A General Discrete-Time Queueing Model: Analysis and Applications

U. Briem, T. H. Theimer and H. Kröner

University of Stuttgart
Institute of Communications Switching and Data Technics
Seidenstraße 36, D – 7000 Stuttgart 1, Federal Republic of Germany

Abstract

Realistic source models reflecting the main characteristics of traffic sources such as their correlation structure are necessary to investigate the performance of ATM network components. In this paper we study a discrete-time single server queueing system with a non-renewal input process which fulfills the above requirements while still being analytically tractable.
An exact analysis of the infinite and of the finite capacity queue yields the state probabilities at departure instants and at an arbitrary time as well as the loss probability. Further, we present an approximate solution for the finite buffer system which allows a significant reduction of the computational effort. The analysis is used to demonstrate the influence of serial correlations in the arrival process. Finally, some examples and numerical results are given to illustrate the accuracy of the approximation.

1 Introduction

Realistic source models reflecting the main characteristics of traffic sources such as video codecs are essential in order to study the performance of ATM network components. For simulation purposes, the source models can in principle be sufficiently detailed (though complex), but for analytical investigations they are usually very simple. Unfortunately, the simulation technique has several limitations which reduce its applicability for the performance study of ATM networks. These limitations are due to the measurement of extremely small probabilities below 10^{-6} and the runtime requirements for complex traffic models. Hence, analytic models which enable a queueing analysis are useful, e.g., when dimensioning a statistical multiplexer to provide a loss probability of 10^{-10}.

Commonly used analytic source models are the Poisson process and the Bernoulli process. However, these models do not incorporate important characteristics of the traffic such as its correlation structure or periodicity [16]. Therefore, more realistic models are required which include the discrete-time nature and other basic properties of the cell traffic, while still being analytically tractable. The Discrete-Time Markovian Arrival Process (DMAP) [1] is a promising approach in this direction, because it covers various source models which have been previously applied to model ATM traffic, e.g., the talkspurt-silence source studied in [7].

In this paper we first give a formal definition of the DMAP and derive some of its important characteristics like the counting function and the interarrival time distribution. Then the analysis of an infinite capacity discrete-time single server queueing system with a general service time distribution is briefly described. Since the arrival process is a discrete-time analog of the MAP introduced in [10], the analysis follows the same concepts as outlined there and in [13, 14]. Based on an embedded Markov chain approach, the queue length distribution and its moments at departure and arrival instants as well as at an arbitrary time are obtained.

The following analysis of the DMAP/G/1 queue with finite buffer capacity yields an exact solution for the state probabilities and the loss probability. In order to reduce the computational effort, an approximation for the state probabilities immediately after a departure instant is derived by taking into account the state probabilities at departures in the system with infinite buffer capacity.

Finally, a number of applications for this general queueing model are presented. We study the characteristics of the Leaky Bucket algorithm used for ATM source monitoring and the influence of serial correlations in the arrival process on the queueing performance. Further, it is shown how the model can be used to analyze a discrete-time finite capacity GI/G/1 queue, a problem which is in general very difficult to solve for arbitrary service time distributions.

2 A Discrete-Time Markovian Arrival Process

The DMAP is a stochastic process which is based on an irreducible discrete-time Markov chain with state space $\{1, \ldots, m\}$. If the process is in state i at time $k \cdot \Delta t$, it

moves to state j at time $(k+1)\cdot\Delta t$ with probability $u_{ij} = c_{ij} + d_{ij}$. Here, c_{ij} is the probability that the transition from state i to state j occurs without an arrival, and d_{ij} is the probability that the same transition generates an arrival event. Thus, the process is completely defined in terms of the matrices $\mathbf{C}$ and $\mathbf{D}$, where $\mathbf{C} = [c_{ij}]$ and $\mathbf{D} = [d_{ij}]$. From this definition it is clear that the transition matrix of the underlying Markov chain is $\mathbf{C} + \mathbf{D}$, and consequently $\sum_{j=1}^{m}(c_{ij} + d_{ij}) = 1$, $\forall\, i$. In the following, the state of the Markov chain describing the DMAP will be called its phase to avoid confusion with the state of the queueing system we consider later on.

Note that the DMAP is the discrete-time analog of a stochastic process described in [10]. It was introduced in [1], and a detailed description of its characteristics can be found in [2]. The next paragraphs summarize the results of [2] which are relevant in the context of this paper.

In order to compute the transition matrix of the embedded Markov chain described in Section 3 it is necessary to evaluate the counting function of the arrival process, i.e. the distribution of the number of arrivals in an interval of k time slots. Considering a single slot, it is clear from the above definition of the process that d_{ij} is the probability that in the next time slot there is an arrival and the process will be in state j, given that the current state is i. Analogous, c_{ij} is the probability that in the next time slot there is no arrival event and the state of the Markov chain will be j, given that the momentary state is i.

Let $P_{ij}(n,k)$ be the conditional probability that there are n arrivals in the interval $(0,k]$ and that the state of the process is j at the end, given that it starts in phase i at time 0. The matrix $\mathbf{P}(n,1) = [P_{ij}(n,1)]$ describing a single time slot is given by

$$\mathbf{P}(n,1) = \begin{cases} \mathbf{C} & \text{if } n = 0 \\ \mathbf{D} & \text{if } n = 1 \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{1}$$

Consequently, the generating function of $\mathbf{P}(n,1)$ is

$$\tilde{\mathbf{P}}(z,1) = \sum_{\nu=0}^{\infty} \mathbf{P}(\nu,1)\, z^{\nu} = \mathbf{C} + \mathbf{D}z. \tag{2}$$

Due to the Markovian properties of a DMAP, the general expression for the generating function of $\mathbf{P}(n,k)$ is easily obtained as

$$\tilde{\mathbf{P}}(z,k) = [\tilde{\mathbf{P}}(z,1)]^{k} = [\mathbf{C} + \mathbf{D}z]^{k}. \tag{3}$$

The general expression for $\mathbf{P}(n,k)$ is rather difficult to derive, since the matrix multiplication is not a commutative operation. However, $\mathbf{P}(n,k)$ can be computed numerically using a recursive scheme similar to the algorithm described in [9].

Let $f_{ij}(k)$ be the conditional probability that the interarrival time is k time slots and that the state of the DMAP at the next arrival is j, given that the state was i at the previous arrival. If the interarrival time is exactly k time slots there must be $k-1$ slots without an arrival, followed by the last slot with an arrival event. Therefore, the probability matrix $\mathbf{f}(k)$ is given by

$$\mathbf{f}(k) = \begin{cases} \mathbf{C}^{k-1}\mathbf{D} & \text{if } k \geq 1 \\ \mathbf{0} & \text{otherwise,} \end{cases} \tag{4}$$

with generating function

$$\tilde{\mathbf{f}}(z) = \sum_{\nu=1}^{\infty} \mathbf{f}(\nu)\, z^{\nu} = z\,(\mathbf{I} - \mathbf{C}z)^{-1}\,\mathbf{D}, \quad |z| < 1, \tag{5}$$

where $\mathbf{A}^{-1}$ denotes the matrix inverse of $\mathbf{A}$. In order to determine the mean value and higher order moments of the interarrival time, we need the derivatives of $\tilde{\mathbf{f}}(z)$ for $z = 1$. It can be proved by complete induction that the n-th derivative of $\tilde{\mathbf{f}}(z)$ is

$$\tilde{\mathbf{f}}(z)^{(n)} = n! \cdot (\mathbf{I} - \mathbf{C}z)^{-(n+1)}\, \mathbf{C}^{n-1}\, \mathbf{D}. \tag{6}$$

From this general relationship, the matrix of mean interarrival times is immediately derived:

$$\frac{d}{dz}\tilde{\mathbf{f}}(z)\,|_{z=1} = (\mathbf{I} - \mathbf{C})^{-2}\, \mathbf{D} \tag{7}$$

3 Analysis of the DMAP/G/1 Queue

The queueing model consists of a discrete-time single server queue with infinite capacity and FIFO service discipline. The arrival process is supposed to be a DMAP characterized by the transition matrices $\mathbf{C}$ and $\mathbf{D}$. The service time has a general discrete-time distribution $h(k)$, $k \geq 0$, with moments $\mu^{(n)} = \sum_{\nu=0}^{\infty} \nu^{n}h(\nu)$, $\mu^{(1)} = \bar{h}$. Mutual independence is assumed with respect to the arrival process and the service times. Since we consider a discrete-time queueing system, an arrival and a departure may occur simultaneously. In this case the arriving customer enters the system immediately before the departing one leaves (arrival first, AF).

Let τ_k be the time until the k-th departure in the above queueing model, given that $\tau_0 = 0$. Note that the phase of the arrival process completely contains its memory. Therefore, if we define R_k and J_k to be the number of customers in the system and the phase of the DMAP at τ_k^+, the sequence $\{(R_k, J_k, \tau_{k+1} - \tau_k) : k \geq 0\}$ forms a semi-Markov chain with state space $\{(r, j) : r \geq 0, 1 \leq j \leq m\}$. The transition probability matrix has the same structure as for the M/G/1 queue and is given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \dots \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \dots \\ \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{8}$$

where the $m \times m$ matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ are determined by

$$\mathbf{A}_n = \sum_{\nu=0}^{\infty} \mathbf{P}(n,\nu) \cdot h(\nu), \tag{9}$$

$$\mathbf{B}_n = \mathbf{U} \cdot \mathbf{A}_n, \quad \text{and} \tag{10}$$

$$\mathbf{U} = \sum_{\nu=0}^{\infty} \mathbf{C}^{\nu} \mathbf{D} = \tilde{\mathbf{f}}(z)\mid_{z=1} = (\mathbf{I} - \mathbf{C})^{-1} \mathbf{D}. \tag{11}$$

$[\mathbf{A}_n]_{ij}$ is the conditional probability that there are n arrivals during a service time and the DMAP is in phase j at the departure instant, given that the system was not empty after the last departure with the DMAP in phase i. In addition, $\mathbf{U}$ gives the phase transition probabilities between the beginning and the end of an idle period, which is the time between a departure leaving the system empty and the next arrival.

The solution method for this queueing model is based on a matrix analytic approach, which can be found in several publications [10, 13, 14]. Our intention is not to reproduce the whole analysis. However, for completeness, we briefly outline the major steps using matrix analytic techniques. Some steps are also required to describe the exact and an approximate calculation of the loss probability for the queue with finite buffer capacity.

In order to derive the state probabilities at departures the stationary state vector $\vec{x}_0$ is needed. Its components $x_{0,i}$ correspond to the probability that a departure leaves the system empty with the DMAP in phase i. We start by defining some additional variables:

EF := mean length of an idle period (in time slots).

EB := mean length of a busy period (time between two successive idle periods).

EK := mean number of departures in a busy period.

$\vec{l}$:= stationary phase vector at departure instants after which the system is empty.

$\vec{e}$:= $(1, 1, \ldots, 1)^T$.

λ := mean arrival rate of the DMAP. It is determined by $\lambda = \vec{\pi} \cdot \mathbf{D} \cdot \vec{e}$, where the stationary phase vector $\vec{\pi}$ is implicitly given by $\vec{\pi} \cdot (\mathbf{C} + \mathbf{D}) = \vec{\pi}, \vec{\pi} \cdot \vec{e} = 1$.

Using the equations

$$\frac{EF}{EF + EB} = 1 - \rho, \quad EB = \bar{h} \cdot EK \quad \text{and}$$

$$EF = \vec{l} \cdot (\mathbf{I} - \mathbf{C})^{-2} \mathbf{D} \cdot \vec{e} = \vec{l} \cdot (\mathbf{I} - \mathbf{C})^{-1} \cdot \vec{e},$$

leads to

$$\vec{x}_0 = \frac{\vec{l}}{EK} = \frac{1 - \rho}{\lambda} \frac{\vec{l}}{\vec{l} \cdot (\mathbf{I} - \mathbf{C})^{-1} \cdot \vec{e}}. \tag{12}$$

The following derivation of $\vec{l}$ is based on the matrix $\mathbf{G}(k)$ with components $[\mathbf{G}(k)]_{ij}$ corresponding to the conditional probability that a busy period ends after exactly k departures with the DMAP in phase j, given that it was in phase i at the beginning of this busy period. Due to the structure of $\mathbf{Q}$ this is equal to the probability that the first departure leaving r customers in the system occurs exactly after k transitions with the DMAP in phase j, given that the embedded Markov chain started in state $(r + 1, i)$. The probability is independent of r, $r \geq 0$. Hence, denoting $\tilde{\mathbf{G}}(z) = \sum_{\nu=0}^{\infty} \mathbf{G}(\nu) z^{\nu}$, we obtain for $|z| \leq 1$

$$\tilde{\mathbf{G}}(z) = z \cdot \sum_{\nu=0}^{\infty} \mathbf{A}_{\nu} \left[\tilde{\mathbf{G}}(z) \right]^{\nu} = z \cdot \sum_{\nu=0}^{\infty} \tilde{\mathbf{P}}(\tilde{\mathbf{G}}(z), \nu) h(\nu)$$

$$= z \cdot \sum_{\nu=0}^{\infty} \left[\mathbf{C} + \mathbf{D}\tilde{\mathbf{G}}(z) \right]^{\nu} h(\nu). \tag{13}$$

Assuming $\mathbf{C} + \mathbf{D}$ to be irreducible and $\rho < 1$ (stable system), the matrix

$$\mathbf{G} = \sum_{\nu=0}^{\infty} \mathbf{G}(\nu) = \tilde{\mathbf{G}}(z)\mid_{z=1} = \sum_{\nu=0}^{\infty} \left[\mathbf{C} + \mathbf{D}\mathbf{G} \right]^{\nu} h(\nu) \tag{14}$$

is stochastic [13] and represents the phase transition matrix of a busy period. Its stationary vector is given by $\vec{g} \cdot \mathbf{G} = \vec{g}, \vec{g} \cdot \vec{e} = 1$. The matrix $\mathbf{G}$ may be computed by successive substitutions in Equation 14. More efficient algorithms are given in [5] to reduce the number of iterations especially at high loads. Since there is always an idle period followed by a busy period between two successive departures leaving the system empty, $\vec{l}$ is determined by

$$\vec{l} \cdot \mathbf{U}\mathbf{G} = \vec{l}, \qquad \vec{l} \cdot \vec{e} = 1. \tag{15}$$

Using Equation 14 it can be shown that $\vec{g}(\mathbf{C} + \mathbf{D}\mathbf{G}) = \vec{g}$. This together with Equation 15 implies

$$\vec{l} = \left[\vec{g}(\mathbf{I} - \mathbf{C}) \vec{e} \right]^{-1} \cdot \vec{g}(\mathbf{I} - \mathbf{C}) \quad \text{and} \tag{16}$$

$$\vec{x}_0 = \frac{1 - \rho}{\lambda} \vec{g}(\mathbf{I} - \mathbf{C}). \tag{17}$$

As described in [15] the vectors $\vec{x}_r$, $r > 0$, can be obtained using the numerically stable recursion

$$\vec{x}_r = \left(\vec{x}_0 \bar{\mathbf{B}}_r + \sum_{\nu=1}^{r-1} \vec{x}_{\nu} \bar{\mathbf{A}}_{r+1-\nu} \right) \cdot (\mathbf{I} - \bar{\mathbf{A}}_1)^{-1}, \quad r \geq 1, \tag{18}$$

where $\bar{\mathbf{A}}_w = \sum_{\nu=w}^{\infty} \mathbf{A}_{\nu} \cdot \mathbf{G}^{\nu-w}$ and

$$\bar{\mathbf{B}}_w = \sum_{\nu=w}^{\infty} \mathbf{B}_{\nu} \cdot \mathbf{G}^{\nu-w}, \quad \text{for } w \geq 0.$$

From the state probabilities at departures the stationary queue length distribution at the end of an arbitrary time unit is now derived. First we focus on the probability $y_{0,j}$ that the system is empty at the end of an arbitrary time slot with the DMAP in phase j. This is

only possible if the last departure has left the system empty and there have been no further arrivals. Hence, the element $[\sum_{\nu=0}^{\infty} \mathbf{C}^\nu]_{ij}/ET_D$ is the conditional probability that the system is in state $(0, j)$ at the end of an arbitrary time unit, given that the last departure left the system empty with the DMAP in phase i. Note that the mean number of time units between successive departures ET_D is equal to the mean interarrival time $ET_A = 1/\lambda$ in a queueing system without losses. This leads to

$$\vec{y}_0 = \lambda \, \vec{x}_0 \sum_{\nu=0}^{\infty} \mathbf{C}^\nu = \lambda \, \vec{x}_0 \, (\mathbf{I} - \mathbf{C})^{-1} = (1 - \rho) \, \vec{g}. \quad (19)$$

In a similar way the vectors $\vec{y}_r$, $r \geq 1$, can be obtained, although the scenarios are more complex. The derivation goes along the line of reference [10], with the only difference that we consider a discrete-time model. Due to space limitations the details are skipped, the resulting expression is

$$\vec{y}_r = \lambda \, \vec{x}_0 \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathbf{C}^l \, \mathbf{D} \, \mathbf{P}(r-1, k) \sum_{\nu=k+1}^{\infty} h(\nu)$$
$$+ \lambda \sum_{w=1}^{r} \vec{x}_w \sum_{k=0}^{\infty} \mathbf{P}(r-w, k) \sum_{\nu=k+1}^{\infty} h(\nu). \quad (20)$$

After some manipulations Equation 20 leads to the following recursion (compare with [10]):

$$\vec{y}_{r+1} = [\vec{y}_r \cdot \mathbf{D} + \lambda(\vec{x}_{r+1} - \vec{x}_r)] \cdot (\mathbf{I} - \mathbf{C})^{-1}, \quad r \geq 0. \quad (21)$$

4 The Finite Capacity Queue

In order to obtain the loss probability of the system with finite buffer capacity we calculate the stationary state probabilities $\vec{x}_r^*$ at departures and the probabilities $\vec{y}_r^*$ at the end of an arbitrary time slot. In the following, the asterisk '*' indicates that the corresponding quantity refers to the finite buffer system.

For a system with finite buffer size s, the transition matrix of the Markov chain is given by

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \dots & \mathbf{B}_{s-1} & \mathbf{B}_{sum}^s \\ \mathbf{A}_0 & \mathbf{A}_1 & \dots & \mathbf{A}_{s-1} & \mathbf{A}_{sum}^s \\ 0 & \mathbf{A}_0 & \dots & \mathbf{A}_{s-2} & \mathbf{A}_{sum}^{s-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{A}_0 & \mathbf{A}_{sum}^1 \end{pmatrix}, \quad (22)$$

where $\quad \mathbf{A}_{sum}^w = \sum_{\nu=w}^{\infty} \mathbf{A}_\nu \quad$ and

$$\mathbf{B}_{sum}^w = \sum_{\nu=w}^{\infty} \mathbf{B}_\nu, \quad \text{for } w \geq 1.$$

A direct or an iterative solution of the system of equations $(\vec{x}_0^*, ..., \vec{x}_s^*) \, \mathbf{Q}^* = (\vec{x}_0^*, ..., \vec{x}_s^*)$ yields the exact values of $\vec{x}_r^*$. However, these methods are either memory or time consuming, especially for large m and s. Therefore, in the second part of this section an approximation is

proposed which reduces the computational effort. If an exact solution is necessary, it can be obtained very efficiently from an additional iteration starting with the approximated values. In this case the iteration converges rapidly, because its initial vectors are already close to the exact solution.

In the next paragraphs the stationary queue length distribution at an arbitrary time slot $\vec{y}_r^*$, $0 \leq r \leq s + 1$, is related to the state vectors $\vec{x}_r^*$ at departures. Except for $\vec{y}_{s+1}^*$, this can be done by considerations similar to Equation 19 and 20 with the difference that the arrival rate λ has to be replaced by the inverse of the mean interdeparture time ET_D^*. Since there is either an idle period and one service time — in the case where a departure leaves the system empty — or exactly one service time between two successive departures, the mean interdeparture time can be calculated from

$$ET_D^* = \vec{x}_0^* \cdot \frac{d}{dz}\tilde{\mathbf{f}}(z) \mid_{z=1} \cdot \vec{e} + \bar{h}$$
$$= \vec{x}_0^* \cdot (\mathbf{I} - \mathbf{C})^{-1} \cdot \vec{e} + \bar{h}. \quad (23)$$

For computing the blocking probability $\vec{y}_{s+1}^*$ the fact is used that the sum of all probabilities $\vec{y}_r^*$ must be equal to the stationary phase vector $\vec{\pi}$ of the arrival process. In agreement with [3] the following expressions for the queue length distribution at an arbitrary time slot are obtained:

$$\vec{y}_0^* = \frac{1}{ET_D^*} \cdot \vec{x}_0^* \cdot (\mathbf{I} - \mathbf{C})^{-1} \quad (24)$$

$$\vec{y}_{r+1}^* = \left[\vec{y}_r^* \cdot \mathbf{D} + \frac{1}{ET_D^*} (\vec{x}_{r+1}^* - \vec{x}_r^*) \right] \cdot (\mathbf{I} - \mathbf{C})^{-1},$$
$$r = 0, ..., s - 1 \quad (25)$$

$$\vec{y}_{s+1}^* = \vec{\pi} - \sum_{r=0}^{s} \vec{y}_r^* \quad (26)$$

Finally, the loss probability B can be computed from

$$B = \frac{\text{P\{queue full, cell arrival\}}}{\text{P\{cell arrival\}}} = \frac{1}{\lambda} \, \vec{y}_{s+1}^* \cdot \mathbf{D} \cdot \vec{e} \quad (27)$$

or by applying $1/ET_D^* = \lambda(1 - B)$ together with Equation 23 which leads to

$$B = 1 - \frac{1}{\lambda \cdot ET_D^*}. \quad (28)$$

Another way to obtain B is based on the following direct approach for the blocking probability (compare with Equation 20):

$$\vec{y}_{s+1}^* = \lambda(1 - B) \sum_{r=s+1}^{\infty} \sum_{k=0}^{\infty} \sum_{\nu=1}^{\infty} h(k + \nu)$$
$$\cdot \left[\vec{x}_0^* \cdot \mathbf{U} \, \mathbf{P}(r-1, k) + \sum_{w=1}^{s} \vec{x}_w^* \cdot \mathbf{P}(r-w, k) \right] (29)$$

Here, the summation $\sum_{r=s+1}^{\infty} ...$ takes into account possible cell losses between the last departure and the arbitrary time slot. Using Equation 27 and 29 some manipulations finally lead to

$$B = \frac{\alpha(\vec{x}_0^*, ..., \vec{x}_s^*, \mathbf{A}_0, \mathbf{A}_1, ...)}{1 + \alpha(\vec{x}_0^*, ..., \vec{x}_s^*, \mathbf{A}_0, \mathbf{A}_1, ...)}, \quad \text{where}$$

$$\alpha(\vec{x}_0^*, ..., \vec{x}_s^*, \mathbf{A}_0, \mathbf{A}_1, ...) =$$

$$\vec{x}_0^* \cdot \mathbf{U} \sum_{r=s+1}^{\infty} \left[\mathbf{I} - \sum_{\nu=0}^{r-1} \mathbf{A}_\nu\right] \cdot \vec{e} + \sum_{r=s+1}^{\infty} \sum_{w=1}^{s} \vec{x}_w^* \cdot \left[\mathbf{I} - \sum_{\nu=0}^{r-w} \mathbf{A}_\nu\right] \cdot \vec{e}. \tag{30}$$

In the sequel, an approximation for the state probabilities at departures is given in order to obtain an efficient estimation for the loss probability. Note, that the state probabilities $\vec{x}_r^*$, $r < s$, are computed from $\vec{x}_0^*$ in the same way as in the system with infinite capacity. The transition probabilities differ only for the vector $\vec{x}_s^*$. Therefore we take the probabilities $\vec{x}_r$, $r < s$, of the infinite capacity system as an approximation for $\vec{x}_r^*$. The vector $\vec{x}_s^*$ is then estimated using the transition probabilities of $\mathbf{Q}^*$. This leads to

$$\vec{x}_s^* \approx \left(\vec{x}_0 \mathbf{B}_{sum}^s + \sum_{r=1}^{s-1} \vec{x}_r \mathbf{A}_{sum}^{s-r+1}\right) \cdot \left(\mathbf{I} - \mathbf{A}_{sum}^1\right)^{-1}. \tag{31}$$

Since the vector $\sum_{r=0}^{s} \vec{x}_r^*$ must be stochastic, the vectors $\vec{x}_r^*$ have to be normalized.

According to our experience an approximation of B using Equations 23–28 together with the approximation for $\vec{x}_r^*$ yields results which tend to be smaller than the exact values. With Equation 30 instead of Equations 23–28 the computational effort is higher, but our investigations have shown that this estimation is pessimistic which is preferred in most cases. Furthermore, its relative approximation error has been smaller. In the next section some examples and numerical results are given, which illustrate also the accuracy of the second approximation.

5 Applications

The DMAP is a very general arrival process and covers many discrete-time processes excluding batch arrivals. For instance, $\mathbf{D} = p$ and $\mathbf{C} = 1 - p$ implies a Bernoulli process with arrival probability p. A very general class of processes called General Modulated Deterministic Processes (GMDP) [8] can be modelled as a DMAP. This class includes the well-known burst silence model and the Markov Modulated Deterministic Process (MMDP) as a special case [2]. In addition, a DMAP allows the modelling of an arbitrary discrete-time renewal process, which has a well-defined maximum interarrival time. Let p_k be the probability of the interarrival time being k slots and $p_k = 0$ for $k = 0$ and $k > 3$. Then the matrices $\mathbf{C}$ and $\mathbf{D}$ are given by

$$\mathbf{C} = \begin{pmatrix} 0 & 1-p_1 & 0 \\ 0 & 0 & \frac{p_3}{p_2+p_3} \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} p_1 & 0 & 0 \\ \frac{p_2}{p_2+p_3} & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{32}$$

Further examples for the modelling of various discrete-time processes without batch arrivals can be found in [2]. For a burst silence arrival process and a deterministic service time the results have been validated by comparison with the results obtained from discrete-time analysis [19].

As a first application, we will study the characteristics of the Leaky Bucket algorithm used for source monitoring in ATM networks. This algorithm can be modelled as a finite capacity single server queueing system with deterministic service time h [17]. As shown in Figure 1
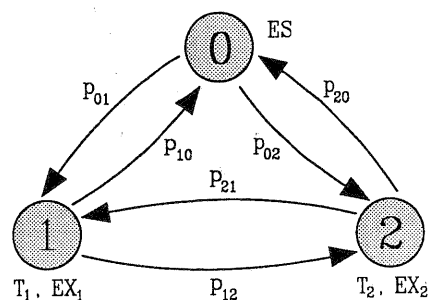

Figure 1: State transition diagram of an MMDP with 3 states

the input process is assumed to be an MMDP with 3 states which may describe the cell generation process of a variable bitrate source (e.g. video codec for ATM networks). In the states 1 and 2 cells have a constant interarrival time T_1 and T_2, respectively. While being in state 0 no cells are emitted. The number of cells generated in the states 1 and 2 and the sojourn time of the silence state have a shifted geometric distribution with mean $EX_1 = EX_2 = EX$ and ES. State changes occur with probability $p_{vw} = 0.5$, $v \neq w$. The offered traffic A and the burstiness BU of this arrival process can be computed from the equations

$$A = \frac{2\,EX\,h}{EX\,(T_1 + T_2) + ES} \quad \text{and}$$

$$BU = \frac{\text{max. cell rate}}{\text{mean cell rate}} \overset{T_1 \leq T_2}{=} \frac{EX\,(T_1 + T_2) + ES}{2\,EX\,T_1}. \tag{33}$$

Figure 2 depicts the cell loss probability of the Leaky Bucket mechanism versus bucket size for different source parameters. Obviously, the cell loss probability increases with the load. Higher burstiness and longer burst duration also lead to higher cell loss probabilities. Note that in all cases considered here T_1 and T_2 have been smaller than h which implies that the proposed approximation yields exact results, since the arrival process must be in the silence phase as long as the system is empty (including the previous departure). The silence phase corresponds to exactly one phase of the DMAP, so
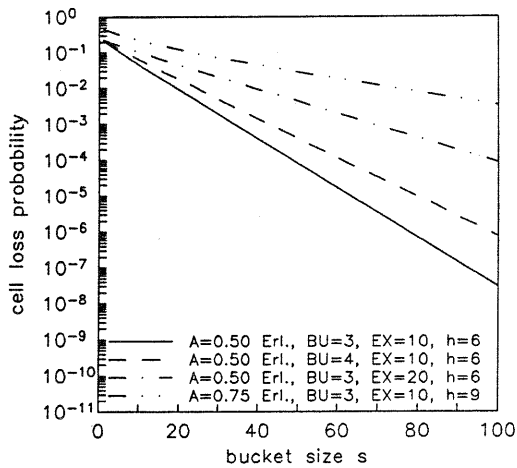
Figure 2: Cell loss probability versus bucket size

Figure 3: Loss probability versus coefficient of correlation

Figure 4: Relative approximation error versus coefficient of correlation

that $\vec{l}^* = \vec{l}$. Hence, the approximation uses the correct phase probability ratio and leads to exact results.

In a second example, the influence of serial correlations within the arrival process is investigated. Since the correlation of consecutive interarrival times is characterized by the coefficient of correlation c_c, we have used an arrival process where c_c is easily adjustable [18]. This process has two states with interarrival times T_v, $v = 1, 2$, being constant in each state. Therefore the process is equivalent to the one shown in Figure 1 without the silence phase, where the number of cells emitted in one state has a shifted geometric distribution with the same mean value $EX = (1-q)^{-1}$ for both states. Here, $1-q$ denotes the probability of changing the state after each cell arrival. For $T_1 \neq T_2$ the mean ET_A, the coefficient of variation c_v, and the coefficient of correlation c_c of the interarrival time T_A are given by

$$ET_A = \frac{1}{2}[T_1 + T_2],$$

$$c_v = \left| \frac{1 - T_1/T_2}{1 + T_1/T_2} \right|, \quad \text{and} \quad c_c = 2q - 1. \qquad (34)$$

These parameters can be chosen independently from each other for $ET_A > 0$, $0 < c_v < 1$, and $-1 < c_c < 1$. A coefficient of variation greater than 1 can be obtained, if the mean number of cells generated in each state are different. Figure 3 illustrates the loss probability versus coefficient of correlation for different buffer sizes, assuming a constant service time $h = 5$. The loss probability increases rapidly with the coefficient of correlation, because the state sojourn times of the arrival process tend to infinity if c_c approaches $+1$. Conversely, the process is a more or less deterministic alternating sequence of T_1 and T_2 if c_c is close to -1.

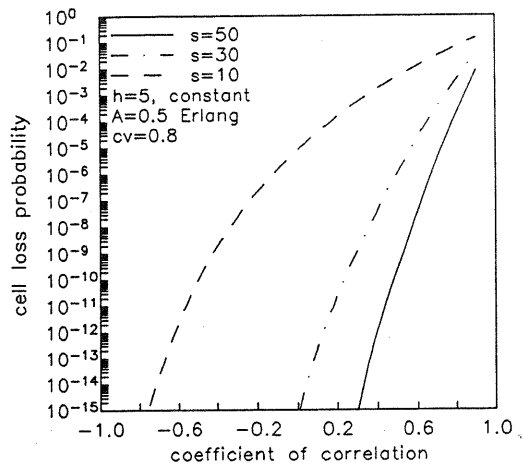The relative error of the proposed approximation shown in Figure 4 is evaluated from exact results, obtained by an iterative solution of the equation system resulting from the transition matrix given in Equation 22. The approximation is based on Equation 30. The curves indicate a good agreement of approximate and exact results. For the case where the relative error is maximal, we have studied the influence of c_v and A. Investigations have shown that the relative approximation error increases slightly with c_v and A. In all cases we studied the approximated cell loss probability has been larger than the exact value.

6 Conclusion

In this paper an exact analysis of a general discrete-time single server queueing model with correlated input is presented. For the system with finite buffer capacity an exact and an approximate solution for the loss probabil-

ity is given. As an example an MMDP arrival process is applied to study the characteristics of the Leaky Bucket algorithm used for source monitoring in ATM networks. It is also shown how the model can be used to analyze a discrete-time GI/G/1-s queue with a well-defined maximum interarrival time.

Further the influence of serial correlations within the arrival process on the cell loss probability is investigated. The study indicates a good agreement of the approximate results with the exact solution, which is either memory or time consuming. The relative approximation error increases with the offered traffic and the coefficient of variation, but within our investigations it has always been small. In all cases the approximated cell loss probability has been larger than the exact value. Future work includes the study of algorithms for matching the parameters of a DMAP to measurements of variable bitrate video codecs.

Acknowledgement

References

[1] C. BLONDIA, "A Discrete-time Markovian Arrival Process", Document RACE 1022, PRLB_123_0015_CC_CD, August 1989.

[2] C. BLONDIA, T. H. THEIMER, "A Discrete-time Model for ATM Traffic", Document RACE 1022, PRLB_123_0018_CC_CD / UST_123_0022_CC_CD, October 1989.

[3] C. BLONDIA, "Finite Capacity Vacation Models with Non-Renewal Input", to appear in Journal of Applied Probability, vol. 28, 1991.

[4] A. GRAVEY, J. R. LOUVION, P. BOYER, "On the Geo/D/1 and the Geo/D/1/n Queue", Performance Evaluation, vol. 11, no. 2, 1990, pp. 117-125.

[5] L. GÜN, "Experimental Results on Matrix-Analytical Solution Techniques — Extensions and Comparisons", Communications in Statistics — Stochastic Models, vol. 5, 1989, pp. 669-682.

[6] H. HEFFES, D. M. LUCANTONI, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance", IEEE Journal on Selected Areas in Communications, vol. SAC-4, no. 6, September 1986, pp. 856-868.

[7] H. ITOH, M. AKIYAMA, Y. TANAKA, "A Study on ATM Multiplexed Characteristics of Bursty Traffic Model", ITC Specialist Seminar, Adelaide, 1989.

[8] P. J. KÜHN, "Input Process Model GMDP", Document RACE 1022, UST_123_0003_CC_CD, June 1988.

[9] D. M. LUCANTONI, V. RAMASWAMI, "Efficient Algorithms for Solving the Non-Linear Matrix Equations Arising in Phase Type Queues", Communications in Statistics — Stochastic Models 1, no. 1, 1985, pp. 29-51.

[10] D. M. LUCANTONI, K. S. MEIER-HELLSTERN, M. F. NEUTS, "A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes", Working Paper 88-008, Systems and Industrial Engineering Department, The University of Arizona, 1988.

[11] M. F. NEUTS, "A Versatile Markovian Point Process" Journal of Applied Probability, vol. 16, 1979, pp. 764-779.

[12] M. F. NEUTS, Matrix-Geometric Solutions in Stochastic Models, The Johns Hopkins University Press, 1981.

[13] M. F. NEUTS, Structured Stochastic Matrices of M/G/1 Type and their Applications, Marcel Dekker Inc., New York and Basel, 1989.

[14] V. RAMASWAMI, "The N/G/1 Queue and its Detailed Analysis", Adv. Appl. Probability 12, 1980, pp. 222-261.

[15] V. RAMASWAMI, "Stable Recursion for the Steady State Vector in Markov Chains of M/G/1 Type", Communications in Statistics — Stochastic Models, vol. 4, 1988, pp. 183-188.

[16] V. RAMASWAMI, W. WILLINGER, "Efficient Traffic Performance Strategies for Packet Multiplexers", ITC Specialist Seminar, Adelaide, 1989.

[17] E. P. RATHGEB, "Policing Mechanisms for ATM Networks — Modelling and Performance Comparison", ITC Specialist Seminar, Morristown, October 1990.

[18] J. SOLÈ, J. DOMINGO, J. GARCÌA, "Traffic Burstiness Influence on ATM Switching Behaviour", Fourth International Conference on Data Communication Systems and their Performance, Barcelona, June 1990, pp. 133-147.

[19] P. TRAN-GIA, H. AHMADI, "Analysis of a Discrete-Time $G^{[x]}/D/1$-S Queueing System with Applications in Packet-Switching Systems", INFOCOM'87, New Orleans, 1987, pp. 861-870.