Discrete-time queues with correlated arrivals and constant service times

Sabine Wittevrongel*1, Herwig Bruneel2

Stochastic Modeling and Analysis of Communication Systems Research Group, Laboratory for Communications Engineering, University of Ghent, Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

Received June 1997; received in revised form June 1998

Scope and purpose

Discrete-time queueing models are suitable for the performance evaluation of Asynchronous Transfer Mode (ATM) multiplexers and switches. In these models, the time axis is divided into fixed-length slots and the service of a customer must start and end at slot boundaries. Most analytical studies of discrete-time queues assume constant service times equal to one slot, an infinite buffer capacity and/or uncorrelated arrival process. The present paper is an attempt to explore whether analytical techniques are still applicable in absence of these restricting assumptions. Specifically, we focus on a discrete-time single-server queueing system with constant service times of arbitrary length, a finite storage capacity and a simple non-independent arrival model. The analysis method is based on an extensive use of generating functions, an approach which has traditionally been reserved for infinite-capacity queues. We show in this paper that generating functions can also be very useful in the finite-capacity case.

Abstract

A discrete-time single-server finite-capacity queue with correlated arrivals and constant service times of arbitrary length is investigated in this paper. Cells are generated by a bursty on/off source, with geometrically...
distributed lengths of the on-periods and the off-periods. The performance of the system is evaluated by means of an analytical technique, based on generating functions, whose computational complexity does not depend on the buffer space. As a result of the analysis, closed-form expressions are obtained for the cell loss ratio, the steady-state probability generating functions of the queue length, the unfinished work and the cell delay and the joint probability generating function of two consecutive interdeparture times at the output of the queue. Some numerical examples illustrate the results. © 1998 Elsevier Science Ltd. All rights reserved.

1. Introduction

Discrete-time queueing models have often been used in the performance analysis of ATM multiplexers and switches [1]. In these models the time axis is assumed to be divided into fixed-length time intervals, referred to as slots, and the servicing of cells is synchronous, i.e. the service of a cell can start or end at slot boundaries only. This implies that the service time of a cell always consists of an integer number of full slots so that it can be considered as a discrete random variable.

Most performance studies of ATM queues assume constant service times equal to one slot. Discrete-time queueing models with constant service times of arbitrary length have been studied in [2, 3]. In [2], a discrete-time single-server queue with constant service times and Bernoulli arrivals is studied. Both the case of an infinite and a finite buffer size are considered. Closed-form expressions are obtained for the steady-state distributions of the queue length and the unfinished work in the system. In [3], a discrete-time multiserver queueing model with constant service times and a general uncorrelated arrival process is studied. A general analysis of the queue-length distribution is presented, under the assumption of an infinite buffer capacity. Infinite-capacity queues with arbitrarily distributed service times and general independent arrivals are investigated in [4, 5]. The performance measures studied in these papers are the unfinished work the cell delay and the buffer occupancy at the start of an arbitrary slot. In [5], the queue length at various observation epochs and the busy and idle periods of the system are studied.

The models considered in [2–5] assume that the cells are generated according to an uncorrelated arrival process. It has been observed however, that in practice, ATM cell arrival streams tend to be correlated. Therefore, in this paper a non-independent cell arrival process is considered. Specifically, we present a method for analyzing the buffer behavior of a finite-capacity ATM queue with first-order Markovian arrivals and constant service times. The queueing system is mainly investigated by analytical means. The analysis method is based on an extensive use of generating functions, an approach which has traditionally been reserved for infinite-capacity case. The performance results for an infinite-capacity queue can also be found from our analysis as a special case. Besides the standard performance quantities such as the queue length, the unfinished work and the cell delay, we investigate in this paper the characteristics of the traffic at the output of the queue.

The outline of the paper is as follows. In Section 2, the queueing model under study is described. Section 3 introduces a three-dimensional Markovian state description for the system and gives the system equations describing the buffer behavior. In Section 4, closed-form expressions are derived
for the cell loss ratio and the steady-state probability generating function (pgf) of the buffer occupancy. The pgf of the unfinished work in the system is derived in Section 5. Section 6 concentrates on the steady-state pgf of the cell delay. The characteristics of the output process are studied in Section 7, and finally, in Section 8, some numerical results are presented.

2. The queueing model

We consider a discrete-time queueing system, i.e. a system in which the time axis is divided into fixed-length time intervals, referred to as slots. The queueing system consists of one single server and a finite waiting room for cells of size $S$ ($S > 0$) (Fig. 1). It should be noted that the total number of cells present in the buffer can hence be at most $S + 1$. Cells arrive at the queueing system in a stochastic manner and are stored in the finite waiting room if the latter is not full; otherwise, they are lost. In the waiting room, the cells wait for some time, until finally, they receive service from the server of the system. It is assumed that each cell requires a constant service time equal to $M$ slots, where $M > 1$. The service time of a cell must start and end at slot boundaries, which implies that the service of a cell which arrives in an empty buffer cannot start before the end of the cell’s arrival slot.

Cells are generated according to a first-order Markovian arrival process, i.e. the number of cell arrivals during an arbitrary slot is a random variable dependent on the number of arrivals during the immediately preceding time slot. During a slot, there can be either one cell arrival or no cell arrival. The cell arrival process is then characterized by two independent parameters $\alpha$ and $\beta$, defined as

$$\alpha = \text{Prob}[\text{one arrival during a slot} | \text{one arrival during previous slot}],$$

$$\beta = \text{Prob}[\text{no arrival during a slot} | \text{no arrival during previous slot}].$$

This means that the arrival process alternates between on-periods, during which exactly one cell is generated per slot, and off-periods, during which no cells are generated. Both the lengths of the on-periods and the lengths of the off-periods are geometrically distributed random variables with mean values $1/(1 - \alpha)$ and $1/(1 - \beta)$, respectively. The steady-state probability $\rho$ of having a cell arrival during an arbitrary slot is then easily seen to be given by

$$\rho = (1 - \beta)/(2 - \alpha - \beta).$$

![Fig. 1. Single-server finite-capacity queueing system.](image-url)
3. System equations and definitions

Let us define the random variable $s_k$ as the buffer occupancy, i.e. the total number of cells stored in the buffer including the possible cell in service, at the beginning of slot $k$. Also, let $a_k$ indicate the number of cells entering the buffer during slot $k$. Furthermore, let us introduce the random variable $r_k$ as follows: $r_k$ denotes the number of slots service received by the cell currently in service at the beginning of slot $k$, if $s_k > 0$, and $r_k = 0$, if $s_k = 0$. This random variable will be referred to in this paper as the “delivered service time” at the start of slot $k$. Note that because of the constant nature of the service times, $r_k$ can be at most $M! - 1$ slots. The following set of system equations can then be established:

- If $s_k > 0$ and $r_k \neq M-1$:
  \[ s_{k+1} = \min(s_k + a_k, S + 1), \]  
  \[ r_{k+1} = r_k + 1. \]  
  \[ (4) \]
  \[ (5) \]

- If $s_k > 0$ and $r_k = M - 1$:
  \[ s_{k+1} = \min(s_k + a_k, S + 1) - 1, \]  
  \[ r_{k+1} = 0. \]  
  \[ (6) \]
  \[ (7) \]

- If $s_k = 0$ (and, hence, $r_k = 0$):
  \[ s_{k+1} = a_k, \]  
  \[ r_{k+1} = 0. \]  
  \[ (8) \]
  \[ (9) \]

These system equations are based on the following observations. First, if at the beginning of slot $k$, the system is non-empty and the delivered service time is less than $M - 1$ slots, there is no departure at the end of slot $k$, the buffer occupancy is augmented with the number of effective cell arrivals during slot $k$ and the number of slots service received by the customer being served is augmented by one. If, on the other hand, the cell in service has already received $M - 1$ slots of service at the start of slot $k$, this cell will leave the buffer at the end of slot $k$. The system becomes empty if the cell in service was the last one in the system and no cells have arrived during slot $k$, while a new cell is taken in service in the opposite case. In both cases, the random variable $r_{k+1} = 0$.

Finally, if the system is empty at the beginning of slot $k$, the buffer occupancy at the start of slot $k + 1$ equals the number of cell arrivals during slot $k$, and the random variable $r_{k+1} = 0$.

From the system Eqs. (4)–(9), it is easy to see that the vector $(r_k, a_{k-1}, s_k)$ constitutes a three-dimensional Markovian state description of the system at the beginning of slot $k$. Let $p(i, n, j)$ indicate the equilibrium probabilities of the Markov chain $(r_k, a_{k-1}, s_k)$, i.e.

\[ p(i, n, j) \triangleq \lim_{k \to \infty} \text{Prob}[r_k = i, a_{k-1} = n, s_k = j], \]  

for all $0 \leq i \leq M - 1$, $n = 0, 1$ and $0 \leq j \leq S + 1$. We define the partial pgf’s $Q_{i,n}(z)$ as

\[ Q_{i,n}(z) \triangleq \sum_{j=0}^{S+1} p(i, n, j) z^j, \quad 0 \leq i \leq M - 1, \quad n = 0, 1, \]  

(11)
and the functions $R_n(y, z)$ as
\[
R_n(y, z) \triangleq \sum_{i=0}^{M-1} Q_{i,n}(z)y^i. \tag{12}
\]

A method to calculate the functions $Q_{0,n}(z)$ and $R_n(y, z)$ is given in the Appendix. In the next sections, these functions will be used to analyze the buffer contents, the unfinished work, the cell delay and the output process of the considered queueing system.

4. Buffer contents

In this section, we determine the pgf $S(z)$ of the steady-state buffer contents $s$ at the beginning of an arbitrary slot. Using the definitions and notations, introduced above, it is easily seen that $S(z)$ can be derived as
\[
S(z) = \sum_{j=0}^{S+1} \text{Prob}[s = j]z^j = \sum_{i=0}^{M-1} \sum_{n=0}^{1} \left( \sum_{j=0}^{S+1} p(i, n, j)z^j \right) = \sum_{i=0}^{M-1} \sum_{n=0}^{1} Q_{i,n}(z) = R_0(1, z) + R_1(1, z).
\]

Next, using Eqs. (A.11)–(A.18), we find the following explicit expression for $S(z)$:
\[
S(z) = \frac{p_0(z-1)}{N(z)} \left\{ z \left( \frac{\lambda_1^{M+1} - \lambda_2^{M+1}}{\lambda_1 - \lambda_2} \right) - z\lambda_1\lambda_2 \left( \frac{\lambda_1^{M} - \lambda_2^{M}}{\lambda_1 - \lambda_2} \right) - (\lambda_1\lambda_2)^M \right\} \\
+ \frac{(z-1)z^{S+1}M-1}{\rho N(z)} \sum_{i=1}^{M-1} p_i \left\{ \lambda_1\lambda_2 \left[ \rho + (1-\rho)z \right] \left( \frac{\lambda_1^{M-i-1} - \lambda_2^{M-i-1}}{\lambda_1 - \lambda_2} \right) \\
- z \left( \frac{\lambda_1^{M-i} - \lambda_2^{M-i}}{\lambda_1 - \lambda_2} \right) - (\lambda_1\lambda_2)^M \left[ \frac{\lambda_1^{i} - \lambda_2^{i}}{\lambda_1 - \lambda_2} \right] \right\} \\
+ \frac{1}{\rho} \sum_{i=1}^{M-1} p_i z^{S+1}
\]

in terms of the eigenvalues $\lambda_1(z)$ and $\lambda_2(z)$ of the matrix $F$, where
\[
F \triangleq \begin{pmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{pmatrix}, \tag{14}
\]
and the $M$ unknown parameters $p_i$ ($0 \leq i \leq M-1$). Here $p_0$ is the steady-state probability of having an empty buffer at the beginning of an arbitrary slot and the parameters $p_i$ are defined as
\[
p_i \triangleq \lim_{k \to \infty} \text{Prob}[r_k = i, a_k = 1, s_k = S + 1], \quad 1 \leq i \leq M - 1. \tag{15}
\]
The function $N(z)$ in expression (13) for $S(z)$ is given by Eq. (A.12).
In order to determine the remaining unknown probabilities \( p_i (0 \leq i \leq M - 1) \), we proceed as follows. In the Appendix, it is shown that both the denominator and the numerator in \( Q_{0,n}(z) \) are polynomials in \( z \), and vanish for \( z = 0 \) and for \( z = 1 \). The denominator is a polynomial of degree \( M + 1 \), and therefore has exactly \( M + 1 \) zeros inside the complex \( z \)-plane. The key step is now to realize that the partial pgf’s \( Q_{0,n}(z) \) are polynomials in \( z \), and therefore are analytic function of \( z \) in the whole complex plane. Each of the zeros of the denominator \( Q_{0,n}(z) \) must therefore be a zero of the numerator of \( Q_{0,n}(z) \) as well, so that \( M - 1 \) linear equations in the \( M \) unknowns are obtained (no equation is obtained for the zeros \( z = 0 \) and \( z = 1 \)). It is worth noting that this reasoning is somewhat different from the one used in classical queueing analysis of infinite-capacity queueing systems, where remaining unknown probabilities in a generating function are determined by invoking the analyticity property of a pgf inside the unit circle of the complex \( z \)-plane, which in general only implies that the zeros of the denominator inside the unit circle must also be zeros of the numerator. The normalizing equation \( S(1) = 1 \) leads to an \( M \)th linear equation in the unknowns, namely

\[
p_0 - M \sum_{i=1}^{M-1} p_i = 1 - M\rho. \tag{16}
\]

Hence, we have a set of \( M \) independent linear equations which can be solved to yield the values for \( p_0 \) and \( p_i (1 \leq i \leq M - 1) \). These can be substituted into Eq. (13) to obtain the pgf \( S(z) \) of the buffer contents in terms of known quantities only.

All the important characteristics of the buffer contents can be derived from Eq. (13). The mean buffer occupancy \( E[s] \) at the start of an arbitrary slot can be found from the formula \( E[s] = S'(1) \), yielding

\[
E[s] = \frac{1}{(1 - M\rho)} \left\{ \frac{\rho}{2} ((M + 1)p_0 - M + 1) + \frac{1 - \rho - p_0}{2 - \alpha - \beta} - \sum_{i=1}^{M-1} p_i [M(S + 1) - i - 1] \right\}. \tag{17}
\]

The probability mass function (pmf) \( s(j) \) of \( s \) can be obtained by taking the inverse \( z \)-transform of Eq. (13). Since \( S(z) \) is a polynomial in \( z \), inverting \( S(z) \) can easily be done by division of the numerator and the denominator polynomials in Eq. (13). The cell loss ratio (CLR) is defined as the fraction of the arriving cells that cannot enter the queue because of buffer overflow. It is clear that a cell will be lost if it is generated during a slot at the start of which the queueing system is full. This observation leads to the following expression for the CLR:

\[
CLR = \frac{1}{\rho} \sum_{i=1}^{M-1} p_i. \tag{18}
\]

It is very interesting to see that the CLR is known, once the unknowns \( p_i \) are determined. This implies that the CLR can be calculated by solving a set of linear equations, whose dimension is independent of the size \( S \) of the waiting room.

Finally, let us note that the performance results for an infinite-capacity queueing system can easily be found from our analysis as a special case. This is achieved by setting \( p_i = 0, 1 \leq i \leq M - 1 \), and \( p_0 = 1 - M\rho \) in the expressions for the finite-capacity case in order to obtain the corresponding results for the infinite-capacity case.
5. Unfinished work

The aim of this section is to obtain the pgf of the unfinished work of the queueing system at the beginning of an arbitrary slot, which is defined as the remaining number of slots needed to serve all the cells present in the system at the start of the slot. Let the random variable \( w_k \) denote the unfinished work of the system at the start of slot \( k \). Then \( w_k \) is given by

\[
w_k = \begin{cases} 
(s_k - 1)M + M - r_k & \text{if } s_k > 0, \\
0 & \text{if } s_k = 0,
\end{cases}
\]

or, equivalently,

\[
w_k = Ms_k - r_k.
\] (19)

Here \( r_k \) and \( s_k \) are two components of the three-dimensional state vector \((r_k, a_k - 1, s_k)\), introduced in Section 3. Consequently, the pgf \( W(z) \) of the unfinished work \( w \) at the beginning of an arbitrary slot in the steady state can be expressed in terms of the functions \( R_0(y, z) \) and \( R_1(y, z) \), defined in Eq. (12), as follows:

\[
W(z) = E[z^w] = \lim_{k \to \infty} E[z^{M s_k - r_k}] = R_0(z^{-1}, z^M) + R_1(z^{-1}, z^M).
\]

Using result (A.18) for \( R_0(y, z) \) and \( R_1(y, z) \), we finally obtain the following closed-form expression for \( W(z) \):

\[
W(z) = \frac{(z - 1)p_0[\beta + (1 - \alpha - \beta)z^{M-1} + (1 - \beta)z^M] + \sum_{i=1}^{M-1} p_i z^{M(S+1)-i-1}(1 - z^M)(z + 1 - \alpha - \beta)}{z - (\beta + \alpha z^M) + (\alpha + \beta - 1)z^{M-1}}.
\] (20)

6. Cell delay

The next step is to compute the statistics of the cell delay. The delay of an arbitrary cell is defined as the total number of slots between the end of the arrival slot of the cell and the departure instant of the cell. Since a cell can only leave the buffer at slot boundaries, the cell delay always consists of an integer number of slots and can hence be described as a discrete random variable. It is assumed here that cells are served according to a first-come-first-served (FCFS) queueing discipline, i.e. cells are transmitted from the buffer in their order of arrival. It may also be noticed that the distribution of the cell delay can only be determined for the fraction of the cells that arrive at and effectively enter the buffer.

Let \( C \) indicate an arbitrary cell, which is generated and effectively enters the buffer during some slot \( J \) in the steady state. Let the random variable \( w^* \) denote the unfinished work at the beginning of slot \( J + 1 \). Owing to the FCFS-queueing rule, the delay \( d \) of the cell \( C \) is equal to the time required for the transmission of all those cells that arrive in the buffer no later than the cell \( C \), except for those that have left the buffer by the end of slot \( J \). In view of the assumption that during a slot, at most one cell can arrive in the queue, it is clear that the delay \( d \) is given by

\[
d = w^*.
\] (21)
Now since the number of cell arrivals in the buffer during a slot is either 0 or 1, selecting an arbitrary effectively arriving cell is equivalent with selecting an arbitrary slot with one cell arrival and a non-full buffer at the start of this slot. This implies that the unfinished work \( w^* \) just after the arrival slot of an arbitrary cell has the same statistics as the unfinished work just after an arbitrary slot with one cell arrival and a non-full buffer. The pgf \( D(z) \) of the random variable \( d \) can hence be obtained as

\[
D(z) = \lim_{k \to \infty} E[z^{w^*} | a_{k-1} = 1, s_{k-1} \neq S + 1].
\]

Making use of the relationship \( w_k = M s_k - r_k \), we can transform the above equation into

\[
D(z) = \frac{R_1(z^{-1}, z^M) - \sum_{i=1}^{M-1} p_i z^{M(S+1)-i-1}}{\rho(1 - \text{CLR})},
\]

or, using Eq. (A.18),

\[
D(z) = \frac{(z - 1)(p_0(1 - \beta)z^M + \sum_{i=1}^{M-1} p_i z^{M(S+2)-i-2}(\alpha + \beta - 1 - z)^i)}{\rho(1 - \text{CLR})[z - \beta - \alpha z^M + (\alpha + \beta - 1)z^{M-1}]}.
\] (22)

The mean cell delay can then be found as \( E[d] = D'(1) \). It has been verified that \( E[d] = E[s]/(\rho(1 - \text{CLR})) \), in accordance with Little’s result. The variance of the cell delay is obtained as

\[
\text{var}[d] = \frac{1}{3(1 - \text{CLR})(1 - \beta)(1 - M\rho)} \left\{ \frac{1}{M} (1 - p_0)(M - 1)(M - 2) [(1 - \beta)M + 3(\alpha + \beta - 1)] + 3p_0(1 - \beta)M(M - 1) + 3 \sum_{i=1}^{M-1} p_i [M(S + 2) - i - 2][(M(S + 2) - i - 2)] + E[d] \left\{ \frac{\rho(M - 1)[(1 - \beta)M + 2(\alpha + \beta - 1)]}{(1 - \beta)(1 - M\rho)} \right. \\
+ \left. 1 \right\} - (E[d])^2.
\] (23)

The pmf \( d(j) \) of the delay \( d \) can be easily derived from Eq. (22) by inverse \( z \)-transformation.

7. Interdeparture times

We are concerned with studying the characteristics of the output process from the queue. Specifically, the purpose of this section is to derive the joint pgf \( I(x, y) \) of two consecutive interdeparture times (IDTs). In order to do so, we refine our state description of the system by introducing so-called “departure states”. We define that the system is in departure state \((i, n, j)^d\) at the beginning of the current slot if the delivered service time at the start of the current slot is equal to \( i \), the number of cell arrivals during the previous slot equals \( n \), \( j \) cells are stored in the system at
the start of the current slot and one cell has left the system at the end of the previous slot. The probabilities \( p^d(i, n, j) \) of finding the system in departure state \((i, n, j)^d\) can be derived as follows. Firstly, we observe that a departure at the end of a slot implies that the delivered service time at the start of the next slot must necessarily be zero, so that

\[
p^d(0, n, j) = 0, \quad i \neq 0. \tag{24}
\]

If at the beginning of the current slot, the buffer contents is at least equal to two cells and the service time of one of those cells starts, it is sure that a cell has left the buffer at the end of the previous slot. Similarly, if at the start of the current slot, exactly one cell is present in the queue, the service time of this cell starts and furthermore this cell did not arrive in the buffer during the previous slot, it is certain that a departure took place at the end of the previous slot. This implies that

\[
p^d(0, n, j) = p(0, n, j), \quad j \geq 2, \tag{25}
\]

\[
p^d(0, 0, 1) = p(0, 0, 1). \tag{26}
\]

Further, if the buffer contents at the start of the current slot is equal to one, the amount of service received by this cell is zero and a cell has arrived during the previous slot, a cell has left the buffer at the end of the previous slot if and only if the buffer was not empty at the start of the previous slot. Therefore, we have

\[
p^d(0, 1, 1) = p(0, 1, 1) - (1 - \beta)p(0, 0, 0). \tag{27}
\]

A cell can never leave the buffer at the end of its arrival slot; so

\[
p^d(0, 1, 0) = 0. \tag{28}
\]

Finally, if the buffer is empty at the start of the current slot, a cell has left the buffer at the end of the previous slot except if the buffer was already empty at the start of the previous slot. Hence

\[
p^d(0, 0, 0) = (1 - \beta)p(0, 0, 0). \tag{29}
\]

From the above equations, it is easily seen that the probabilities \( p^d(i, n, j) \) are known once the probabilities \( p(0, n, j) \) are calculated. The latter can be done by taking the inverse \( z \)-transform of \( Q_{0,n}(z) \), given in Eq. (A.11).

The next step in the derivation of the pgf \( I(x, y) \) is to derive for each of the departure states, the conditional joint pgf of the next two IDTs. For departure states \((0, n, j)^d, j \geq 2\), at least 2 cells are present in the buffer immediately after the last departure. This implies that the next two IDTs are both equal to the constant service time \( M \). The corresponding pgf is then given by

\[
I_1(x, y) = x^M y^M. \tag{30}
\]

Secondly, for departure states \((0, n, 1)^d\), exactly one cell is present in the buffer immediately after the last departure. Consequently, the first IDT is equal to \( M \). The second IDT equals \( M \) if at least one cell arrives during the first \( M \) slots, and equals \( g, g > M \), if no cells arrive during the first \((g - 1)\) slots and one cell arrives during slot \( g \). For departure state \((0, 0, 1)^d\), this leads to the following pgf:

\[
I_2(x, y) = (1 - \beta^M)x^M y^M + (1 - \beta)\beta^M x^M y^{M+1} \frac{1}{1 - \beta y}, \tag{31}
\]
whereas for departure state \((0, 1, 1)^d\), we obtain
\[
I_3(x, y) = [1 - (1 - x)\beta^{M-1}] x^M y^M + (1 - x)(1 - \beta) \beta^{M-1} \frac{x^{M+1} y^{M+1}}{1 - \beta y}.
\]
Finally, for departure state \((0, 0, 0)^d\), we observe that both the first and the second IDT are determined by the arrival process of cells to the queue. Specifically, the first IDT equals \(f, f > M\), if no cells arrive during the first \((f - M - 1)\) slots and one cell arrives during slot \(f - M\). Given that the first IDT is equal to \(f\), the second IDT equals \(M\) if at least one cell arrives during the slots \(f - M + 1\) to \(f\), and equals \(g, g > M\), if no cells arrive during the slots \(f - M + 1\) to \(f - M + g - 1\) and one cell arrives during slot \(f - M + g\). Therefore, the pgf \(I_4(x, y)\), corresponding to the departure state \((0, 0, 0)^d\), is obtained as
\[
I_4(x, y) = (1 - \beta) \frac{x^{M+1}}{(1 - \beta x)} \left\{ [1 - (1 - x)\beta^{M-1}] y^M + (1 - x)(1 - \beta) \beta^{M-1} \frac{y^{M+1}}{1 - \beta y} \right\}.
\]
The final step is then to average over all the possible departure states, which leads to the following expression for the joint pgf \(I(x, y)\) of two consecutive IDTs:
\[
I(x, y) = \frac{1}{b} \left\{ \sum_{n=0}^{S} \sum_{j=0}^{S} p^d(0, n, j) I_1(x, y) + p^d(0, 0, 1)I_2(x, y) + p^d(0, 1, 1)I_3(x, y) + p^d(0, 0, 0)I_4(x, y) \right\}.
\]
Here \(b\) is the steady-state probability that a cell leaves the buffer at the end of a slot, i.e.
\[
b = \sum_{n=0}^{S} \sum_{j=0}^{S} p^d(0, n, j).
\]

From Eq. (34), the various moments of the two-dimensional interdeparture-time distribution can be obtained by evaluating the consecutive mixed partial derivatives of \(I(x, y)\) with respect to \(x\) and \(y\), for \(x = y = 1\). In the next section, we will present numerical results for the variance of the IDTs and the coefficient of correlation \(\text{corr}(i_1, i_2)\) between two consecutive IDTs \(i_1\) and \(i_2\), i.e.
\[
\text{corr}(i_1, i_2) = \frac{E[i_1 i_2] - E[i_1]E[i_2]}{(\text{var}[i_1]\text{var}[i_2])^{1/2}}.
\]

8. Numerical results

Eq. (3) shows that different choices for \(\alpha\) and \(\beta\) may lead to the same value of the load \(\rho\). It has been observed before (see e.g. [6]) that even for a constant value of \(\rho\), the buffer behavior is strongly influenced by the actual values of \(\alpha\) and \(\beta\). In order to investigate the influence of the parameters \(\alpha\) and \(\beta\), as in [6], we define the burstiness factor \(K\) of the source as
\[
K \triangleq \frac{1 - \rho}{1 - \alpha} = \frac{\rho}{1 - \beta}.
\]
Here $1/(1 - \alpha)$ and $1/(1 - \beta)$ are the mean lengths of the on-periods and the off-periods of the source. It is clear from Eq. (36) that the loads are a measure for the ratio of the mean lengths of the on-periods and the off-periods, whereas the parameter $K$ is a representative for the absolute lengths of these periods. For a constant value of $\rho$, the correlation in the cell arrival process is described by the value of $K$. The case of an uncorrelated Bernoulli arrival process is found for $K = 1$. In this section, we will characterize the cell arrival process to the queue by means of the parameters $\rho$ and $K$ instead of $\alpha$ and $\beta$.

Fig. 2. Minimum size $S$ of the waiting room required to have a cell loss ratio less than $10^{-9}$ versus the burstiness factor $K$, for $\rho = 0.13$ and $M = 2, 4, 6$.

Fig. 3. Variance of cell delay versus the load $\rho$, for $K = 3$, $S = 15$ and $M = 2, 4, 6, 8, 10$. 
Fig. 4. Ratio of the variance of the interdeparture times to the variance of the interarrival times $\frac{\text{var}[\text{IDT}]}{\text{var}[\text{IAT}]}$ versus the burstiness factor $K$, for $\rho = 0.13, S = 15$ and $M = 2, 4, 6, 8, 10$.

Fig. 5. Coefficient of correlation between two consecutive interdeparture times versus the load $\rho$, for $K = 3, S = 15$ and $M = 2, 4, 6, 8, 10$.

In Fig. 2, we have plotted the minimum size $S$ of the waiting room which is required to have a cell loss ratio less than $10^{-9}$, against the burstiness factor $K$ of the source, for a load $\rho = 0.13$ and various values of the constant service time $M$. From this figure we observe that the required buffer size is an increasing function of both $K$ and $M$. The more bursty the source is and the longer the service times, the larger the minimum buffer size that reduces the cell loss ratio to a specified value. Fig. 3 shows the variance of the cell delay $\text{var}[d]$ versus the load $\rho$, for $K = 3, S = 15$ and various values of the service time $M$. Apparently, for a fixed value of $M$, the delay variance reaches
a maximum as a function of $\rho$ for $\rho = 1/M$, and the maximum value of $\text{var}[d]$ increases with increasing $M$.

Next, we present some numerical results about the characteristics of the output traffic. The ratio of the variance of the interdeparture times to the variance of the interarrival times $\text{var}[\text{IDT}]/\text{var}[\text{IAT}]$ is shown in Fig. 4, for $\rho = 0.13$, $S = 15$ and various values of $K$ and $M$. In Fig. 5, the coefficient of correlation between two successive interdeparture times is displayed as a function of the load $\rho$, for $K = 3$, $S = 15$ and various values of $M$.

9. Conclusions

In this paper, we have presented an analytical technique, based on generating functions, for the performance evaluation of a discrete-time single-server finite-capacity queueing system with non-independent arrivals and constant service times of arbitrary length. The obtained results include the pgf’s of the queue length, the unfinished work, the cell delay and the interdeparture times in the steady state. The paper illustrates that generating functions, which traditionally are only used in the queueing analysis of infinite-capacity queueing systems, can be very useful for the performance analysis of finite-capacity queues as well.

Appendix. Calculation of the functions $Q_{i,n}(z)$ and $R_n(y, z)$

The functions $Q_{i,n}(z)$ and $R_n(y, z)$, defined in Eqs. (11) and (12), can be calculated as follows. First, using the system equations (4) and (5) and the assumption that the durations of the on-periods and the off-periods of the source are geometrically distributed, we easily arrive at the following relationships between the functions $Q_{i,n}(z)$ and the functions $Q_{i-1,n}(z)$:

$$
\begin{align*}
\begin{pmatrix}
Q_{i-1,0}(z)
\end{pmatrix}
&= \begin{pmatrix}
\beta & 1 - \alpha \\
(1-\beta)z & \alpha z
\end{pmatrix}
\begin{pmatrix}
Q_{i-1,1}(z)
\end{pmatrix}
+ p_{i-1}z^{S+1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad 2 \leq i \leq M - 1, \quad \text{(A.1)}
\end{align*}
$$

$$
\begin{align*}
\begin{pmatrix}
Q_{1,0}(z)
\end{pmatrix}
&= \begin{pmatrix}
\beta & 1 - \alpha \\
(1-\beta)z & \alpha z
\end{pmatrix}
\begin{pmatrix}
Q_{0,0}(z)
\end{pmatrix}
- p_0 \begin{pmatrix} \beta \\ (1-\beta)z \end{pmatrix}, \quad \text{(A.2)}
\end{align*}
$$

Here $p_0$ is the steady-state probability of having an empty buffer at the beginning of an arbitrary slot and the parameters $p_i$ are defined as

$$
p_i \triangleq \lim_{k \to \infty} \text{Prob}[r_k = i, a_k = 1, s_k = S + 1], \quad 1 \leq i \leq M - 2. \quad \text{(A.3)}
$$

Analogously, from the system equations (6)–(9), the following equations can be derived relating the functions $Q_{M-1,n}(z)$ and $Q_{M-1,n}(z)$:

$$
\begin{align*}
\begin{pmatrix}
Q_{M-1,0}(z)
\end{pmatrix}
&= \begin{pmatrix}
\beta & 1 - \alpha \\
(1-\beta)z & \alpha z
\end{pmatrix}
\begin{pmatrix}
Q_{M-1,1}(z)
\end{pmatrix}
+ p_{M-1}z^{S+1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{(A.4)}
\end{align*}
$$
where
\[ p_{M-1} \triangleq \lim_{k \to \infty} \text{Prob}[r_k = M - 1, a_k = 1, s_k = S + 1]. \] (A.5)

Eqs. (A.1), (A.2) and (A.4) can then be used to derive a set of linear equations with the unknown functions \( Q_{0,0}(z) \) and \( Q_{0,1}(z) \). After some repeated substitutions, we find
\[
(zI - F^M) \begin{pmatrix} Q_{0,0}(z) \\ Q_{0,1}(z) \end{pmatrix} = (zI - F^{M-1}) \begin{pmatrix} \beta \\ (1 - \beta)z \end{pmatrix} p_0 + \sum_{i=1}^{M-1} p_i z^{S+1} (1 - z) F^{M-i-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},
\] (A.6)
where
\[
I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]
and the matrix \( F^m \) is defined as
\[
F^m \triangleq \begin{pmatrix} \beta & 1 - z \\ (1 - \beta)z & az \end{pmatrix}^m, \quad 0 \leq m \leq M.
\] (A.7)

Let us now denote by \( \lambda_1(z) \) and \( \lambda_2(z) \) the two eigenvalues of the matrix \( F \). These eigenvalues are the solutions of the characteristic equation \( \lambda^2 - (xz + \beta) \lambda + (z + \beta - 1)z = 0 \), i.e.
\[
\lambda_1(z) = u(z) + v(z) \quad \text{and} \quad \lambda_2(z) = u(z) - v(z),
\] (A.8)
where
\[
u(z) = \frac{1}{2}(xz + \beta) \quad \text{and} \quad [v(z)]^2 = \frac{1}{4}[(xz + \beta)^2 - 4(z + \beta - 1)z].
\] (A.9)

It is also clear that \( \lambda_1(z) + \lambda_2(z) = xz + \beta \) and \( \lambda_1(z) \lambda_2(z) = (x + \beta - 1)z \). The matrix \( F^m \), \( 0 \leq m \leq M \), can then be expressed as follows:
\[
F^m = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} \lambda_1^{m+1} - \lambda_2^{m+1} - xz(\lambda_1^m - \lambda_2^m) & (1 - z)(\lambda_1^m - \lambda_2^m) \\ (1 - \beta)z(\lambda_1^m - \lambda_2^m) & \lambda_2^m - \lambda_1^m + (1 - \beta)(\lambda_1^m - \lambda_2^m) \end{pmatrix}.
\] (A.10)

Consequently, the functions \( Q_{0,0}(z) \) and \( Q_{0,1}(z) \) can be derived explicitly from the set of equations (A.6) as
\[
Q_{0,n}(z) = \frac{1}{N(z)} \left\{ p_0 L_n(z) + \sum_{i=1}^{M-1} p_i z^{S+1}(1 - z) T_{i,n}(z) \right\}, \quad n = 0, 1,
\] (A.11)
where the functions \( N(z), L_n(z) \) and \( T_{i,n}(z) \) are expressed in terms of the eigenvalues \( \lambda_1(z) \) and \( \lambda_2(z) \) as
\[
N(z) = z[z - (\lambda_1^M + \lambda_2^M) + (x + \beta - 1)^M z^{M-1}],
\] (A.12)
\[
L_0(z) = \frac{1}{(\lambda_1 - \lambda_2)} \left\{ [\beta x^2 + (\lambda_1 \lambda_2)^M](\lambda_1 - \lambda_2) + (1 + \beta)z \lambda_1 \lambda_2(\lambda_1^{M-1} - \lambda_2^{M-1}) - (\lambda_1^M - \lambda_2^M) \right\},
\] (A.13)
\[
L_1(z) = \frac{(1 - \beta) x^2}{(\lambda_1 - \lambda_2)} \left\{ \lambda_1 \lambda_2(\lambda_1^{M-1} - \lambda_2^{M-1}) - (\lambda_1^M - \lambda_2^M) + z(\lambda_1 - \lambda_2) \right\},
\] (A.14)
\[ T_{i,0}(z) = \frac{(1 - z)}{(\lambda_1 - \lambda_2)} \left\{ \left( z(\lambda_1^{M-i-1} - \lambda_2^{M-i-1}) + (\lambda_1 \lambda_2)^{M-i-1}(\lambda_1^{i+1} - \lambda_2^{i+1}) \right) \right\}, \quad (A.15) \]

\[ T_{i,1}(z) = \frac{1}{(\lambda_1 - \lambda_2)} \left\{ -\beta z(\lambda_1^{M-i-1} - \lambda_2^{M-i-1}) + (\lambda_1 \lambda_2)^{M-i-1}(\lambda_1^{i} - \lambda_2^{i}) \right. \]
\[ + z(\lambda_1^{M-i} - \lambda_2^{M-i}) - \beta(\lambda_1 \lambda_2)^{M-i-1}(\lambda_1^{i+1} - \lambda_2^{i+1}) \right\}, \quad (A.16) \]

respectively. In expression (A.11) for \(Q_{0,n}(z)\), the \(M\) probabilities \(p_0\) and \(p_1\) \((1 \leq i \leq M - 1)\) still remain to be determined. Note that under the assumption of an infinite storage capacity for the queueing system, i.e. \(S = \infty\), the following expression is obtained for \(Q_{0,n}(z)\):

\[ Q_{0,n}(z) = p_0L_n(z)/N(z), \quad n = 0, 1, \quad S = \infty, \quad (A.17) \]

which contains one unknown probability \(p_0\) only. A comparison of Eq. (A.11) and (A.17) reveals that the assumption of a finite buffer size results in an extra term (which contains the unknowns \(p_i\) \((1 \leq i \leq M - 1)\)) in the numerator of \(Q_{0,n}(z)\).

From Eqs. (A.8) and (A.9), it follows that \(\lambda_1^M + \lambda_2^M\) is a polynomial in \(z\) of degree \(M\), and \((\lambda_1^m - \lambda_2^m)/(\lambda_1 - \lambda_2), m \geq 1\), is a polynomial in \(z\) of degree \(m - 1\). Consequently, the denominator \(N(z)\) of \(Q_{0,n}(z)\) is a polynomial in \(z\) of degree \(M + 1\), and therefore has exactly \(M + 1\) zeros inside the complex \(z\)-plane. Also, the numerator in Eq. (A.11) is a polynomial in \(z\). Furthermore, it can be shown that \(N(0) = N(1) = 0\) and \(L_n(0) = L_n(1) = 0\), which implies that both the denominator and the numerator in \(Q_{0,n}(z)\) vanish for \(z = 1\).

Finally, by using definition (12) and Eqs. (A.1) and (A.2), we can establish a set of two linear equations for the functions \(R_0(y, z)\) and \(R_1(y, z)\), which has the following solution:

\[ \begin{pmatrix} R_0(y, z) \\ R_1(y, z) \end{pmatrix} = \frac{ \begin{pmatrix} 1 - xyz & (1 - z)y \\ (1 - \beta)y & 1 - \beta y \end{pmatrix} }{ \begin{pmatrix} (1 - \beta y)(1 - xyz) & (1 - \beta)(1 - z)y^{2}z \end{pmatrix} } \begin{pmatrix} (1 - y^Mz)Q_{0,0}(z) \\ Q_{0,1}(z) \end{pmatrix} \]
\[ - (y - y^Mz)p_0 \begin{pmatrix} \beta \\ (1 - \beta)z \end{pmatrix} + \sum_{i=1}^{M-1} p_i y^{i+1}z^{S+1}(1 - z)^{0} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (A.18) \]

Acknowledgements

The authors wish to thank the Fund for Scientific Research-Flanders (Belgiur F.W.O.) for support of this research.

References


