes the system contents \( s_{k+1} \) to the next slot and accounts for the relation (8). All three arrows together, i.e. the vectors \( \{t_k, m_{n,k} \ (n \geq 1), s_{k+1}\} \) are sufficient to fully describe the state of the multiplexer after each slot. Therefore, we can conclude that these vectors—again—form an infinite-dimensional Markov chain, and we shall often refer to the set of variables \( t_k, m_{n,k} \ (n \geq 1), s_{k+1} \) as the system state variables for slot \( k \).
4. Preliminary results

Let us assume now that for \( k \to \infty \), the system reaches a stochastic equilibrium. During an arbitrary slot in this steady state, the joint distribution of the variables comprising the system state after slot \( k \), will no longer vary with \( k \), and we shall use \( t, m_n \) (\( n \geq 1 \)), \( e \) and \( s \) to denote the steady-state versions of \( t_k \), \( m_{n,k} \) (\( n \geq 1 \)), \( e_k \) and \( s_{k+1} \), respectively. To start the analysis of the multiplexer behaviour, we define the joint pgf of the system state vector \((t, m_{n,n} (n \geq 1)), s)\) in the steady state:

\[
P(x, y_1, y_2, \ldots, z) \equiv \mathbb{E}[x^{t}y_1^{m_{n,1}}y_2^{m_{n,2}}\ldots z^{s}].
\]

In [15,16], we have established the following fundamental equation for \( P(x, y_1, y_2, \ldots, z) \), based on the system equations (1)-(9):

\[
P(x, y_1, y_2, \ldots, z) = \frac{1}{z}M_0(z) \left( P(R(x \mu (y_1 z)), D_1(y_2 z), D_2(y_3 z), \ldots, z) + p_0(z-1) \right),
\]

where the quantity \( p_0 \) indicates the steady-state probability of having an empty system at the start of an arbitrary slot. Obviously, solving Eq. (10) for the function \( P \) directly is not easy, save perhaps for some very simple choices for the pgfs \( L(z), M_1(z), M_0(z) \) and the parameters \( \sigma \) and \( K \). Nevertheless, in Appendix A, we show how an iterative solution of the functional equation can be constructed. Furthermore, it is possible to derive from (10) general closed-form results regarding the distributions of the marginal

---

Fig. 2. Flow chart of the packet arrival process: primary and secondary correlation.
processes in the steady state, as shown in [15,16]. Specifically, for the pgf $T(x)$ of $t$, we have $T(x) = P(x, 1, \ldots, 1) = \sigma x + (1 - \sigma)$ and the pgf $\hat{E}(y)$ of the number of leading packet arrivals $m_1$ in an arbitrary steady-state slot is given by

$$\hat{E}(y) = P(1, y_1, 1, \ldots, 1) = \sigma M_1(y_1) + (1 - \sigma)M_0(y_1).$$

(11)

Intuitively, one understands that the average number of packet arrivals $E[\sigma]$ per slot in the steady state is given by the average number of new messages $E[m_1] = \hat{E}(1)$ multiplied by the mean length $L(1)$ of a message. Since the multiplexer has only one server, this quantity $E[\sigma]$ also equals the load $\rho$ of the system. Moreover, it is shown that the probability $p_0$ is given by $1 - \rho$. The stability condition for the system can thus be written as

$$\rho = E[\sigma] = \hat{E}(1)L(1) = 1 - p_0 < 1.$$  

Let the pgf of the system contents $s$, i.e. the number of packets in the system at the end of an arbitrary slot in the steady state, be denoted by $S(z) \triangleq P(1, 1, \ldots, 1, z)$. It is not possible to derive the mean system contents $E[s] = S'(1)$ from (10) by merely differentiating both sides with respect to $z$ and letting all arguments be equal to 1, since this method leaves $(\partial/\partial z)P(1, 1, \ldots, 1) = S'(1)$ undetermined. In [15,16], we have used the following technique instead: consider (10) for only those values of $x$ and $y_n$, $n \geq 1$, for which the respective arguments of the functions $P$ on both sides of the equation are equal to each other. This yields

$$y_n = y_n(z) = \lim_{N \to \infty} D_n(zD_{n+1}(\ldots zD_N(z) \ldots)) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{z^m}{m!} n!,$$

(12)

and $s = \chi(z)$ is given implicitly by

$$\chi(z) = R(\chi(z)\mu(L(z))).$$

(13)

These values of $(s, y_n, n \geq 1, z)$ form a one-dimensional trajectory in the domain of $P$ which we can choose to contain the point $(1, 1, 1, 1)$ and in which the functional equation is reduced to a simple linear equation that can easily be solved:

$$P(s, \eta_1(z), \eta_2(z), \ldots, z) = \frac{p_0(z - 1)G(z)}{z - G(z)}$$

(14)

with $G(z) \triangleq (1 - \beta)\chi(z)\mu(L(z)) + \beta M_0(L(z))$. Now, total differentiation of (14) and evaluation for $z = 1$ yields an expression for the partial derivative (PD) of $P(x, y_1, y_2, \ldots, z)$ with respect to $z$ in the point $(1, 1, 1, 1, \ldots)$, i.e. for $S'(1)$, in terms of all other (infinitely many, in fact) first-order PDs of $P$, also in the point $(1, 1, 1, \ldots, 1)$. However, unlike $S'(1)$, the latter PDs can consecutively be derived from (10) in a direct way. Doing so, we eventually find for the mean system contents:

$$E[s] = \hat{E}(1)L(1) + \frac{(\hat{E}(1))^2L'(1) + \hat{E}(1)L'(1)}{2(1 - \hat{E}(1)L'(1))} - L'(1)$$

$$+ \sigma(K - 1)(M_1(1) - M_0(1))L'(1) \left(1 - \frac{M_1(1) - M_0(1)L'(1)}{1 - \hat{E}(1)L'(1)} \right).$$

The above scheme for the calculation of $E[s]$, using the consecutive PDs of the joint pgf, can also be extended to obtain higher-order moments of $s$. For example, finding $\text{Var}[s]$ involves the calculation of all
second-order PDs of $P$, again in the point $(1, 1, \ldots, 1)$. We shall need some of these PDs later on for the calculation of the mean message delay and transmission time in the next sections. In particular, we need the following PDs of $\frac{\partial}{\partial y_j}P(1, 1, \ldots, 1)$:

$$\frac{\partial^2}{\partial x \partial y_j}P(1, 1, \ldots, 1) = \sigma M'_1(1), \quad \frac{\partial^2}{\partial y_j^2}P(1, 1, \ldots, 1) = \mathcal{E}'_1(1), \quad \text{for } j \geq 2, \quad \text{(15)}$$

$$\frac{\partial^2}{\partial z \partial y_j}P(1, 1, \ldots, 1) = \mathcal{E}'_1(1) + \mathcal{E}''_1(1) + \mathcal{E}'(1)(1 - \mathcal{E}'_1(1))\left(1 - \frac{L(\phi)}{\phi}\right), \quad \text{(17)}$$

5. Mean message delay

In this section, we deal with the delay experienced by messages, rather than individual packets. We define the message delay $c$ as the time period between the end of the slot in which the first packet of a message is generated and the end of the slot during which the last packet of this message is transmitted. In what follows, we extend the technique used in [5,9] to obtain the mean message delay $E[c]$. To start with, let us first consider an arbitrary but tagged message $M$, of which the leading packet enters the multiplexer during some slot $J$ in the steady state. Let $t(J+n), m_1(J+n), m_2(J+n), \ldots, s(J+n)$ be the system state variables for slot $J+n$. Also, by $e(J+n)$, we indicate the total number of packet arrivals during slot $J+n$. Furthermore, if we denote by $L$ the length of message $M$, then for $0 \leq n \leq L-1$, one of the packets contained in $e(J+n)$ will be the $(n+1)$th packet of the tagged message $M$, and by $e^*(J+n)$ we shall count those packets arriving in slot $J+n$ that are to be transmitted no later than this particular packet of $M$. Now, using similar methods as in [3,5,17], it can be shown that the joint mass function of the system state variables in slot $J$ is given by

$$\text{Pr}[t(J) = i, m_1(J) = j_1, m_2(J) = j_2, \ldots, s(J) = k] = \frac{j_1}{\mathcal{E}'_1(1)} \text{Pr}[t = i, m_1 = j_1, m_2 = j_2, \ldots, s = k], \quad \text{(18)}$$

and the corresponding pgf $\hat{P}(x, y_1, y_2, \ldots, z)$ is given by

$$\hat{P}(x, y_1, y_2, \ldots, z) = \frac{y_1}{\mathcal{E}'_1(1)} \frac{\partial}{\partial y_1}P(x, y_1, y_2, \ldots, z). \quad \text{(19)}$$

From (18) and (19) it is clear that the statistics of the system in the steady-state depend on the events on which one chooses to observe the multiplexer: the joint pgf of the system state ‘on slot boundaries’ ($P$) differs from the pgf of the system state ‘as seen by new messages’ ($\hat{P}$). In Appendix A, we shall use (19) to give a more explicit expression for $\hat{P}(x, y_1, y_2, \ldots, z)$. 
In order to calculate the mean value of the message delay $c$, we first condition on the length of $M$, i.e.

$$E[c] = \sum_{k=1}^{\infty} E[c_k] p_k,$$  \hspace{1cm} (20)

where $c_k$ denotes the delay of an arbitrary message of length $k$. To derive the conditional mean delay $E[c_k]$, we make a distinction between the cases for $L = 1$ and for $L > 1$. In Fig. 3, we have depicted the packet arrivals in the consecutive slots during which message $M$ is delivered to the buffer. Note that we have drawn the arriving packets in the order they are stored in the buffer and that in each slot, the position of the packet belonging to $M$ is random (i.e. uniformly distributed). In the case $L = 1$, it is clear that the message delay is given by

$$c_1 = s(J) - e(J) + e^*(J),$$ \hspace{1cm} (21)

which corresponds to the shaded packets in Fig. 3(a). If the system is non-empty at the beginning of slot $J$, then $s(J) - e(J)$ is the number of packets in the system at that moment minus the one packet being served during slot $J$; otherwise, it is zero. Also, $e^*(J)$ is the number of packet arrivals in slot $J$ to be transmitted no later than the packet of $M$. Next, if $L = k > 1$, then the message delay is given by the number of shaded packets in Fig. 3(b):

$$c_k = s(J) + \sum_{n=1}^{k-2} e(J+n) + e^*(J+k-1), \quad k > 1,$$ \hspace{1cm} (22)

where the second term is dropped if $k = 2$, by convention. Due to the fact that the packet belonging to $M$ could be any of the $e(J+n)$ packets arriving in slot $J+n$, it can be seen that

$$E[e^*(J+n)] = \frac{z}{z-1} \frac{z^{e(J+n)} - 1}{e(J+n)} \quad \text{and} \quad E[e^*(J+n)] = \frac{1}{2} E[e(J+n)] + \frac{1}{2}, \quad n \geq 0.$$ \hspace{1cm} (23)
Then, after using (19) again, together with (11) and (16), we find
\[
E[c] = E[s(J)] - \frac{1}{2} E[e(J)] + \frac{1}{2} + \frac{1}{2} \sum_{k=2}^{\infty} E[e(J + k)] + \frac{1}{2} \sum_{k=2}^{\infty} E[e(J + k - 1)] \frac{1}{2}.
\]
(24)

The mean system contents \( E[s(J)] \) as seen by an arbitrary message \( M \) can be calculated from (19) in terms of the PD of \( P \) given by (17)
\[
E[s(J)] = \frac{1}{b(J)} \sum_{\nu=1}^{\infty} \frac{\beta^\nu}{\gamma^\nu}. \]
(25)

Similarly, to find the mean number of arrivals \( E[e(J)] \) in slot \( J \), we can first use the system equation (9):
\[
E[e(J)] = \sum_{i=0}^{\infty} E[n_i(J)] = \sum_{i=0}^{\infty} \frac{\beta}{\gamma^\nu} \phi(N_i). \]

Then, after using (19) again, together with (11) and (16), we find
\[
E[e(J)] = \frac{E_s(1) + (E_s(1)^2 - E_s(1))}{E_s(1)} + \frac{E_s(1) - E_s(1)}{E_s(1)} = \frac{1}{1 - \phi}. \]
(26)

Now, the only quantities in (24) that remain to be determined are the mean numbers of arrivals \( E[e(J + n)] \) in the slots \( J + n \) \( (n \geq 1) \), given that \( M \) is still active in those slots. To this end, we shall first derive in the following paragraphs the expected values of the environment state \( t(J + n) \) and of the numbers of users \( m_i(J + n) \) sending their \( i \)th packet \( (i \geq 1) \) in slot \( J + n \). Furthermore, in the remainder of this section, we always assume that \( n \geq 1 \) and that a packet from message \( M \) arrives in slot \( J + n \).

5.1. Environment state in slot \( J + n \)

From (1), we have for \( T_{J+n}(z) \), the pgf of \( t(J + n) \), that
\[
T_{J+n}(z) = a(z)P[t(J + n - 1) = 1] + b(z)P[t(J + n - 1) = 0] = b(z)T_{J+n-1}(R(z)), \quad (27)
\]
\[
E[(z)^{J+n} \phi] = E[b(z)(R(z))^{J+n-1}]. \]
(28)

Differentiating both sides of (27) and evaluation in \( z = 1 \) yields
\[
E[t(J + n)] - \beta + E[t(J + n - 1)] \phi = (1 - \beta)\frac{1 - \phi^n}{1 - \phi} + E[t(J)] \phi^n, \quad (29)
\]
where we have recursively applied the first equation to obtain the latter. Next, considering the fact that \( E[t(J)] = (\partial/\partial x)P(1, 1, \ldots, 1) \), we can conclude from (15), (19) and (29) that
\[
E[t(J + n)] = \sigma + \sigma(1 - \sigma)\frac{M(1) - M_1(1)}{E_1(1)} \phi^n. \quad (30)
\]

Note that \( E[t(J + n)] = \sigma \), as would be the case for the environment state in an arbitrary slot. However, slot \( J + n \) is not just any slot, but the \( n \)th slot after the arrival of the tagged message \( M \), which is statistically very different! The same remark can be made for the other expressions in this section regarding averages taken in slot \( J + n \).
5.2. New messages in slot $J + n$

Let $E_{1, J + n}(z)$ be the pgf of $m_1(J + n)$. In a similar way as for (27), it follows from (5) that

$$E_{1, J + n}(z) = E[z^{m_{1}(J + n)}] = E[M_0(z)(\mu(z))^{(J + n)}] = M_0(z)T_{J + n}(\mu(z)).$$  \hfill (31)

Then, after differentiating (31) and using (30), we find

$$E[m_1(J + n)] = E'_{1}(1) + \sigma(1 - \sigma) \frac{(M'_1(1) - M'_0(1))^2}{E'_1(1)} \phi^*.$$  \hfill (32)

5.3. Active messages initiated before slot $J + n$

In view of Eq. (6), it follows that for $i > 1$:

$$m_i(J + n) = \sum_{j = 1}^{m_{i-1}(J + n - 1)} d_{i-1,j}, \quad i \neq n + 1, \quad m_{n+1}(J + n) - 1 = \sum_{j = 1}^{m_{n}(J + n - 1) - 1} d_{n,j},$$  \hfill (33)

where, for given $i$, the $d_{i-1,j}$s are i.i.d. with common pgf $D_{i-1}(z)$, given in (7), and with expected value $q(i - 1)$. The term $-1$ in (33) accounts for the $(n + 1)$th packet of $M$ that is certain to arrive in slot $J + n$.

Taking expected values of (33), we obtain

$$E[m_i(J + n)] = q(i - 1)E[m_{i-1}(J + n - 1)], \quad 1 < i \neq n + 1,$n

$$E[m_{n+1}(J + n)] - 1 = q(n)(E[m_{n}(J + n - 1)] - 1).$$

Recursive application of the above equations and using definition (4), then yields

$$E[m_i(J + n)] = \left(\sum_{r = 2}^{+\infty} l_r\right) E[m_1(J + n - i + 1)], \quad 1 < i < n + 1,$$  \hfill (34)

$$E[m_{n+1}(J + n)] - 1 = \left(\sum_{r = 2}^{+\infty} l_r\right) (E[m_1(J)] - 1),$$  \hfill (35)

$$E[m_i(J + n)] = \sum_{r = 1}^{+\infty} l_r E[m_{i-1}(J)] + \sum_{r = 2}^{+\infty} l_r E[m_{n+1}(J)], \quad i > n + 1.$$  \hfill (36)

Now, in (34), the averages on the right-hand side can directly be obtained from (32). On the other hand, in (35) and (36), the quantities $E[m_i(J)]$ are given by $(\partial/\partial y_i) \hat{P}(1, 1, \ldots, 1)$ for each $i \geq 1$, respectively.

Then, using (19) and the needed PDs of $P$, we find

$$E[m_i(J + n)] = \left(\sum_{r = 1}^{+\infty} l_r\right) \left[\frac{E'_1(1) + \sigma(1 - \sigma) \frac{(M'_1(1) - M'_0(1))^2}{E'_1(1)} \phi^{n+i}}{n+i}\right], \quad 1 < i \neq n + 1,$$  \hfill (37)

$$E[m_{n+1}(J + n)] = 1 + \left(\sum_{r = 1}^{+\infty} l_r\right) \frac{E'_1(1)}{E'_1(1)}.$$  \hfill (38)
Since the mean value of the total number of packet arrivals $E[e(J + n)]$ is given by the sum of all $E[m_i(J + n)], i \geq 1$, we can now appropriately substitute the terms of this sum by the expressions (32), (37) and (38), depending on whether $i = 1, 1 < i \neq n + 1$ or $i = n + 1$, respectively. This yields

$$E[e(J + n)] = 1 + E'_1(1)L'(1) + \frac{E''_1(1) - (e''(1))^2}{E'_1(1)} \left( \sum_{r=1}^{+\infty} l_r \right) + \sigma(1 - \sigma) \frac{(M'_1(1) - M''_1(1))^2}{E'_1(1)}$$

$$\times \frac{1}{\phi} \left[ -\phi^n + \sum_{r=1}^{+\infty} l_r + \sum_{r=1}^{+\infty} l_r + \sum_{r=1}^{+\infty} \phi^{n-1} l_r - \sum_{r=1}^{+\infty} \phi^{n-1} l_r \right].$$

Finally, substitution of the results (25), (26) and (39) in (24) leads to an expression for the mean message delay $E[c]$. After some very tedious calculations, we find

$$E[c] = E[s] + L'(1) + \frac{L'(1)}{2E'_1(1)} \left( E''_1(1) \right) \left( 2L'(1) - 5 \right) + 2E''_1(1)$$

$$- \frac{E''_1(1) - (e''(1))^2}{E'_1(1)} \sum_{r=1}^{+\infty} \sum_{j=0}^{+\infty} \left( j + \frac{1}{r} \right) l_{r,j} + \sigma(K - 1) \frac{M'_1(1) - M''_1(1)}{E'_1(1)}$$

$$\times \left[ M'_1(1) - 1 + E'_1(1)(L'(1) - 1) + (1 - \sigma)(M'_1(1) - M''_1(1)) \right]$$

$$\times \left[ 2L'(1) - 2 - (K - 1) \left( 1 - \frac{L}{\phi} \right) \left( 1 - \frac{1}{\phi^j} \right) \right] + 2\sigma(1 - \sigma)(K - 1)$$

$$\times \frac{(M'_1(1) - M''_1(1))^2}{E'_1(1)} \sum_{r=1}^{+\infty} \sum_{j=1}^{+\infty} \left( K - j - \frac{(K - 1/2)^2}{K} \phi^{j-r} \right) l_{r,j}.$$  

(40)

6. Mean message transmission time

We define the transmission time of a message as the time period between the beginning of the slot in which the leading packet of the message is transmitted and the end of the slot in which the last packet is transmitted. Let $h$ denote the transmission time of an arbitrary message in the steady state, then we show in this section, how the previously obtained results can be used to compute the mean transmission time $E[h]$. Specifically, let us consider again the tagged message $M$ and condition on the length $L$ of $M$, i.e.

$$E[h] = \sum_{i=1}^{k} E[h_i]l_i,$$

(41)

where $h_i$ denotes the transmission time of an arbitrary message of length $k$. Obviously, if $L = 1$, then $h_i = 1$. On the other hand, if $L = k > 1$, then

$$h_i = e^\sigma(J) + \sum_{i=1}^{k-1} e(J + n) + e(J + k - 1), \quad k > 1,$$

(42)
where \( e^*(J) \) indicates the number of packets arriving in slot \( J \), but to be transmitted no sooner than the packet of \( M \) that arrives in slot \( J \) (see Fig. 4). Note that, since the position of this packet among the arrivals in slot \( J \) is random, \( e^*(J) \) has the same distribution as \( e^*(J) \) and its mean value is given by (23).

Thus, it follows from (41) and (42) that

\[
E[h] = 1 + \frac{1}{2} E[e(J)](1 - l_i) + \sum_{k=2}^{\infty} \sum_{i=1}^{k-1} E[e(J + n)]l_i - \frac{1}{2} \sum_{k=2}^{\infty} E[e(J + k - 1)]l_i
\]

\[
= E[e] - E[s(J)] + \frac{1}{2} + \frac{1}{2} E[e(J)],  
\]  

(43)

where we have used (24). All terms appearing in (43) were studied in the previous section, and we arrive at the following closed-form expression for the mean message transmission time \( E[h] \):

\[
E[h] = L'(1) + \frac{1}{2} E_1(1) + E_1'(1) L'(1)(L'(1) - 2) + \frac{E_1'(1)}{E_1(1)} (L'(1) - \frac{1}{2})
\]

\[
- \frac{E_1'(1)}{E_1(1)} \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \left( j + \frac{1}{2} \right) l_i l_{i+j} + 2\sigma(1 - \sigma)(K - 1) \frac{(M'_1(1) - M'_2(1))^2}{E_1(1)}
\]

\[
\times \left[ L'(1) - 1 - (K - \frac{1}{2}) \left( 1 - \frac{L(\phi)}{\phi} \right) + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left( K - j - \frac{(K - 1/2)^2}{K} \phi^{j-1} \right) l_{i+j} \right] .
\]

(44)

In addition to the message delay and transmission time, many authors (e.g., [5,8,9]) also study the message waiting time. We define the waiting time \( w \) of a message in the steady state as the time period between the end of the slot during which the first packet of the message arrives and the time instant at which the transmission of this packet is about to start. From this definition, it is clear that \( w = c - h \). Therefore,
from (43), the mean waiting time $E[w]$ is given by

$$E[w] = E[s(J)] - \frac{1}{2}E[e(J)] = E[s] + \frac{E'(1)}{2E'(1)} \cdot \frac{1}{2}E'(1) (L'(1) + 1)$$

$$+ \sigma (K - 1) \frac{M'(1) - M_0(1)}{E'(1)} \left[ M'(1) - 1 + E'(1) (L'(1) - 1) + (1 - \sigma) \right] \beta^{-1}.$$  \hspace{1cm} (45)

7. Tail behaviour of the message delay

Whereas in Section 5, we have considered only the first moment of the message delay $c$, we now want to study the $z$-transform of its complete distribution and give an approximation for the distribution of the tail that can be implemented numerically. Specifically, we first construct an expression for the pgf $c_k(z)$ of $k$, the delay of an arbitrary message that consists of $k$ packets. Obviously, the pgf of the unconditional message delay $c$ is then given by $c(z) = \sum_{k=1}^{+\infty} c_k(z) k$. To assess the tail distribution of $c_k$, we use the following approximation technique described in [17,18], which is known to yield very accurate results:

From the inversion formula of $z$-transforms, it follows that the probability mass function $\text{Pr}[c_k = n]$ can be expressed as a weighted sum of negative $n$th powers of the poles of $c_k(z)$. Since all these poles have a modulus larger than 1, $\text{Pr}[c_k = n]$ is dominated by the contribution of the pole $z_0$ with the smallest modulus. It is shown in [18] that this ‘dominant’ pole $z_0$ must necessarily be real and positive in order to ensure a non-negative mass function $\text{Pr}[c_k = n]$. Therefore, the probability for a message of length $k$ to experience a delay of more than $C$ slots can be expressed by the following geometric form for sufficiently large values of $C$:

$$\text{Pr}[c_k > C] \approx -\frac{\theta^{c_k}}{z_0 C} \left( \frac{1}{z_0 C} \right)^{C+1}.$$  \hspace{1cm} (46)

where $\theta^{c_k}$ is the residue of $c_k(z)$ in the point $z = z_0$. To identify $z_0$ and $\theta^{c_k}$ we thus need an expression for $c_k(z)$.

Let us consider again an arbitrary message $M$ of $L$ packets of which the first one arrives during slot $J$. We first treat the case where $L = k > 1$. Then it follows from (22) and (23) that

$$c_k(z) = E[z^{s(J)} z^{e(J+1)} z^{e(J+2)} \cdots z^{e(J+k-2)} z^{e(J+k-1)}]$$

$$= \frac{z}{z - 1} E \left[ \frac{z^{e(J+1)} z^{e(J+2)} \cdots z^{e(J+k-2)} z^{e(J+k-1)} - 1}{z^{J+k-1} - 1} z^{e(J)} \right]$$

$$= \frac{z}{z - 1} \int_{Y_k} \Delta_k(z, y_2, \ldots, y_k, z) \, dy_k.$$  \hspace{1cm} (47)

where we have introduced the joint generating function of the system contents at the end of slot $J$, together with the numbers of packet arrivals in each of the following $k - 1$ slots:

$$\Delta_k(y_2, y_3, \ldots, y_k, z) \equiv E[y_2^{s(J+1)} y_3^{s(J+2)} \cdots y_k^{s(J+k-1)} z^{e(J)}].$$  \hspace{1cm} (48)
Now, following a similar line of thought as in Section 5, it is possible to relate the random variables in (48) to the system state variables for slot $J$ only (this is demonstrated in Appendix B in case all arguments of $\Lambda_k$ are equal to $z$). We can therefore rely on result (A.7) of Appendix A for the joint pgf $\hat{P}$ as to find an exact expression for $\Delta_k(z, z, \ldots, y_k, z)$ and hence, for $\theta_n^\circ$. We have carried out the necessary calculations and found that the dominant pole $z_0$ is given implicitly by $z_0 = G(z_0)$, where $G(z)$ is defined in (14). Its value can thus easily be obtained numerically by using, e.g., the Newton–Raphson scheme.

However, to find the tail behaviour, we also need $\theta_n^\circ$, which follows from (47) as

$$\theta_n^\circ = \operatorname{Res}_{z=0} \theta_n(z) = \frac{z_0}{z_{0} - 1} \int \frac{1}{y_k} \Delta_k(z, z, \ldots, z, y_k, z) \, dy_k,$$

(49)

As such, the evaluation of $\theta_n^\circ$ involves the integration over a very complicated function of $y_k$. Indeed, as it turns out, the integrand of (49) is a product of an infinite number of factors in $y_k$, which makes it hard to implement the integration numerically. Similarly, for the case $L = 1$, it follows from (21) and (23) that

$$\theta_n^\circ = E[e^{\sum_{j=1}^{n} e^{\gamma_j} n} e^{\sum_{i=1}^{k-1} e^{\gamma_i} n}] = \frac{z}{z - 1} \int \frac{1 - e^{-\gamma_i} n z^{e(J)}}{e(J)} \, dz = \frac{z}{z - 1} \int \frac{1}{y} \hat{P}(1, y, y, \ldots, z) \, dy.$$

(50)

Here we can draw the same conclusions regarding $z_0$ and $\theta_n^\circ$ as above.

In short, although we might be able to give an analytically exact expression for $\theta_n^\circ$ ($k \geq 1$) and its residue $\theta_n^\circ$ in $z_0$, the actual evaluation of the latter would be very cumbersome due to the integrations in (47) and (50). Note that this integral arises from the fact that the arrivals in the last arrival slot of message $M$ are only partially comprised within $c_k$, i.e. it is the last term in (21) and (22) which gives difficulties. In the following paragraphs we try to avoid this problem by proposing an appropriate lower and upper bound for $c_k$.

### 7.1. An upper bound $\hat{c}_k$ for $c_k$

Let us call $\hat{c}_k$ the delay of a message $M$ with length $k$ in the worst-case scenario that of all the arrivals in a certain slot, the packet of $M$ is always the last one to be transmitted. This means that the term $e^{\gamma} (J + n)$, $n \geq 0$, in (21) and (22) is always equal to its maximum value $e(J + n)$ and we have that

$$c_k \leq \hat{c}_k \leq s(J) + \sum_{n=1}^{k-1} c(J + n), \quad k \geq 1$$

with generating function

$$\hat{c}_k(z) = E \left[ \sum_{i=1}^{k-1} e^{\gamma_i} n \right] = \Lambda_k(z, z, \ldots, z, z).$$

In Appendix B, we calculate the pgf $\Lambda_k(z, z, \ldots, z, z)$ in terms of the joint pgf $\hat{P}$, which in turn is calculated in Appendix A. Using the obtained expressions (A.7) and (B.7) then yields an explicit expression for $\hat{c}_k(z)$ in terms of the parameters of the arrival process only. Again, we can show that the dominant pole $z_0$ of $\hat{c}_k(z)$ is given by $G(z_0) = z_0$. Its residue $\theta_n^\circ$ in the point $z = z_0$ then follows as

$$\theta_n^\circ = A_{k-1}(z_0) A_{k-2}(z_0) h_1(z_0) A_{k-3}(z_0) h_2(z_0) \cdots A_1(z_0) h_{k-2}(z_0) \left( \frac{e^{\gamma}}{E(1)} \right) \frac{p_0(z_0 - 1)}{1 - G(z_0)} \times \left( \prod_{i=0}^{\infty} \frac{M_i(g_i)(y_i)}{M_i(0) h_i(y_i)} \right) \left( \frac{z_0 M_k(z_0) h_{k-1}(z_0)}{M_k(z_0) h_{k-1}(z_0)} + \sum_{i=1}^{\infty} \frac{1 - \beta}{\theta_{y_i}^\circ} \right),$$

(51)
where we make use of these particular evaluations of the functions (A.2) and (A.3) from Appendix A:

\[ g_n = g_n(h_{n,k+1}^{-1}(z_0), z_0) = \sum_{r=1}^{k} l_r z_0^{r-1} + z_0^{r-1} h_{n,k+1}^{-1}(z_0) \sum_{r=n+1}^{\infty} l_r, \quad (52) \]

\[ \gamma_0 = \gamma_0(f_k(z_0), h_{1,k+1}(z_0), z_0) = f_k(z_0) \mu(z_0 h_{1,k+1}(z_0)), \]

\[ \gamma_n = \gamma_n(f_k(z_0), h_{1,k+1}(z_0), h_{2,k+1}(z_0), \ldots, h_{n+1,k+1}(z_0), z_0) = R(\gamma_{n-1}) \mu(\gamma_n), \quad n > 0, \quad (53) \]

\[ \frac{\partial}{\partial y_n} \gamma_n = f_k(z_0) \mu(z_0 h_{1,k+1}(z_0)), \quad \frac{\partial}{\partial y_n} \gamma_n = R(\gamma_{n-1}) \mu(\gamma_n), \quad n > 0. \quad (54) \]

In (51)-(54) we also make use of the functions \( h_i(z) \) (i, j \( \geq 0 \)) and \( f_k(z) \) (k \( \geq 1 \)) defined in Appendix B, by (B.3), (B.5) and (B.6), respectively. For large n, the factors of the infinite product in (51) converge to \( G(z_0)/z_0 = 1 \). Likewise, one can prove the infinite series to be also convergent. Therefore, it is not too difficult to evaluate \( \theta^c \) numerically based on the expressions (51)-(54).

### 7.2. A lower bound \( c_\theta \) for \( c_\theta \)

We can also consider the scenario in which the packets of M always get transmitted first, i.e. the term \( \epsilon^c(J + n), n \geq 0 \) in (21) and (22) always equals its minimum value 1. Under this condition, the message delay \( c_\theta \) is given by

\[ c_\theta \geq c_\theta = \begin{cases} x(J) + \sum_{n=1}^{k-2} \epsilon(n + J) + 1, & k \geq 2, \\ x(J) - \epsilon(J) + 1, & k = 1. \end{cases} \]

In the case where \( L = k > 1 \), we can remark that \( c_\theta = \tilde{c}_k - 1 \). Hence, the pgf of \( c_\theta \) is given by \( \tilde{c}_k(z) = z \tilde{c}_k(z) \) and has the same dominant pole \( z_0 \) as \( \tilde{c}_k(z) \). Its residue in the point \( z = z_0 \) is then simply given by

\[ \theta^c = z_0 \theta^c, \quad k \geq 2, \]

so we can reuse the result of (51). In case \( L = 1 \), we find for \( c_\theta (z) \):

\[ c_\theta (z) = E[c^{(J)}]z^{-c-1} = z \tilde{P} \begin{pmatrix} 1, & 1, & 1, & z, & z, & \ldots \end{pmatrix}. \]

As such, it is possible to devise a similar algorithm for \( \theta^c \) as in (51)-(54), albeit much simpler here.

To conclude, we note that once the values \( z_0, \theta^c, \) and \( \theta^c, k \geq 1 \), are obtained, one can calculate from (46) the following upper and lower bounds for the probability of the unconditional message delay \( c \) to be larger than \( C \):

\[ \frac{\theta^c}{z_0 - 1} z_0^{-C-1} \leq P[c > C] \leq \frac{\theta^c}{z_0 - 1} z_0^{-C-1}, \]

where \( \theta^c = \sum_{n=1}^{\infty} \theta^c n! \) and \( \theta^c = \sum_{n=1}^{\infty} \theta^c n! \). In practice, these bounds prove to be reasonably tight, as will be demonstrated by the examples in the next section.
8. Numerical examples and discussion of results

In order to illustrate how $E[c]$, $E[h]$ and $Pr[c > C]$ are influenced by the parameters of the arrival process, we now consider some practical examples. First, let us introduce four possible choices for the pgf $L(z)$ of the message length:

\begin{align}
L_1(z) &= z^n, \\
L_2(z) &= \frac{(1 - \theta)\theta z}{(1 - \theta z)^2}, \\
L_3(z) &= \frac{(1 - \lambda)\lambda z}{1 - \lambda z} + \frac{(1 - \lambda)\lambda z}{1 - \lambda z}, \\
L_4(z) &= \frac{(1 - \lambda_1)\lambda_1 z}{1 - \lambda_1 z} + \frac{(1 - \lambda_2)\lambda_2 z}{1 - \lambda_2 z},
\end{align}

i.e. fixed-length messages, a negative binomial distribution, a geometric distribution and a mixture of two geometric distributions, respectively. The parameters of the distributions are chosen such that the mean message length $L(1)$ is equal to a given value $m$. Additionally, in case of $L_4(z)$, a value for $\text{Var}[L_4]$ must be specified. The variances of the other message-length distributions are given by

\begin{align}
\text{Var}[L_1] &= 0, \\
\text{Var}[L_2] &= \frac{1}{2}(m - 1)(m + 1), \\
\text{Var}[L_3] &= m(m - 1).
\end{align}

Figs. 5–7 all consist of two plots. In part (a) of these figures, the mean delay $E[c]$ and the mean transmission time $E[h]$ of the messages are plotted versus the total load $\rho = \frac{\pi M(1) + (1 - \sigma)M(1)l'(1)}{1}$ of the system. The variation of $\rho$ along the abscis is brought about only by varying $\sigma$ between 0 and a maximum value implied by the stability condition. Hence, $\rho$ can range from $M(1)l'(1)$ to 1. In part (b) of each figure, $E[c]$ and $E[h]$ are plotted versus the burst-length factor $K$, for a particular fixed choice of the load (or equivalently, of $\sigma$).

In Fig. 5, the message-length distribution is negative binomial according to $L_3(z)$, with $m = 10$. The number of new messages in the '0'-slots is assumed to have a Poisson distribution with intensity $q = 0.05$, $M_0(z) = e^{0.05} \frac{z}{z - 1}$ and $M_1(z) = z$. In (a) these measures are plotted versus the total load $\rho$ for various values of $K$, whereas in (b) they are plotted versus $K$, for different values of the load.

![Fig. 5. Mean message delay $E[c]$ and transmission time $E[h]$ in case of a negative binomial message-length distribution (mean 10), $M_0(z) = e^{0.05} \frac{z}{z - 1}$ and $M_1(z) = z$. In (a) these measures are plotted versus the total load $\rho$ for various values of $K$, whereas in (b) they are plotted versus $K$, for different values of the load.](image-url)
Fig. 6. Mean message delay $E[c]$ and transmission time $E[h]$ in case of a deterministic message length $L = m$, $M_0(z) = 1$ and $M_1(z) = z^2$, for various values of $L$. In (a) these are plotted versus the load $\rho$ for $K = 3$ and (b) versus $K$ for $\rho = 0.5$.

whereas exactly one new message is generated in each ‘1’-slot, i.e. $M_0(z) = e^{\sigma(z-1)}$ and $M_1(z) = z$. Note that in this case, the load cannot become less than $mq_0 = 0.5$ by merely varying $\sigma$. The first thing we notice on these plots, is the rapid growth of the mean message delay as the burst-length factor $K$ increases, even though the load is unchanged. Specifically, Fig. 5(b) shows that this growth is almost linear, especially for large $K$. Therefore, we can conclude that in the analysis of multiplexers like the one studied here, it is very important to take into account possible correlation in the message generation process. In fact, suppose

Fig. 7. Mean message delay $E[c]$ and transmission time $E[h]$ for various distributions of the message length $L$ (mean 10), $M_0(z) = 0.06z + 0.94$ and $M_1(z) = z$. In (a) these are plotted versus the load $\rho$ for $K = 1$, 5 and (b) versus $K$ for $\rho = 0.75$. 
we would choose to neglect this secondary correlation and analyse the system under the assumption that the numbers of messages generated in each slot are no longer correlated, but are i.i.d. variables with pgf $E(z)$. Of course, this pgf would then be: $\sigma M_1(z) + (1 - \sigma) M_0(z)$. It can be seen that the results of the analysis for this model with only primary correlation can be obtained from our model (with both primary and secondary correlation) by assuming that the correlation coefficient $\phi$ of the environment state in two consecutive slots is zero, i.e. $K = 1$. Indeed, if $K = 1$ (and $\phi = 0$), our results (40) and (45) for $E[c]$ and $E[h]$ reduce to the ones obtained in [9], where an uncorrelated message generation process was considered. The next observation we make from Fig. 5 is that the influence of the environment parameters $K$ and $\sigma$ (or $\rho$) on the mean transmission time $E[h]$ is rather limited (as opposed to the very significant influence on $E[c]$). For instance, in Fig. 5(a) we see that, for given $K$, $E[h]$ increases only slightly with the load $\rho$ and remains bounded, even if the load is maximal. Indeed, during the transmission time of a message $M$, the only packets that are transmitted, are those that arrived during the $L$ arrival slots of $M$. On the average, the amount of those packets is bounded for all $0 \leq \sigma \leq 1$, even for values of $\sigma$ for which $\rho > 1$.

Additionally, in Fig. 5(b), we observe that, for given $\rho$, $E[h]$ reaches a limit value if $K \to \infty$. Naturally, all of the above considerations are valid for Figs. 6 and 7 too.

In Fig. 6, the message length is deterministic according to $L_1(z)$, and we have plotted $E[c]$ and $E[h]$ for $m = 1, 2, 3, 4, 5$. In the ‘0’-slots, no new messages are generated, whereas in each ‘1’-slot, there are exactly 2 new messages, i.e. $M_0(z) = 1$ and $M_1(z) = z^2$. In Fig. 6(a) and (b) we have assumed $K = 3$ and $\rho = 0.5$, respectively. We observe that, for a given $K$ and $\sigma$, both $E[c]$ and $E[h]$ increase with the message length $m$, as is intuitively clear.

In Fig. 7, in each ‘1’-slot exactly one new message is generated, while in each ‘0’-slot this only happens with probability 0.06, i.e. $M_0(z) = 0.06z + 0.94$ and $M_1(z) = z$. To illustrate the influence of the distribution of the message lengths, we have plotted $E[c]$ and $E[h]$ for the four pgfs given in (55). Their parameters are chosen such that $L_1(z) = 10$, and $\text{Var}[L_1] = 150$. The other variances are then given by $\text{Var}[L_2] = 0$, $\text{Var}[L_3] = 49.5$ and $\text{Var}[L_4] = 90$. In Fig. 7(a) and (b) we chose $K = 1.5$ and $\rho = 0.75$, respectively. Although it cannot directly be proven from (40), we make the empirical observation that the mean message delay $E[c]$ appears to increase with $\text{Var}[L_1]$ (in [15,16] we proved that this is certainly the case for $E[h]$). However, for the mean transmission time $E[h]$, quite the opposite seems to hold: message-length distributions with higher variances yield a lower, hence better, mean transmission time.

Finally, in Fig. 8, we show two plots of the tail distribution of the message delay $c$, with $M_0(z) = e^{\phi L_0(z)}$ and $M_1(z) = z$. First, in Fig. 8(a), the messages are exactly 10 packets long. Both $\Pr[c > C]$ and $\Pr[c > C]$ are plotted for $K = 5$ and various values of the load $\rho$. We also used these parameters to run a simulation according to the mathematical model described in Section 3. The delay distribution of the messages generated during a $2 \times 10^5$-slot simulation for $\rho = 0.7$ is added to the figure, matching exactly the result predicted in Section 7. Secondly, in Fig. 8(b), the upper and lower bounds for $\Pr[c > C]$ are plotted for the same distributions of the message length as in Fig. 7, for $\rho = 0.8$ and $K = 1, 5$. Again, we see that the message-length distribution with the highest variance gives the highest tail probabilities for the delay. Both figures show that the upper and lower bounds for the tail distribution of $c$ which we obtained in Section 7 lie extremely close together.

As one of the reviewers noted, the influence of the parameters of the model as observed in the above examples, is in accordance with the nowadays well-established rule that buffer performance worsens as the autocorrelation function of the arrival stream extends to larger time scales. With regard to our particular model, it would be useful to be able to quantify the ‘amount’ of correlation that can be seen as primary correlation (messages arrive as trains) and the amount qualified as secondary correlation (non-independent
environment) apart from each other, as well as their respective impact on the multiplexer performance. For the secondary correlation, this not too difficult. It is clear from (3) that the autocorrelation of the environment \( \{ t_k \} \) at lag \( n \) behaves like \( \phi^n \), so a higher \( \phi \) (or \( K \)) means a higher secondary correlation. Also, the impact of \( K \) on the performance measure is clearly discussed above. A similar characterisation for the primary correlation is not so easy to provide. A possible candidate could be the mean message length \( L'(1) \), since it reveals how long an average message affects the arrival stream. Unfortunately, unlike \( K \), \( L'(1) \) also has influence on the first-order statistics of the arrival stream, which makes it difficult to use this parameter to quantify correlation.

9. Conclusions

In this paper, we have presented some further results on the queueing model studied in [15,16], i.e. a discrete-time buffer system with correlated variable-length messages arriving at the rate of one packet per slot. By means of a generating-functions approach and the use of an infinite-dimensional state description, we were able to obtain explicit formulas for the mean delay and mean transmission time of a message, as well as a tight lower and upper bound for the tail probabilities of the message delay. This allowed us to investigate the impact of the correlation in the message generation process (the secondary correlation, as we have called it). Specifically, it turns out that if this correlation is neglected in the analysis, the message delay may be severely underestimated.

Acknowledgements

The second author is a postdoctoral fellow of the Fund for Scientific Research-Flanders (Belgium) (FWO). We also wish to thank the reviewers for their useful remarks.
Appendix A. An iterative solution for the functional equation

We now derive an explicit formula for the joint pgf of system state variables, observed both on arbitrary slot boundaries \((P)\) and at the start of new messages \((\hat{P})\). To this end, we repeatedly apply functional equation (10) as follows:

\[
P(x, y_1, y_2, \ldots, z) = \frac{1}{z} M_0(x, z) b(x \mu(y_1, z)) p_0(z - 1) + P(R(x \mu(y_1, z)), D_1(y_2, z))
\]

\[
= \frac{1}{z} M_0(y_1, z) b(x \mu(y_1, z)) p_0(z - 1)
\]

\[
+ \frac{1}{z} M_0(y_1, z) b(x \mu(y_1, z)) \frac{1}{z} M_0(D_1(y_2, z)) b(R(x \mu(y_1, z)) \mu(z D_1(y_2, z)))
\]

\[
\times \left[ p_0(z - 1) + P(R(x \mu(y_1, z)) \mu(z D_1(y_2, z))), D_1(z D_1(y_2, z)), \ldots, z \right] = \ldots
\]

\[
= p_0(z - 1) \sum_{j=0}^{\infty} \left( \prod_{r=0}^{j} M_0(g_r) b(p_j) \right) + \left( \prod_{r=0}^{\infty} M_0(g_r) b(p_j) \right)
\]

\[
\times \lim_{N \to \infty} P(R(y_N), D_1(\ldots D_3(\ldots), \ldots, D_N(y_{N+2}) \ldots \ldots, D_N(y_{N+3}) \ldots \ldots, \ldots).
\]  

(A.1)

In this expression, we defined the functions \(g_n, n \geq 0\), as

\[
g_n = g_n(y_{n+1}, z) = z D_1(\ldots D_n(y_{n+1}) \ldots) = \sum_{r=1}^{n} l_r z^r + y_{n+1} z^{n+1} \sum_{r=n+1}^{\infty} l_r,
\]  

(A.2)

i.e. \(g_n\) is a polynomial of degree \(n + 1\) in \(z\), converging to \(L(z)\) as \(n\) goes to infinity. Secondly, the functions \(y_n, n \geq 0\), in (A.1) are defined recursively as

\[
y_n = y_n(x, y_1, z) = x \mu(y_1, z), \quad y_n = y_n(x, y_1, y_2, \ldots, y_{n+1}, z) = R(y_n-1) \mu(g_n), \quad n \geq 1.
\]  

(A.3)

A remarkable thing about this iterative procedure is that the arguments of \(P\) on the right-hand side of (A.1) not only become independent of all variables other than \(z\), but also seem to converge to the one-dimensional trajectory \((\chi(z), \eta_1(z), \eta_2(z), \ldots, z)\) for which (14) holds. For the variables \(y_n, n \geq 1\), this follows directly from (12). On the other hand, to prove that \(R(y_n) = \chi(z)\), we first remark that if the row \(y_n\) converges to some limiting function \(y_n\), then definition (A.3) implies that

\[
R(y_n) = R(R(y_{n-1}) \mu(L(z))), \quad R(y_n(1, 1, \ldots, 1)) = 1,
\]  

(A.4)

since \(\lim_{n \to \infty} y_n = L(z)\). Now, when comparing (13) and (A.4), we see that both \(\chi(z)\) and \(R(y_n)\) are implicitly determined by the same relations. Hence, we can conclude on their equality.

We can also give a more ‘constructive’ proof for the convergence of \(R(y_n(x, y_1, \ldots, y_{n+1}, z))\) to \(\chi(z)\), in the case where the message length is bounded by \(N\) (see Appendix C). From (A.2) and (A.3), we then have that \(y_n = R(y_{n-1}) \mu(L(z))\), for all \(n \geq N\). Therefore, the functions \(y_n, n \geq N\), depend only on the variables \((x, y_1, \ldots, y_N, z)\). Moreover, for a particular value of \(z\), the dynamic behaviour of the row \(\{y_n\}_{n \geq N}\) is determined by the following linear fractional transformation (or Moebius-transform):

\[
y_n = \left( \frac{a}{b} y_{n-1} + 1 - a \right) \left[ \frac{1 - b}{1 - \beta} y_{n-1} + \beta \mu(L(z)) \right].
\]  

(A.5)
This well-known type of transformations generally has one repelling and one attracting fixed point, the latter here being equal to \( \chi(z)\mu(L(z)) \). This means that, whatever the starting value \( \gamma_{N-1} \), eventually the row \{R(\gamma_n)\} will always converge to the value \( \chi(z) \).

Finally, following the above discussion, we can substitute the linear solution (14) of the functional equation into (A.1), which yields an explicit expression for the joint pgf of the system state variables in an arbitrary slot:

\[
P(x, y_1, y_2, \ldots, z) = p_0(z - 1) \sum_{j=0}^{+\infty} \left( \prod_{n=0}^{j-1} \frac{1}{z} M_0(g_j) b(\gamma_j) \right) \left( \frac{z}{G(z)} \right)^{j} \cdot \left( \prod_{n=0}^{j} \frac{1}{z} M_0(g_j) b(\gamma_j) \right).
\]

(A.6)

To obtain also the distribution as seen by an arbitrary new message, we need to differentiate this expression to \( y_1 \), as indicated by (19). Here it is useful to remark that \( g_n, n \geq 1, \) is independent of \( y_1 \). We obtain

\[
\hat{P}(x, y_1, y_2, \ldots, z) = \frac{y_1}{E_1(1)} p_0(z - 1) \sum_{j=0}^{+\infty} \left( \prod_{n=0}^{j-1} \frac{1}{z} M_0(g_j) b(\gamma_j) \right) \left( \frac{z}{G(z)} \right)^{j} \cdot \left( \prod_{n=0}^{j} \frac{1}{z} M_0(g_j) b(\gamma_j) \right) \left( \frac{1}{G(z)} \right)^{+\infty} \sum_{n=0}^{j} 1 - \beta \frac{b(\gamma_j)}{b(\gamma_1)} \frac{\partial}{\partial y_1} \gamma_n.
\]

(A.7)

Note that since \( g_n \to L(z) \) and \( \gamma_n \to \chi(z)\mu(L(z)) \) as \( n \) goes to infinity, the factors of the infinite product in both (A.6) and (A.7) converge to \( G(z)/z \).

Appendix B. The upper bound \( \bar{c} \)

In this appendix, we elaborate on the pgf \( \bar{c} \) = \( \Delta_0(z, z, \ldots, z, z) \) of \( \bar{c} \), an upper bound for the total delay experienced by a message of length \( L = k \). From definition (48), we have that

\[
\Delta_k(z, z, \ldots, z, z, z) = E[x^{(J+1)} z^{(J+2)} \ldots z^{(J+k-1)} x^{(J)}].
\]

(B.1)

The first \( k - 1 \) factors in (B.1) can be expanded using \( e(J+n) = \sum_{i=n}^{+\infty} m_i(J + n), 1 \leq n \leq k - 1 \). Additionally, based on system equation (33), one can derive the following relations concerning the expansion of the variables \( m_i(J + n), i \geq 1 \):

\[
E[x^{(J+n)}] = \begin{cases} 
E[(D_1(D_2(D_3(\ldots D_{n}(z) \ldots)))^{m_{i}(J+n-1)}), & 1 < i < n + 1, \\
E[(z(D_1(D_2(D_3(\ldots D_{n-1}(z) \ldots)))^{m_{i}(J-1)}), & i = n + 1, \\
E[(D_{i-n}(D_{i-n+1}(\ldots D_{n-1}(z) \ldots)))^{m_{i-1}(J)}), & i > n + 1.
\end{cases}
\]
After using these in (B.1) and collecting all factors with the same exponent, we find

\[
\Delta_k(z, \ldots, z) = E[z^m_1 \cdot (zh_1)^{m_1} \cdot (zh_2)^{m_2} \cdot \ldots \cdot (zh_{J+1})^{m_{J+1}}]
\]

\[
= \sum_{i=0}^{n} \left[ \frac{(\sigma+1-\sigma \phi^i)}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right) \right]
\]

\[= \frac{1}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

where we have defined the functions \(\mu_1\), \(\mu_2\), and where \(\mu_1\) can also be proved by induction. We then finally find:

\[(28):
\]

\[
E[-\Delta_k(z, \ldots, z)] = E[\sum_{i=0}^{n} \left[ \frac{(\sigma+1-\sigma \phi^i)}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right) \right]
\]

\[
= \frac{1}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

\[(B.2)\]

where we have defined the functions \(h_{ij}, i, j > 0\), as

\[
h_{ij} = D_i(z)D_j(z) \cdots D_i(D_j(\ldots(D_i(D_j(z))))\ldots)
\]

\[
= \prod_{r=0}^{i-1} D_i(D_j(\ldots(D_i(D_j(z))))\ldots)),
\]

\[(B.3)\]

and where \(h_{ij} = 1\) if either of the indices is zero, by convention. Next, we can use (31) to relate the number of new messages \(m_1(J+n)\) generated in the slots \(J+1, \ldots, J+k-1\) to the environment state \(t(J+n)\) in those slots. Also, if we let \(R^n(z)\) denote \(n\) consecutive applications of \(R(z)\), we find from (28):

\[
E[z^{t(J+n)}] = E\left[ \left( \prod_{r=0}^{n-1} b(R^t(z)) \right)(\frac{R^n(z)}{t(J)})^{\phi_r} \right] = E[\sum_{i=0}^{n} \left[ \frac{(\sigma+1-\sigma \phi^i)}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right) \right]
\]

\[
= \frac{1}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

\[(B.4)\]

The only remaining stochastic quantities in (B.4) are the system state variables for slot \(J\), of which the joint pgf is given by \(\hat{P}(z, y_1, y_2, \ldots, z)\). Let us further define

\[
A_n(z) = v_n M_1(z) + (1 - v_n) M_0(z), \quad n \geq 1,
\]

\[(B.5)\]

and

\[
f_k(z) = \prod_{r=0}^{n-1} F(R^{t(J+n)}(z)) = \prod_{r=0}^{n-1} \frac{(\sigma+1-\sigma \phi^i)}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

\[
= \frac{1}{\mu(zh_{i+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

\[(B.6)\]

which can also be proved by induction. We then finally find:

\[
\Delta_k(z, \ldots, z) = A_k(z)A_{k-2}(zh_1)A_{k-3}(zh_2) \cdots A_1(z)zh_{k-2}
\]

\[
\times \frac{1}{\mu(zh_{k+1})} \cdot \left( \frac{(1-\sigma)(1-\phi^i)}{\mu(zh_{i+1})} \right)
\]

\[(B.7)\]
Now, one can obtain an explicit expression for $\Delta_k(z, z, \ldots, z, z)$ by simply substituting result (A.7) of Appendix A.

Appendix C. Finite-length messages

We now point out some useful simplifications to our analysis in the particular case that the length of the messages cannot exceed a certain bound $N \geq 1$, as may sometimes be assumed in practical situations. In fact, all it takes to incorporate this observation is to substitute $l_r = 0$ for all $r > N$ into the results of the above analysis and no further problems will arise. For instance, the infinite sums in expressions (40) and (44) now reduce to finite sums and are therefore much easier to evaluate.

Furthermore, if it is known from the start that the messages are bounded, the analysis no longer requires the use of an infinite-dimensional description of the system state. Indeed, if no message can contain more than $N$ packets, it is seen that $m_{n,k}$ must be 0 for all $n > N$. Therefore, the set of system state variables for slot $k$, as defined at the end of Section 3, reduces to $t_{1,k}, m_{1,k}, \ldots, m_{N,k}, s_{k+1}$, so no more than $N + 2$ random variables are required. As a consequence, the pgf $P$ (and also $\hat{P}$) of the system state has only $N + 2$ arguments and the main functional equation (10) becomes

$$P(x, y_1, y_2, \ldots, y_N, z) = E[x^y_1 y_2 \cdots y_N z]$$

$$= \frac{1}{z} \left[ M_0(y_1 z) b(x \mu(y_1 z), D_1(y_1 z), D_2(y_2 z), \ldots, D_{N-1}(y_N z), 1, z) + p_0(z - 1) \right],$$

since, from (7) and (12), $D_n(z) = 1$ and $\eta_n(z) = 1$ for $n \geq N$. All further calculations change accordingly. Also, the iterative procedure of Appendix A to obtain an explicit expression for $P$ and $\hat{P}$ still holds. Contrary to what one might expect, an infinite number of iterations are still required, even for bounded message lengths. For $n \geq N$, the functions $g_n$ in (A.2) become equal to $L(z)$ and no longer vary, but the functions $\gamma_n$ keep changing according to (A.5). Thus, we still need infinite products in the results (A.6) and (A.7).

Concerning $E[c]$ and $E[h]$, Eqs. (20) and (41) reveal that it might be a good idea to concentrate mainly on the calculation of those $E[c_k]$ and $E[h_k]$ for which $l_k$ is large (separate expressions for these conditional means can easily be derived from our discussion in Sections 5 and 6). Anyhow, for bounded message lengths, no more than $N$ of these terms have to be considered. Likewise, when evaluating the bounds $\bar{c}$ and $c$ to assess the tail distribution of the message delay, only those residues $\theta_k^n$ have to be determined for which $k \leq N$. Note that the algorithm (51)-(54) can only begin to converge beyond $N$ iterations.

References


S. De Vuyst was born in Eeklo, Belgium, in 1973. He obtained his M.S. degree in Electrical Engineering from Ghent University, Belgium, in 1997. Since October 1998, he is a member of the SMACS Research Group, Department of Telecommunications and Information Processing, also at Ghent University. His main interest involves the study of discrete-time queuing models.

S. Wittevorgel was born in Gent, Belgium, in 1969. She received her M.S. degree in electrical engineering and Ph.D. degree in applied sciences from Ghent University, Belgium, in 1992 and 1998, respectively. Since September 1992, she has been with the SMACS Research Group, Department of Telecommunications and Information Processing, Ghent University, first in the framework of various projects, and since October 1994, as a researcher of the Fund for Scientific Research-Flanders (Belgium) (FWO). Her main research interests include discrete-time queuing theory, the performance evaluation of IP networks, the study of traffic control mechanisms, and the analysis of ARQ protocols.
H. Bruneel was born in Zottegem, Belgium, in 1954. He received his M.S. degree in electrical engineering, the degree of Licentiate in computer science, and Ph.D. degree in computer science in 1978, 1979 and 1984, respectively, all from Ghent University, Belgium. From 1979 to 1998, he has been a researcher of the Fund for Scientific Research Flanders (Belgium) (FWO) at Ghent University. He has also been a part time Professor in the Faculty of Applied Sciences at the same university from 1987 to 1998. Currently, he is a full time Professor and the head of the Department of Telecommunications and Information Processing. He also leads the SMACS Research Group within this department. His main research interests include stochastic Modeling of digital communication systems, discrete-time queuing theory, and the study of ARQ protocols. He has published more than 150 papers on these subjects and is coauthor of the book H. Bruneel and B.G. Kim, “Discrete-Time Models for Communication Systems Including ATM” (Kluwer Academic Publishers, Boston, 1993).