ANALYSIS OF DISCRETE-TIME MULTISERVER QUEUES WITH CONSTANT SERVICE TIMES AND CORRELATED ARRIVALS

Peixia Gao, Sabine Wittevrongel and Herwig Bruneel
Stochastic Modeling and Analysis of Communication Systems Research Group
Department of Telecommunications and Information Processing
Ghent University
Sint-Pietersnieuwstraat 41, B-9000 Gent, BELGIUM
E-mail: pg@telin.UGent.be

KEYWORDS
Queueing Systems and Network Models, Performance Modelling

ABSTRACT

We investigate the behavior of a discrete-time multiserver buffer system with infinite buffer size. Packets arrive at the system according to a two-state correlated arrival process. The service times of the packets are assumed to be constant, equal to multiple slots. The behavior of the system is analyzed by means of an analytical technique based on probability generating functions (pgf's). Explicit expressions are obtained for the pgf's of the system contents and the packet delay. From these, the moments and the tail distributions of the system contents and the packet delay can be calculated. Numerical examples are given to show the influence of various model parameters on the system behavior.

1 INTRODUCTION

Discrete-time queueing models have been used for many years to analyze the behavior and performance of digital communication networks, where buffers are used for the temporary storage of information packets awaiting transmission. In such discrete-time models, time is divided into fixed-length slots and the service or transmission of packets starts and ends at slot boundaries only. In the scientific literature, many results can be found with respect to the analysis of discrete-time single-server queues with various types of (uncorrelated or correlated) packet arrival processes and various types of service-time distributions. For systems with multiple servers however fewer results are available. Firstly, most studies of multiserver systems assume constant service times of one slot, see e.g. Bruneel et al. (1992) and Bruneel and Kim (1993). Only a limited number of papers consider more general service-time distributions. Multiserver systems with geometrically distributed service times have been studied in Rubin and Zhang (1991), Gao et al. (2004b), Gao et al. (2003) and Gao et al. (2004c); queues with multiple servers and constant service times of arbitrary length have been studied in Bruneel and Wuyts (1994) and Gao et al. (2004a). Secondly, in case of multiple servers, mostly an uncorrelated packet arrival process is considered, i.e., the numbers of packet arrivals during the consecutive slots are assumed to be independent, see e.g. Bruneel et al. (1992), Bruneel and Kim (1993), Rubin and Zhang (1991), Gao et al. (2004b), Bruneel and Wuyts (1994) and Gao et al. (2004a). In Gao et al. (2003) and Gao et al. (2004c), for the case of geometric service times, more general, so-called correlated packet arrival processes are considered, which are more adequate to describe the bursty nature of the traffic in nowadays communication networks.

In the present paper, we investigate the behavior of a discrete-time multiserver buffer system with constant service times of multiple slots and correlated arrivals. From the above survey, the paper can be seen as a generalization of Bruneel and Wuyts (1994) and Gao et al. (2004a) to the case of correlated arrivals. It is also an extension of Wittevrongel and Bruneel (1999) to the multiserver case.

The paper is organized as follows. In Section 2, we describe the system under study and introduce some notations. In Section 3, the pgf's of the partial system contents and system contents are derived, and the mean value, the variance and the tail distribution of the system contents are calculated. In Section 4, the characteristics of the packet delay are analyzed. In Section 5, some numerical examples are

Proceedings 12th International Conference ASMTA 2005
Khalid Al-Begain, Gunter Bolch, Miklos Telek © ECMS, 2005
ISBN 1-84233-112-4 (Set) / ISBN 1-84233-113-2 (CD)

given. Finally, the paper is concluded in Section 6.

2 SYSTEM DESCRIPTION AND NOTATIONS

We consider a discrete-time queueing system with c ($c \geq 1$) servers (or output channels) and an infinite buffer capacity for the storage of packets. Time is divided into fixed-length slots. Packets arrive at the input of the system and are queued in a buffer until they can be transmitted via one of the c output channels based on a FCFS (first-come-first-served) discipline. The service (or transmission) of a packet can start or end at slot boundaries only. The service times of the packets are assumed to be constant, equal to s ($s \geq 1$) slots.

The packet arrival process is modelled as follows. The traffic source has a bursty nature and alternates between two states, state 0 and state 1. Transitions between the states are assumed to occur at slot boundaries. The numbers of consecutive slots during which the source state is 0 or 1 are called 0-times and 1-times respectively. The 0-times and 1-times are assumed to be independent geometrically distributed random variables with parameters α and β respectively, i.e.,

$$\text{Prob}[\text{0-time} = n \text{ slots}] = (1-\alpha)\alpha^{n-1};$$
$$\text{Prob}[\text{1-time} = n \text{ slots}] = (1-\beta)\beta^{n-1}, \quad n \geq 1.$$

Note that this assumption implies there is a first-order Markovian correlation in the state of the source, meaning that the probability that the source is in state 0 or state 1 in any given slot is fully determined by the state of the source in the previous slot. In particular, if the source is in state 0 during a slot, it will remain in state 0 with probability α or turn to state 1 with probability $1-\alpha$ during the next slot; if the source is in state 1 during a slot, it will remain in state 1 with probability β or turn to state 0 with probability $1-\beta$ during the next slot. The case of uncorrelated source states from slot to slot corresponds to $\gamma = \alpha + \beta - 1 = 0$, where γ is the coefficient of correlation between the source states in two consecutive slots in the steady state. The number of packet arrivals during a slot has an arbitrary distribution which depends only on the source state during the slot. We denote the probability mass functions (pmf's) of the numbers of arrivals during an arbitrary slot where the source state is 0 or 1 by $a_0(n)$ or $a_1(n)$, i.e.,

$$a_m(n) \stackrel{\triangle}{=} \text{Prob}[n \text{ arrivals in a slot where the source state is } m], \quad n \geq 0, \; m = 0, 1,$$

and the corresponding pgf's by $A_0(z)$ and $A_1(z)$, respectively. Moreover, the service and arrival processes are assumed to be mutually independent.

Finally, we assume that the queueing system can reach a steady state, i.e., we assume

$$\rho = \left[p_0 \; A_0^{'}(1) + p_1 \; A_1^{'}(1)\right] s/c < 1.$$

Here ρ denotes the load of the system, p_0 and p_1 denote the probabilities that the source is in state 0 or state 1, respectively, during an arbitrary slot in the steady state:

$$p_0 = \frac{1-\beta}{2-\alpha-\beta} = \frac{1-\beta}{1-\gamma};$$
$$p_1 = \frac{1-\alpha}{2-\alpha-\beta} = \frac{1-\alpha}{1-\gamma},$$

and $A_0^{'}(1)$ and $A_1^{'}(1)$ are the average arrival rates of packets when the source state is 0 or 1, respectively.

3 SYSTEM CONTENTS AND PARTIAL SYSTEM CONTENTS

Let us denote by v_k the system contents (i.e., the total number of packets in the buffer system, including the packets under transmission, if any) at the beginning of slot k, by a_k the number of packet arrivals during slot k, and by t_k the state of the source during slot k. Furthermore, let $u_{j,k}$ be the partial system contents of degree j at slot k, i.e., the number of packets in the system at the beginning of slot k whose service has progressed for at most j slots at the end of slot k. Note that no packets in the system at the beginning of a slot have received more than s slots of service at the end of the slot due to the constant nature of the service times. Then, we have

$$v_k = u_{s,k}; \tag{1}$$

$$u_{j,k+1} = u_{j-1,k} + a_k, \quad 1 \leq j \leq s, \tag{2}$$

$$u_{0,k} = (u_{s,k} - c)^+, \tag{3}$$

where $(\cdots)^+ = max(0, \cdots)$. Indeed, the right-hand side of (3) is the queue length at the beginning of slot k, i.e., the number of packets present in the system at the beginning slot k whose service has not yet started by the end of the slot. In the steady state, the distributions of the above random variables become independent of the time index k. We denote by $V(z)$ and $U_j(z)$ the pgf's of the random variables v_k and $u_{j,k}$ respectively when steady state is reached.

Let us now define the joint pgf of the random variables $(t_k, u_{j,k})$ as

$$Y_{j,k}(x,z) \stackrel{\triangle}{=} E[x^{t_k} z^{u_{j,k}}] = \sum_{m=0}^{1} \sum_{n=0}^{\infty} \text{Prob}[t_k = m, u_{j,k} = n] x^m z^n. \tag{4}$$

Using system equation (2), we then obtain

$$Y_{j,k+1}(x,z) = E[x^{t_{k+1}} z^{a_k} z^{u_{j-1,k}}], \quad 1 \le j \le s. \quad (5)$$

From the arrival process description in Section 2, it follows that $\{t_k\}$ is a homogeneous two-state Markov chain and the distribution of a_k depends only on the value of t_k. More specifically, the joint pgf of the random variables (t_{k+1}, a_k) can be written in terms of the pgf of the random variable t_k:

$$E[x^{t_{k+1}} z^{a_k}] = T_0(x,z)\ E\left[\left(\frac{T_1(x,z)}{T_0(x,z)}\right)^{t_k}\right], \quad (6)$$

where

$$T_0(x,z) = [\alpha + (1-\alpha)x]A_0(z);$$
$$T_1(x,z) = [1-\beta+\beta x]A_1(z).$$

Combining equations (4)-(6), when steady state is reached, we have

$$Y_j(x,z) = T_0(x,z)\ Y_{j-1}\left(\frac{T_1(x,z)}{T_0(x,z)}, z\right), \quad 1 \le j \le s. \quad (7)$$

Next, let us introduce the following partial pgf's:

$$Y_{j;m}(z) \triangleq \lim_{k\to\infty} \sum_{n=0}^{\infty} \mathrm{Prob}[u_{j,k} = n, t_k = m]\ z^n.$$

Then the function $Y_j(x,z)$ is expressed as

$$Y_j(x,z) = Y_{j;0}(z) + xY_{j;1}(z). \quad (8)$$

Substitution of (8) in the functional equation (7) and identification of the coefficients of equal powers of x on both sides of the resulting equation then yields the following set of two recursive equations for $Y_{j;0}(z)$ and $Y_{j;1}(z)$ (expressed in a matrix form):

$$\begin{bmatrix} Y_{j;0}(z) \\ Y_{j;1}(z) \end{bmatrix} = \begin{bmatrix} \alpha A_0(z) & (1-\beta)A_1(z) \\ (1-\alpha)A_0(z) & \beta A_1(z) \end{bmatrix} \cdot \begin{bmatrix} Y_{j-1;0}(z) \\ Y_{j-1;1}(z) \end{bmatrix}, \ 1 \le j \le s. \quad (9)$$

By repeated use of equation (9), we find

$$\begin{bmatrix} Y_{j;0}(z) \\ Y_{j;1}(z) \end{bmatrix} = M^j \begin{bmatrix} Y_{0;0}(z) \\ Y_{0;1}(z) \end{bmatrix}, \quad (10)$$

where

$$M^j = \begin{bmatrix} M_j^{00}(z) & M_j^{01}(z) \\ M_j^{10}(z) & M_j^{11}(z) \end{bmatrix} \triangleq \begin{bmatrix} \alpha A_0(z) & (1-\beta)A_1(z) \\ (1-\alpha)A_0(z) & \beta A_1(z) \end{bmatrix}^j.$$

In a similar way, now using the system equations (1) and (3), we get

$$\begin{bmatrix} Y_{0;0}(z) \\ Y_{0;1}(z) \end{bmatrix} = z^{-c}\left\{\begin{bmatrix} Y_{s;0}(z) \\ Y_{s;1}(z) \end{bmatrix} + \begin{bmatrix} \sum_{n=0}^{c-1} v_{n0}(z) \\ \sum_{n=0}^{c-1} v_{n1}(z) \end{bmatrix}\right\}, \quad (11)$$

where

$$v_{nm}(z) \triangleq v(n,m)(z^c - z^n);$$
$$v(n,m) \triangleq \lim_{k\to\infty} \mathrm{Prob}[v_k = n, t_k = m] = \lim_{k\to\infty} \mathrm{Prob}[u_{s,k} = n, t_k = m],$$
$$m = 0,1; \quad 0 \le n \le c-1.$$

Combination of equations (10) and (11) finally gives

$$z^c \begin{bmatrix} Y_{j;0}(z) \\ Y_{j;1}(z) \end{bmatrix} = M^j \left\{\begin{bmatrix} Y_{s;0}(z) \\ Y_{s;1}(z) \end{bmatrix} + \begin{bmatrix} \sum_{n=0}^{c-1} v_{n0}(z) \\ \sum_{n=0}^{c-1} v_{n1}(z) \end{bmatrix}\right\}, \quad 0 \le j \le s. \quad (12)$$

The entries of the matrix M^j can be expressed in terms of the 2 eigenvalues λ_1 and λ_2 of the matrix M, by using the property that ${\lambda_1}^j$ and ${\lambda_2}^j$ are the 2 eigenvalues of the matrix M^j, as follows:

$$M_j^{00}(z) = \frac{\lambda_1^{j+1} - \lambda_2^{j+1} + \beta A_1(z)(\lambda_2^j - \lambda_1^j)}{\lambda_1 - \lambda_2};$$
$$M_j^{01}(z) = \frac{(1-\beta)A_1(z)(\lambda_1^j - \lambda_2^j)}{\lambda_1 - \lambda_2};$$
$$M_j^{10}(z) = \frac{(1-\alpha)A_0(z)(\lambda_1^j - \lambda_2^j)}{\lambda_1 - \lambda_2};$$
$$M_j^{11}(z) = \frac{\lambda_1^{j+1} - \lambda_2^{j+1} + \alpha A_0(z)(\lambda_2^j - \lambda_1^j)}{\lambda_1 - \lambda_2},$$

where

$$\lambda_\tau = \frac{\alpha A_0(z) + \beta A_1(z)}{2} \pm \frac{\sqrt{[\alpha A_0(z) + \beta A_1(z)]^2 - 4\gamma A_0(z)A_1(z)}}{2}, \quad \tau = 1,2,$$

with $\pm$ being $+$ for $\tau = 1$ and $-$ for $\tau = 2$. Note that λ_1 and λ_2 are functions of z. However, we write λ_τ instead of $\lambda_\tau(z)$ to ease the notation. When $j = s$, (12) leads to a set of linear equations for $Y_{s;0}(z)$ and $Y_{s;1}(z)$, from which the partial pgf's $Y_{s;0}(z)$ and $Y_{s;1}(z)$, as well as the pgf of the system contents $V(z) = Y_{s;0}(z) + Y_{s;1}(z)$ can be calculated. Substitution of the results for $Y_{s;0}(z)$ and $Y_{s;1}(z)$ in (12) moreover enables the calculation of the pgf $U_j(z) = Y_{j;0}(z) + Y_{j;1}(z)$ of the partial system contents of degree j, $0 \le j \le s$. As a result, the following expressions are obtained:

$$V(z) = \frac{1}{\lambda_1 - \lambda_2} \sum_{n=0}^{c-1} \left\{ \left[\frac{\lambda_1 \lambda_2^s}{z^c - \lambda_2^s} - \frac{\lambda_2 \lambda_1^s}{z^c - \lambda_1^s} \right] v_n(z) + \frac{z^c(\lambda_1^s - \lambda_2^s)[A_0(z)v_{n0}(z) + A_1(z)v_{n1}(z)]}{(z^c - \lambda_1^s)(z^c - \lambda_2^s)} \right\}; \tag{13}$$

$$U_0(z) = z^{-c} \left[V(z) + \sum_{n=0}^{c-1} v(n)(z^c - z^n) \right]; \tag{14}$$

$$U_j(z) = \frac{\lambda_1^j - \lambda_2^j}{\lambda_1^s - \lambda_2^s}\, V(z) - \frac{\lambda_1^j \lambda_2^s - \lambda_2^j \lambda_1^s}{\lambda_1^s - \lambda_2^s}\, U_0(z), \quad 0 \le j \le s, \tag{15}$$

where

$$v_n(z) \triangleq v_{n0}(z) + v_{n1}(z) = v(n)(z^c - z^n);$$
$$v(n) \triangleq v(n,0) + v(n,1), \quad 0 \le n \le c-1.$$

In order to determine $V(z)$ completely, we need to find the $2c$ unknown constants $v(n,0)$ and $v(n,1)$ $(0 \le n \le c-1)$ in (13). These can be obtained by invoking the analyticity of the pgf $V(z)$ inside the unit disk $(z : |z| < 1)$ of the complex z-plane and the normalization condition $V(1) = 1$. Specifically, by means of Rouché's theorem (Kleinrock (1975)), it can be shown that the factor $(z^c - \lambda_1^s)(z^c - \lambda_2^s)$ in the denominator of $V(z)$ has exactly $2c-1$ roots inside the unit disk. We denote these roots by z_i, $1 \le i \le 2c-1$. Since $V(z)$ is analytic for $|z| < 1$, the numerator of $V(z)$ must also be zero at these points. Thus, we have

$$(\lambda_1^s - \lambda_2^s) z^c \sum_{n=0}^{c-1} \Big\{ [\lambda_1 \delta(z^c - \lambda_2^s) + \lambda_2 \delta(z^c - \lambda_1^s)] v_n(z) - A_0(z) v_{n0}(z) - A_1(z) v_{n1}(z) \Big\} \Big|_{z=z_i} = 0, \quad 1 \le i \le 2c-1, \tag{16}$$

where $\delta(\cdot)$ is the Kronecker delta function, which is 1 when its argument is zero and 0 otherwise. From the normalization condition $V(1) = 1$ and equation (13), we moreover find that

$$\sum_{n=0}^{c-1} (c-n) v(n) = c - s\lambda_1'(1) = c(1-\rho), \tag{17}$$

where $\lambda_1'(1) = p_0 A_0'(1) + p_1 A_1'(1)$ is the first-order derivative of λ_1 at $z = 1$, which also denotes the mean number of packet arrivals during an arbitrary slot. With equations (16) and (17), the constants $v(n,0)$ and $v(n,1)$ $(0 \le n \le c-1)$ can be calculated.

Once $V(z)$ is determined, some important performance measures for the system, such as the mean value, the variance and the tail distribution of the system contents, can be calculated. The mean system contents $E[v]$ can be obtained by taking the first-order derivative of equation (13) with respect to z in $z = 1$. Using de l'Hospital's rule twice, we get

$$\begin{aligned} E[v] &= V'(1) \\ &= \frac{\sum_{n=0}^{c-1} [A_0'(1)\ v(n,0) + A_1'(1)\ v(n,1)](c-n)}{c\ (1-\gamma)(1-\rho)} \\ &\quad - \frac{\rho c}{s(1-\gamma)} + \frac{s\lambda_1''(1) + \sum_{n=0}^{c-1}(c^2 - n^2) v(n)}{2c(1-\rho)} \\ &\quad - \frac{c(1-\rho)}{2} + \frac{\rho(s - c\rho)}{2s(1-\rho)}, \end{aligned}$$

where $\lambda_1''(1)$ is the second-order derivative of λ_1 with respect to z at $z = 1$. Higher-order moments of the system contents can be derived in a similar way. For instance, the variance of the system contents follows from the relation

$$Var[v] = V''(1) + V'(1) - V'(1)^2.$$

Another important performance characteristic for a buffer is the tail distribution of the system contents, i.e., the probability that the system contents equals a given value n, for sufficiently large n. In principle, the tail distribution of a discrete random variable can be determined by applying the inversion formula for z-transforms and Cauchy's residue theorem from complex analysis (see e.g. Kleinrock (1975)) on its generating function and keeping only the contribution of the pole (or poles) of the pgf with smallest modulus outside the unit disk. As argued in Bruneel and Kim (1993), the system-contents distribution exhibits a geometric tail behavior. That is, for sufficiently large values of n, the tail distribution of the system contents can be approximated as

$$\mathrm{Prob}[v = n] \approx -C_v\ z_v^{-n-1}, \tag{18}$$

where z_v is the pole of $V(z)$ with the smallest modulus (outside the unit disk), and the constant C_v is the residue of $V(z)$ at $z = z_v$. The dominant pole z_v must necessarily be real and positive in order to ensure that the tail distribution is nonnegative anywhere (Bruneel and Kim (1993)). From (13), it follows that z_v is a real positive zero of the denominator of $V(z)$. The residue C_v can be calculated from (13) as

$$C_v = \begin{cases} \frac{\sum_{n=0}^{c-1}\{A_0(z)v_{n0}(z)+A_1(z)v_{n1}(z)-\lambda_2 v_n(z)\}}{(\lambda_1-\lambda_2)\ [c/z-s\lambda_1'(z)/\lambda_1]}\Big|_{z=z_v}, & \text{when } {z_v}^c=\lambda_1(z_v)^s; \\ \frac{\sum_{n=0}^{c-1}\{A_0(z)v_{n0}(z)+A_1(z)v_{n1}(z)-\lambda_1 v_n(z)\}}{(\lambda_2-\lambda_1)\ [c/z-s\lambda_2'(z)/\lambda_2]}\Big|_{z=z_v}, & \text{when } {z_v}^c=\lambda_2(z_v)^s. \end{cases} \tag{19}$$

From (18), the probability that the system contents exceeds a given threshold N, for large N, follows as

$$\text{Prob}[v > N] \approx -C_v\ \frac{z_v^{-N-1}}{z_v-1}.$$

This probability (for an infinite buffer model) is often used to estimate the packet loss probability or buffer overflow probability that would be observed in case of a buffer with a finite storage capacity N (see e.g. Bisdikian et al. (1993)).

4 PACKET DELAY

The delay of a packet is defined as the total number of slots between the end of the slot during which the packet arrives in the system and the end of the slot where the packet finishes its transmission and leaves the system. Let $D(z)$ be the pgf of the delay d that an arbitrary packet experiences in the system. In this section, we analyze the characteristics of the packet delay by means of a general relationship between partial system contents and packet delay established in Gao et al. (2005). Specifically, it has been shown in Gao et al. (2005) that for any discrete-time multiserver system with constant service times of multiple slots and a FCFS queueing discipline, the pgf $D(z)$ can be expressed in terms of the pgf's of the partial system contents as

$$D(z^c) = \frac{1-z^c}{c\lambda_1'(1)}\sum_{j=0}^{c-1}\frac{\theta^j z^s}{(1-\theta^j z^s)^2}\cdot \sum_{i=0}^{s-1} z^{ci}\Big[U_{s-i-1}(\theta^j z^s)-U_{s-i}(\theta^j z^s)\Big], \tag{20}$$

where $\theta = \exp(2\pi I/c)$ with $I^2=-1$. The relationship (20) holds regardless of the exact nature of the arrival process, and therefore it can also be applied to derive the delay characteristics for the considered system with a two-state (first-order Markovian) correlated traffic source. Combination of (20) and (13)-(15) finally gives

$$D(z^c) = \frac{1-z^c}{c\lambda_1'(1)}\sum_{j=0}^{c-1}\frac{\theta^j z^s}{(1-\theta^j z^s)^2(z^c-\lambda_1)(z^c-\lambda_2)} \cdot \sum_{n=0}^{c-1}\Big\{(z^c+\lambda_1\lambda_2-\lambda_1-\lambda_2)v_n(\theta^j z^s)+(1-z^c) \cdot \Big[A_0(\theta^j z^s)v_{n0}(\theta^j z^s)+A_1(\theta^j z^s)v_{n1}(\theta^j z^s)\Big]\Big\}. \tag{21}$$

Note that in (21) λ_1 and λ_2 are functions of $\theta^j z^s$, i.e., functions $\lambda_1(\theta^j z^s)$ and $\lambda_2(\theta^j z^s)$.

The mean value of the packet delay can be found from (21) by evaluation of the first-order derivative of the pgf $D(z^c)$ with respect to z at $z=1$. Specifically, we get

$$E[d] = D'(1) = \frac{1}{c}\frac{dD(z^c)}{dz}\Big|_{z=1} = \frac{E[v]}{\lambda_1'(1)},$$

in agreement with Little's theorem. In a similar way, we can also obtain higher-order moments of the packet delay, by calculating the appropriate higher-order derivatives of $D(z^c)$ at $z=1$. For instance, the variance of the packet delay (delay jitter) can be obtained as

$$\begin{aligned} Var[d] &= D''(1)+D'(1)-D'(1)^2 \\ &= \frac{1}{c^2}\frac{d^2D(z^c)}{dz^2}\Big|_{z=1}+\frac{1}{c}D'(1)-D'(1)^2. \end{aligned}$$

In order to derive the tail distribution of the delay of a packet, we use a similar procedure as for the system contents. However, from expression (21) for $D(z^c)$, we note that this function does not satisfy the condition that it has only one pole with minimal modulus. Indeed, if z_v is the dominant pole of $V(z)$, i.e., the zero of $[z^c-\lambda_1(z)^s]\cdot[z^c-\lambda_2(z)^s]$ outside the unit disk with the smallest modulus, then $z_d(0) \triangleq z_v^{1/s}$ is the zero with minimal modulus outside the unit disk of the factor $[z^c-\lambda_1(z^s)][z^c-\lambda_2(z^s)]$ in the denominator of $D(z^c)$. Due to $\theta^{mc}=1$ for any integer value of m, z^c remains unchanged when z is multiplied by θ^{-m}, and therefore $z_d(m) = \theta^{-m}z_v^{1/s}$ $(0 \le m \le c-1)$ is also a pole of $D(z^c)$ with the same modulus ${z_v}^{1/s}$. In particular, it can be shown that the pole $z_d(m)$ is a zero of the factor $[z^c-\lambda_1(\theta^j z^s)][z^c-\lambda_2(\theta^j z^s)]$ in the denominator of $D(z^c)$ for which $j=(ms)$ mod c, i.e., for which j equals the remainder of the division of ms by c. Taking into account all the poles $z_d(m)$, $0 \le m \le c-1$, and keeping in mind that Prob$[d=n]$ is the coefficient of z^{cn} in the series expansion of $D(z^c)$, we finally get

$$\begin{aligned}\text{Prob}[d=n] &\approx -\sum_{m=0}^{c-1} \frac{b_m}{z_d(m)} \left[z_d(m)\right]^{-cn} \\ &= -\sum_{m=0}^{c-1} \frac{b_m}{z_d(m)} \; z_v^{-cn/s} \\ &= -C_d \; {z_v}^{-cn/s},\end{aligned} \tag{22}$$

for sufficiently large n. In (22), b_m is the residue of $D(z^c)$ at the point $z = z_d(m)$ and is given by

$$b_m = \frac{N_m(z_d(m))}{{R_m}'(z_d(m))},$$

where $N_m(z)$ and $R_m(z)$ are the numerator and the denominator, respectively, of the term in (21) corresponding to the index value $j = (ms) \bmod c$. Using the expressions (21) and (19), we find

$$\begin{aligned}C_d &= \sum_{m=0}^{c-1} \frac{b_m}{z_d(m)} \\ &= \frac{z_v^{-c/s}}{\lambda_1'(1)} \left(\frac{1-z_v^{c/s}}{1-z_v}\right)^2 C_v.\end{aligned}$$

The probability that the packet delay exceeds a given threshold T follows from (22) as

$$\text{Prob}[d > T] \approx -C_d \frac{z_v^{-cT/s}}{{z_v}^{c/s}-1}.$$

5 NUMERICAL RESULTS

We now present a number of numerical examples in order to illustrate the influence of various parameters of the model, such as the degree of correlation in the arrival process, the number of servers and the length of the service times, on the system behavior. Throughout this section, we assume that the packet arrivals during states 0 and 1 are governed by the sets of distributions shown in Table 1. In the first set, packet arrivals are governed by a geometric distribution with arrival rate λ during state 0 and there are no packet arrivals when the source is in state 1. In the second set, packet arrivals are governed by a Bernoulli distribution with rate λ during state 0 and a geometric distribution with rate 2λ during state 1. In the third set, the arrival distributions are of the same type as for the second set, but with the same arrival rate λ during both states.

Table 1: The Three Sets of Arrival Distributions

Set	1	2	3
$A_0(z)$	$\frac{1}{1+\lambda-\lambda z}$	$1-\lambda+\lambda z$	$1-\lambda+\lambda z$
$A_1(z)$	1	$\frac{1}{1+2\lambda-2\lambda z}$	$\frac{1}{1+\lambda-\lambda z}$

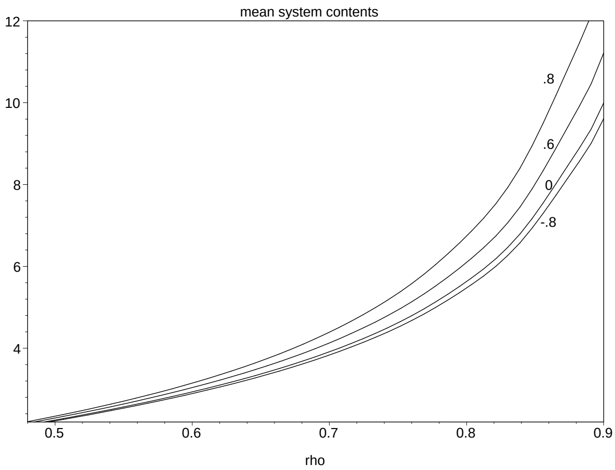

Figure 1: Mean System Contents vs. Load ρ.

In Fig.1, we have plotted the mean system contents versus the load ρ, for $c = 4,\ s = 4,\ \alpha = \beta$, arrival distributions of set 2, and various values of the source state correlation coefficient γ, namely $\gamma = -0.8,\ 0,\ 0.6,\ 0.8$. The figure clearly shows that for a given ρ, the mean system contents increases as γ increases. Especially, for higher loads ρ, the system contents may be heavily underestimated when the (positive) correlation between the source states in two consecutive slots is not taken into account.

In Figs.2-4, we assume $\alpha = 0.7$ and $\beta = 0.8$. The source state correlation coefficient γ then equals 0.5. In Fig.2, the overflow probability Prob$[v > N]$ is shown as a function of N, for $\rho = 0.8$, $c = 4$, $s = 8$ and the three sets of arrival distributions. We note that the first set gives the highest overflow probability, while the third set gives the smallest value. This observation can be understood intuitively from the fact that the variance of the number of arrivals per slot decreases in the order of set 1, set 2 and set 3. Indeed, the higher the variance of the number of arrivals and, hence, the more fluctuation of the arrival process, the higher we expect the buffer contents to be. The required buffer size N to satisfy a given loss bound can also be estimated from Fig.2.

In Fig.3, the mean packet delay is plotted versus the load ρ, for the arrival distributions of set 1, for $s = 1,\ 3,\ 5$ and $c = 4,\ 8$. For given values of c and s, we see that the mean packet delay increases as ρ increases. For a given ρ, the mean delay increases as the service times become longer and/or the number of servers decreases. We also observe that the longer

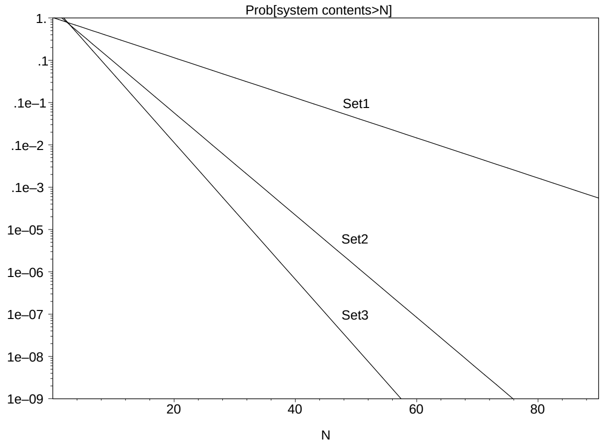

Figure 2: Prob$[v > N]$ vs. N.

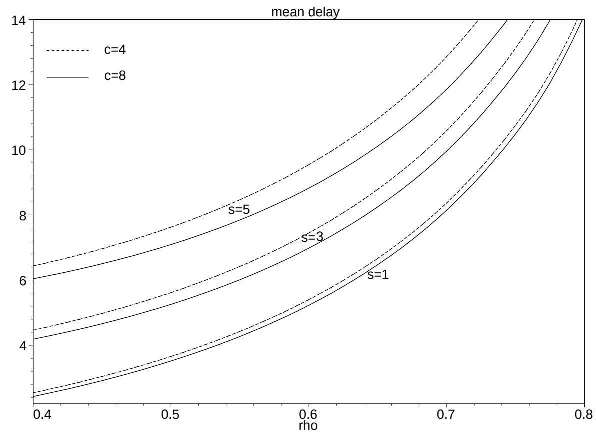

Figure 3: Mean Packet Delay vs. Load ρ.

the service times, the higher the impact of the number of servers on the packet delay, especially when the load gets higher.

In Fig.4, the variance of the packet delay is shown versus ρ, for the arrival distributions of set 3, for $s = 8$ and $c = 1,\ 4,\ 8$. Clearly, for a given value of ρ, the delay jitter decreases as the number of servers increases.

6 CONCLUSIONS

In this paper, we have studied the behavior of a discrete-time infinite-capacity buffer system with multiple servers and constant service times of multiple slots. Packets are generated by a two-state traffic source with a first-order Markovian correlation in the state of the source. We have presented an analytical technique based on generating functions for the analysis of the system. As a result, closed-form expressions have been derived for such performance measures as the mean values, the variances and the tail distributions of the system contents and the packet delay. Some numerical results have been presented to illustrate the analysis. The results indicate that

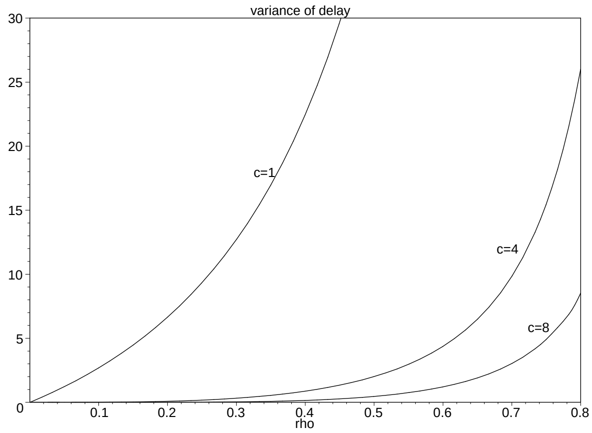

Figure 4: Variance of the Packet Delay vs. Load ρ.

the characteristics of the system contents and the packet delay are sensitive to both the arrival process and the service mechanism.

References

Bisdikian, C., Lew, J., and Tantawi, A. (1993). On the tail approximation of the blocking probability of single server queues with finite buffer capacity. In *Proceedings of the Second International Conference on Queueing Networks with Finite Capacity*, pages 267–80, Research Triangle Park.

Bruneel, H. and Kim, B. (1993). *Discrete-Time Models for Communication Systems Including ATM*. Kluwer Academic Publishers, Boston.

Bruneel, H., Steyaert, B., Desmet, E., and Petit, G. (1992). An analytical technique for the derivation of the delay performance of ATM switches with multiserver output queues. *International Journal of Digital and Analog Communication Systems*, 5:193–201.

Bruneel, H. and Wuyts, I. (1994). Analysis of discrete-time multiserver queueing models with constant service times. *Operations Research Letters*, 15:231–6.

Gao, P., Wittevrongel, S., and Bruneel, H. (2003). Delay against system contents in discrete-time G/Geom/c queue. *Electronics Letters*, 39(17):1290–1292.

Gao, P., Wittevrongel, S., and Bruneel, H. (2004a). Delay analysis for a discrete-time GI-D-c queue with arbitrary-length service times. In *Proceedings of EPEW 2004*, Lecture Notes in Computer Science 3236, pages 184–195, Toledo.

Gao, P., Wittevrongel, S., and Bruneel, H. (2004b). Discrete-time multiserver queues with geometric service times. *Computers & Operations Research*, 31(1):81–99.

Gao, P., Wittevrongel, S., and Bruneel, H. (2004c). On the behavior of multiserver buffers with geometric service times and bursty input traffic. *IEICE Transactions on Communications*, E87-B(12):3576–3583.

Gao, P., Wittevrongel, S., and Bruneel, H. (2005). Relationship between delay and partial system contents in multiserver queues with constant service times. In *Booklet of Abstracts of ORBEL 19*, pages 72–74, Louvain-la-Neuve.

Kleinrock, L. (1975). *Queueing Systems, Volume I: Theory*. Wiley, New York.

Rubin, I. and Zhang, Z. (1991). Message delay and queue-size analysis for circuit-switched TDMA systems. *IEEE Transactions on Communications*, 39:905–14.

Wittevrongel, S. and Bruneel, H. (1999). Discrete-time queues with correlated arrivals and constant service times. *Computers & Operations Research*, 26(2):93–108.

AUTHOR BIOGRAPHIES

PEIXIA GAO has been working as a Ph.D. student at the Stochastic Modeling and Analysis of Communication Systems (SMACS) Research Group, Department of Telecommunication and Information Processing (TELIN), Faculty of Applied Sciences, Ghent University, Belgium since 2000. Her main research interests include discrete-time queueing theory, performance evaluation and stochastic modeling of communication networks.
Her e-mail address is: pg@telin.ugent.be
Her Web-page is: http://telin.ugent.be/∼pg

SABINE WITTEVRONGEL was born in Gent, Belgium, in 1969. She received the M.S. degree in Electrical Engineering and the 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 TELIN, Faculty of Applied Sciences, Ghent University, first in the framework of various projects, from October 1994 to September 2001, as a researcher of the Fund for Scientific Research - Flanders (Belgium) (F.W.O.), and since October 2001 as a full time Professor. Her main research interests include discrete-time queueing theory, the performance evaluation of communication networks, the study of traffic control mechanisms, and the analysis of ARQ protocols.
Her e-mail address is: sw@telin.ugent.be
Her Web-page is: http://telin.ugent.be/∼sw

HERWIG BRUNEEL was born in Zottegem, Belgium, in 1954. He received the M.S. degree in Electrical Engineering, the degree of Licentiate in Computer Science, and the Ph.D. degree in Computer Science in 1978, 1979 and 1984 respectively, all from Ghent University, Belgium. He is full time Professor in the Faculty of Applied Sciences and head of the Department TELIN at the same university. He also leads the SMACS Research Group within this department. His main personal research interests include stochastic modeling and analysis of communication systems, discrete-time queueing theory, and the study of ARQ protocols. He has published more than 200 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). His e-mail address is: hb@telin.ugent.be