A batch-service queueing model with a discrete batch Markovian arrival process

Dieter Claeys, Joris Walraevens, Koenraad Laevens, Bart Steyaert, and Herwig Bruneel

Stochastic Modelling and Analysis of Communication Systems (SMACS) Research Group, Department of Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Tel.: +32 9 264 3411
Fax: +32 9 264 4295
Dieter.Claeys@telin.ugent.be

Abstract. Queueing systems with batch service have been investigated extensively during the past decades. However, nearly all the studied models share the common feature that an uncorrelated arrival process is considered, which is unrealistic in several real-life situations. In this paper, we study a discrete-time queueing model, with a server that only initiates service when the amount of customers in system (system content) reaches or exceeds a threshold. Correlation is taken into account by assuming a discrete batch Markovian arrival process (D-BMAP), i.e. the distribution of the number of customer arrivals per slot depends on a background state which is determined by a first-order Markov chain. We deduce the probability generating function of the system content at random slot marks and we examine the influence of correlation in the arrival process on the behavior of the system. We show that correlation merely has a small impact on the threshold that minimizes the mean system content. In addition, we demonstrate that correlation might have a significant influence on the system content and therefore has to be included in the model.

Keywords: queueing systems and networks, performance modelling

1 Introduction

Whereas servers in traditional queueing systems serve individual customers, batch servers process batches consisting of several customers. Batch servers appear for instance as elevators in high buildings, transport vehicles, recreational devices in amusement parks, ovens in production processes, et cetera. Furthermore, in telecommunications, it is often the case that information packets are aggregated into larger entities (batches) and these batches are transmitted instead of all packets individually. This is mainly done for efficiency reasons, since only one header per aggregated batch has to be constructed instead of one header

per single information unit, thus leading to an increased throughput. Technologies using packet aggregation include Optical burst switched (OBS) networks [10], [25], IEEE 802.11n WLANs [23]. A more general model can be found in [3].

Batch-service queueing models have been studied extensively during the past decades [2], [6]–[9], [11]–[12], [16]–[17], [19], [21], [24], [26], [28]–[29], [31]. However, these surveys share the common feature that only models with an uncorrelated arrival process are considered, which is unrealistic in several real-life situations. For instance, in telecommunications, a traffic source which is inactive in a given time slot is very likely to remain inactive for a long time (or during a large number of time slots). In order to capture most traffic characteristics up to any desired precision, one often adopts the discrete batch Markovian arrival process (D-BMAP) [4], [13], [18], [20], [22], [27], [30]. In [30], the queue length is studied in a multiserver system with finite buffer space and deterministic service times. [4] analyses the queue length at departure instants in a single-server, finite-buffer system with general service times, whereas [18] investigates the queue length at arbitrary instants and the waiting time for this system. [22] examines the queue length for a single-server system with infinite buffer space and general service times. In [27], a single-server, finite-buffer system with vacations is considered. [13] analyses an infinite-buffer, single-server system with vacations, whereas [20] evaluates an infinite-buffer retrial queue. Mark that, although, in theory, self-similar or long-range dependent (LRD) traffic is precluded from Markov modelling, many studies show that this need not be restrictive, see e.g. [1], in realistic scenarios.

The main contribution of this paper is that we study the combination of batch-service and D-BMAP. As a starting point, we consider a model with geometrically distributed service times and a server that solely initiates service if the system contains at least as many customers as a threshold l (more details of the model are given in section 2). We deduce the probability generating function (PGF) of the system content, i.e. the number of customers in the system, those in service included, in section 3. Next, we investigate the influence of correlation on the mean system content in section 4. We demonstrate that ignoring correlation can lead to a severe under- or overestimation of the mean system content and therefore has to be included in the model. In addition, as determining the value of l that minimizes the mean system content is a crucial part in batch service, we pay special attention to the influence of correlation on this issue.

2 Model description

- The time axis is divided into fixed-length slots.
- The queue is infinitely large.
- There is one batch server of capacity c (c a constant), which means that the server can process up to c customers simultaneously. The available server solely starts service if the system contains at least as many customers as the threshold l ($1 \leq l \leq c$). Hence, if the server finds less than l customers upon becoming available, it waits to initiate service until the first slot mark

whereby the system contains at least l customers. We further assume that the already present customers remain in the queue when the server waits to start service. Hence, during each slot, the system content consists of the customers being served (the server content) and the customers waiting in the queue (the queue content).

- A service period is the period between the start and end of the service of one batch of customers. The service can only start and end at slot boundaries, implying that an arriving customer has to wait for service at least until the next slot mark.
- The consecutive service times - a service time is the length of a service period, expressed in a number of slots - are independent and identically distributed (IID) and have a geometric distribution, with mass function $t(n) = (1 - \alpha)\alpha^{n-1}(n \geq 1)$, whereby α $(0 \leq \alpha < 1)$ represents the probability that an ongoing service during a random slot is not finished at the end of that slot. The corresponding PGF is equal to $T(z) = (1 - \alpha)z/(1 - \alpha z)$, and the mean service time $\mathrm{E}\,[T]$ equals $1/(1 - \alpha)$ slots.
- Customers arrive in the buffer according to a homogeneous irreducible D-BMAP. The number of background states is finite and denoted by N. We designate the state during slot k by τ_k and during a random slot by τ. The arrival process is completely defined by the values $a(n, j|i); n \geq 0; i, j \in \{1, \ldots, N\}$, denoting the probability that if the background state is i during a slot, there are n arrivals during this slot and the background state during the next slot is j. We put these probabilities in a matrix generating function $\mathbf{A}(z)$ with dimension $N \times N$, whose entries are defined as follows:

$$[\mathbf{A}(z)]_{ij} \triangleq \sum_{n=0}^{\infty} a(n, j|i)z^n \ .$$

Finally, we define $a_i(n)$ as

$$a_i(n) \triangleq \lim_{k \to \infty} \Pr\,[A_k = n|\tau_k = i] \ ,$$

with A_k the amount of customers arriving in slot k, and

$$A_i(z) \triangleq \sum_{n=0}^{\infty} a_i(n)z^n = [\mathbf{A}(z)\mathbf{1}]_i \ ,$$

with $\mathbf{1}$ a $N \times 1$ column vector whose N entries are equal to 1.

The stability condition requires that the load $\rho \triangleq \frac{\lambda}{c(1-\alpha)} < 1$, whereby λ represents the mean number of customer arrivals during a slot, i.e.

$$\lambda \triangleq \sum_{i=0}^{N} \Pr\,[\tau = i]\,A_i^{'}(1) \ ,$$

with $A_i^{'}(1)$ the mean number of customer arrivals in a slot if the background state equals i (we use primes to indicate derivatives).

3 PGF of the system content

In this section, we deduce the PGF $U(z)$ of the system content at random slot boundaries. Therefore, we first compute the vector generating function $\mathbf{U}(z)$, whose j-th entry equals the partial generating function of the system content at a random slot mark when the background state during the slot equals j. We then have that

$$U(z) = \mathbf{U}(z)\mathbf{1} \ . \tag{1}$$

In order to deduce $\mathbf{U}(z)$, we first compute the vector joint generating function $\mathbf{P}(z, x)$ of the queue content Q and the server content S at random slot boundaries. Then, we find $\mathbf{U}(z)$ by letting $x \to z$ in $\mathbf{P}(z, x)$. Let now Q_k represent the queue content at slot mark k and S_k be the server content at slot boundary k. The slot-by-slot evolution of the pair (queue content, system content) is governed by the following system equations:

$$(Q_{k+1}, S_{k+1}) = \begin{cases} ([Q_k + A_k - c]^+, \min[Q_k + A_k, c]) & \text{if } S_k = 0 \text{ and } Q_k + A_k \geq l \\[2mm] (Q_k + A_k, 0) & \text{if } S_k = 0 \text{ and } Q_k + A_k < l \\[2mm] ([Q_k + A_k - c]^+, \min[Q_k + A_k, c]) & \begin{array}{l}\text{if } S_k > 0, \text{ the service period} \\ \text{finishes at the end of slot } k \\ \text{(with probability } 1 - \alpha) \\ \text{and } Q_k + A_k \geq l\end{array} \\[2mm] (Q_k + A_k, 0) & \begin{array}{l}\text{if } S_k > 0, \text{ the service period} \\ \text{finishes at the end of slot } k \\ \text{(with probability } 1 - \alpha) \\ \text{and } Q_k + A_k < l\end{array} \\[2mm] (Q_k + A_k, S_k) & \begin{array}{l}\text{if } S_k > 0 \text{ and the service} \\ \text{period continues at the end of} \\ \text{slot } k \text{ (with probability } \alpha)\end{array} \end{cases}$$

whereby $(X)^+ \triangleq \max(X, 0)$. Note that, as opposed to the case of an IID arrival process, the sequence $\{(Q_k, S_k)\}_{k \in \mathbb{N}}$ does not constitute a Markov chain anymore. Let us now introduce the $1 \times N$ row vectors $\mathbf{d}_k(n)$ and $\tilde{\mathbf{d}}_k(n)$:

$$[\mathbf{d}_k(n)]_j \triangleq \Pr[Q_k + A_k = n, S_k = 0, \tau_{k+1} = j] \ , \qquad 1 \leq j \leq N \ , \ n \geq 0 \ ,$$

and

$$[\tilde{\mathbf{d}}_k(n)]_j \triangleq \Pr[Q_k + A_k = n, \tau_{k+1} = j] \ , \qquad 1 \leq j \leq N \ , \ n \geq 0 \ .$$

In addition, $\mathbf{P}_k(z, x)$ denotes the $1 \times N$ joint vector generating function of Q_k and S_k, i.e.

$$[\mathbf{P}_k(z, x)]_j \triangleq \mathrm{E}\left[z^{Q_k} x^{S_k} \mathbf{1}_{\{\tau_k = j\}}\right] \ .$$

The above system equations can then be translated into

$$
\begin{aligned}
\mathbf{P}_{k+1}(z,x) =& \alpha \left[\sum_{n=0}^{l-1} \mathbf{d}_k(n)z^n + \sum_{n=l}^{c-1} \mathbf{d}_k(n)x^n + \sum_{n=c}^{\infty} \mathbf{d}_k(n)z^{n-c}x^c \right] \\
& + (1-\alpha) \left[\sum_{n=0}^{l-1} \tilde{\mathbf{d}}_k(n)z^n + \sum_{n=l}^{c-1} \tilde{\mathbf{d}}_k(n)x^n + \sum_{n=c}^{\infty} \tilde{\mathbf{d}}_k(n)z^{n-c}x^c \right] \\
& + \alpha \left[P_k(x,z) - P_k(z,0) \right] \mathbf{A}(z) \ .
\end{aligned}
\tag{2}
$$

Next, taking into account that

$$
\sum_{n=c}^{\infty} \mathbf{d}_k(n)z^{n-c}x^c = \frac{x^c}{z^c} \left[P_k(z,0)\mathbf{A}(z) - \sum_{n=0}^{l-1} \mathbf{d}_k(n)z^n - \sum_{n=l}^{c-1} \mathbf{d}_k(n)z^n \right] \ ,
$$

and

$$
\sum_{n=c}^{\infty} \tilde{\mathbf{d}}_k(n)z^{n-c}x^c = \frac{x^c}{z^c} \left[P_k(z,1)\mathbf{A}(z) - \sum_{n=0}^{l-1} \tilde{\mathbf{d}}_k(n)z^n - \sum_{n=l}^{c-1} \tilde{\mathbf{d}}_k(n)z^n \right] \ ,
$$

going to the steady state and defining $\mathbf{f}(n) \triangleq \lim_{k\to\infty} \left[\alpha \mathbf{d}_k(n) + (1-\alpha)\tilde{\mathbf{d}}_k(n) \right]$, equation (2) transforms into

$$
\begin{aligned}
z^c \mathbf{P}(z,x) \left(I - \alpha \mathbf{A}(z) \right) =& (z^c - x^c) \sum_{n=0}^{l-1} \mathbf{f}(n)z^n + \sum_{n=l}^{c-1} \mathbf{f}(n)(z^c x^n - z^n x^c) \\
& + \alpha \left(x^c - z^c \right) \mathbf{P}(z,0)\mathbf{A}(z) \\
& + (1-\alpha)x^c \mathbf{P}(z,1)\mathbf{A}(z) \ .
\end{aligned}
\tag{3}
$$

Letting $x \to 0$ in (3) produces

$$
\mathbf{P}(z,0) = \sum_{n=0}^{l-1} \mathbf{f}(n)z^n \ .
\tag{4}
$$

Letting $x \to 1$ in (3) and appealing to (4), we obtain

$$
\begin{aligned}
\mathbf{P}(z,1) =& \left[(z^c - 1) \sum_{n=0}^{l-1} \mathbf{f}(n)z^n (\mathbf{I} - \alpha \mathbf{A}(z)) + \sum_{n=l}^{c-1} \mathbf{f}(n)(z^c - z^n) \right] \\
& \times \left[z^c(\mathbf{I} - \alpha \mathbf{A}(z)) - (1-\alpha)\mathbf{A}(z) \right]^{-1} \ ,
\end{aligned}
\tag{5}
$$

with $\mathbf{I}$ the $N \times N$ identity matrix. Finaly, letting $x \to z$ in (3) and relying on (5), we find the expression for $\mathbf{U}(z)$

$$
\begin{aligned}
\mathbf{U}(z) =& \left[(z^c - 1) \sum_{n=0}^{l-1} \mathbf{f}(n) z^n (\mathbf{I} - \alpha \mathbf{A}(z)) + \sum_{n=l}^{c-1} \mathbf{f}(n)(z^c - z^n) \right] \\
& \times \left[z^c (\mathbf{I} - \alpha \mathbf{A}(z)) - (1-\alpha) \mathbf{A}(z) \right]^{-1} (1-\alpha) \mathbf{A}(z)[I - \alpha \mathbf{A}(z)]^{-1} \\
=& \left[(z^c - 1) \sum_{n=0}^{l-1} \mathbf{f}(n) z^n + \sum_{n=l}^{c-1} \mathbf{f}(n)(z^c - z^n)(\mathbf{I} - \alpha \mathbf{A}(z))^{-1} \right] \\
& \times \left[z^c \mathbf{I} - (1-\alpha) \mathbf{A}(z)(\mathbf{I} - \alpha \mathbf{A}(z))^{-1} \right]^{-1} (1-\alpha) \mathbf{A}(z)[\mathbf{I} - \alpha \mathbf{A}(z)]^{-1} \\
=& \left[(z^c - 1) \sum_{n=0}^{l-1} \mathbf{f}(n) z^n + \sum_{n=l}^{c-1} \mathbf{f}(n)(z^c - z^n)(\mathbf{I} - \alpha \mathbf{A}(z))^{-1} \right] \\
& \times \left[z^c \mathbf{I} - T(\mathbf{A}(z)) \right]^{-1} T(\mathbf{A}(z)) \ .
\end{aligned}
\tag{6}
$$

Equation (6) together with (1) produces the PGF $U(z)$ of the system content. More details concerning the computation of the unknowns $\mathbf{f}(n)$ in (6) are mentioned in appendix A.

4 Numerical examples

In this section, we evaluate the influence of combining correlation and batch service on the behavior of the system. To this end, we consider some numerical examples. We assume throughout this section that the number of states N equals 2. We denote the probability that if the background state is i during a slot, the background state remains i during the next slot by $p_i, i = 1, 2$. Note that $p_i = [\mathbf{A}(1)]_{ii}$. In view of the above assumptions, we define the coefficient of correlation γ between the states of two consecutive slots as

$$
\gamma \triangleq \lim_{k \to \infty} \frac{\mathrm{E}\left[\tau_k \tau_{k+1}\right] - \mathrm{E}\left[\tau_k\right] \mathrm{E}\left[\tau_{k+1}\right]}{(\mathrm{Var}\left[\tau_k\right] \mathrm{Var}\left[\tau_{k+1}\right])^{1/2}} = p_1 + p_2 - 1 \ .
$$

In Fig.1, the mean system content $\mathrm{E}\left[U\right]$ is depicted versus the load ρ for several values of γ. The left pane corresponds to $l = 1$, whereas in the right pane $l = c$. It is assumed that $p_1 = p_2$, $a_1(0) = 1$, $a_1(n) = 0$ if $n \geq 1$, $a_2(n) = [1/(1+2\lambda)][2\lambda/(1+2\lambda)]^n$, $c = 10$ and $\mathrm{E}\left[T\right] = 10$. Fig.1 learns us that, regardless of l, positive correlation ($\gamma > 0$) leads to a significant larger $\mathrm{E}\left[U\right]$ as compared to the uncorrelated case ($\gamma = 0$). Hence, disregarding positive correlation can lead to a severe underrating of the mean system content. Fig.1 also exhibits that ignoring negative correlation leads to some overestimation of $\mathrm{E}\left[U\right]$. We further perceive that these observations manifest themselves more as ρ increases.
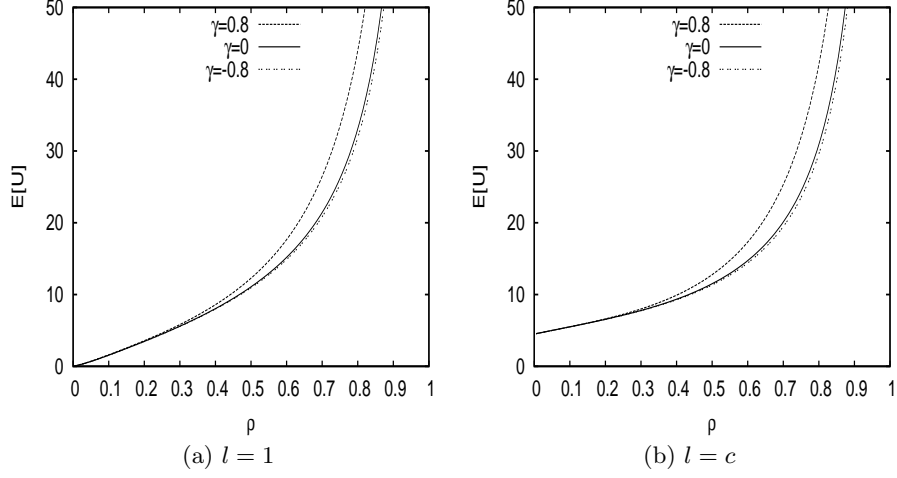
Fig. 1. $\mathrm{E}[U]$ versus ρ for several values of γ; $p_1 = p_2$, $a_1(0) = 1$, $a_1(n) = 0$ if $n \geq 1$, $a_2(n) = [1/(1 + 2\lambda)][2\lambda/(1 + 2\lambda)]^n$, $c = 10$, $\mathrm{E}[T] = 10$

Next, we take a look at the influence of the distributions $a_i(n), i = 1, 2$ on the behavior of the system. Therefore, $\mathrm{E}[U]$ is represented versus ρ in Fig.2 for the sets for $a_1(n)$ and $a_2(n)$ given in Table 1. It is again assumed that $p_1 = p_2$, $c = 10$ and $\mathrm{E}[T] = 10$ and the panes each correspond to a unique combination of γ and l. We notice that, regardless of l, the exact expressions for $a_1(n)$ and $a_2(n)$ have an undeniable impact on $\mathrm{E}[U]$. In general, it appears that a larger variance in the number of customer arrivals leads to a larger $\mathrm{E}[U]$. We also perceive that the influence is more pronounced in case of positive correlation. Note that the preceding conclusions are similar to those in multiserver systems with correlated arrivals (see e.g. [5], [15]).

Table 1. Overview of the different sets of expressions for $a_1(n)$ and $a_2(n)$

	$a_1(n)$	$a_2(n)$
set 1	1 if $n = 0$, else 0	$\frac{1}{1+2\lambda}\left(\frac{2\lambda}{1+2\lambda}\right)^n$
set 2	$\frac{1}{1+5\lambda/3}\left(\frac{5\lambda/3}{1+5\lambda/3}\right)^n$	$e^{-\lambda/3}\frac{(\lambda/3)^n}{n!}$
set 3	$e^{-2\lambda/3}\frac{(2\lambda/3)^n}{n!}$	$e^{-4\lambda/3}\frac{(4\lambda/3)^n}{n!}$

As determining the optimal value of l (i.e. the value that minimizes $\mathrm{E}[U]$) is of the utmost importance in batch-service systems, we study whether correlation affects this optimal value. For this purpose, the optimal value of l is shown versus ρ in Fig.3, for several values of γ. Each pane characterizes a distinct combination of c, $\mathrm{E}[T]$ and expressions for $a_1(n)$ and $a_2(n)$. We perceive that correlation merely has a slight influence on the optimal value of l. In view of this, we now
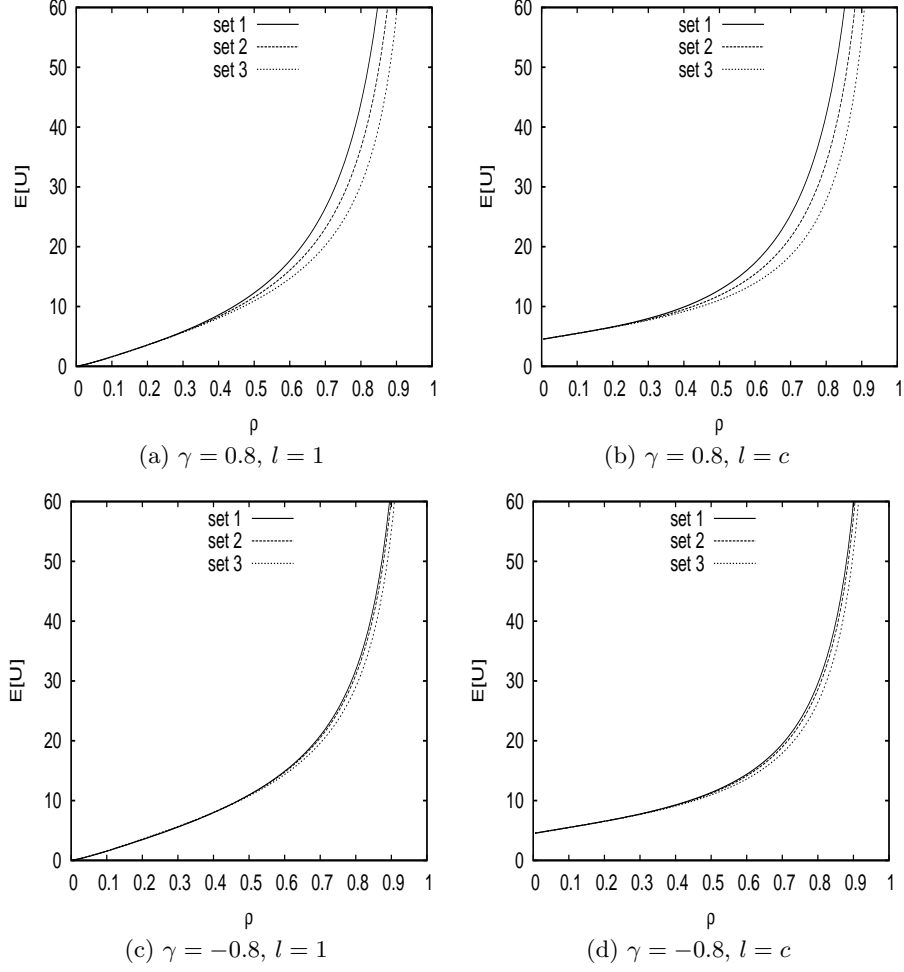
Fig. 2. $E[U]$ versus ρ for the sets of $a_1(n)$ and $a_2(n)$ given in Table 1; $p_1 = p_2$, $c = 10$, $E[T] = 10$; each pane corresponds to a unique combination of γ and l

investigate the impact of adopting l_{opt} of the uncorrelated case in the correlated system. Therefore, we define the relative differences as

$$\frac{E[U]_{l_{\mathrm{opt}}} - E[U]_{\tilde{l}_{\mathrm{opt}}}}{\left(E[U]_{l_{\mathrm{opt}}} + E[U]_{\tilde{l}_{\mathrm{opt}}}\right)/2} \; ,$$

with $E[U]_{l_{\mathrm{opt}}}$ the mean system content in the correlated case when l_{opt} is adopted and $E[U]_{\tilde{l}_{\mathrm{opt}}}$ the mean system content in the correlated system when the optimal value of l of the corresponding uncorrelated system is adopted.

In Fig.4, the relative differences are depicted both in the cases $\gamma = 0.8$ and $\gamma = -0.8$, for the examples from Fig.3. We observe that even when the optimal value of l is different, $E[U]$ is approximated very well. In view of this, the existing results of the corresponding uncorrelated system can be used to determine a near-optimal value of l. Adopting this near-optimal value has nearly no impact on $E[U]$.
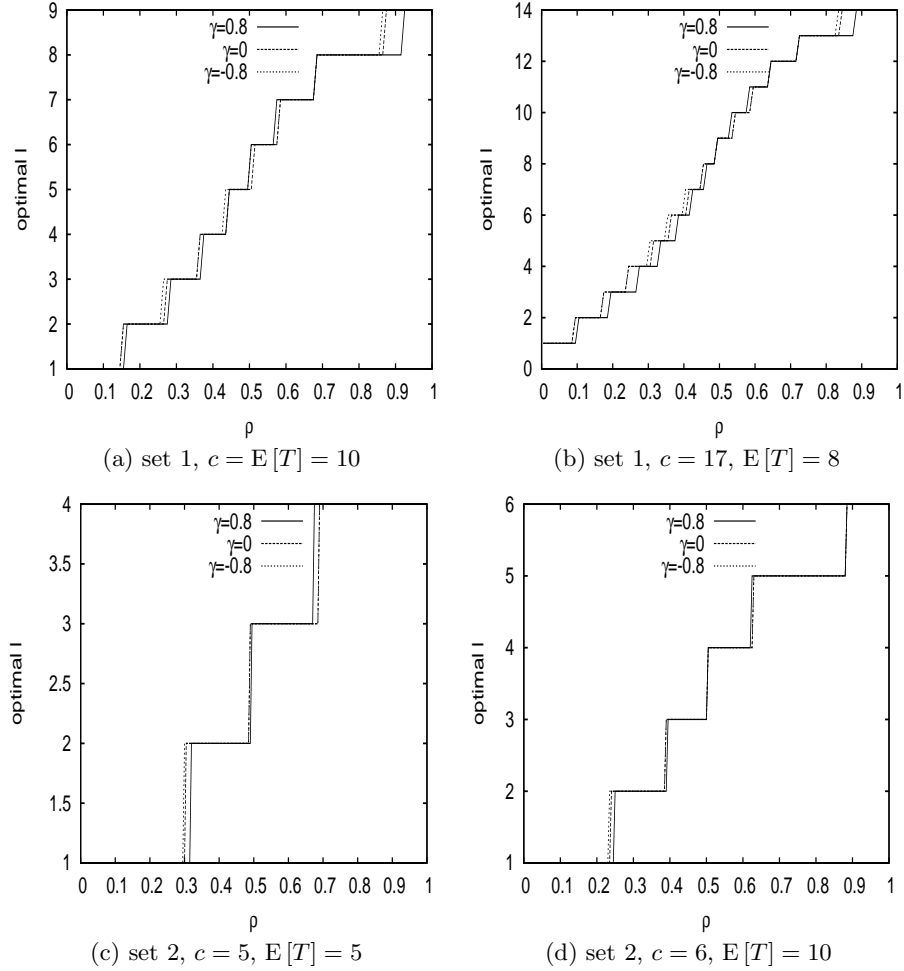

(a) set 1, $c = E[T] = 10$

(b) set 1, $c = 17$, $E[T] = 8$

(c) set 2, $c = 5$, $E[T] = 5$

(d) set 2, $c = 6$, $E[T] = 10$

Fig. 3. The optimal value of l versus ρ for various values of γ; $p_1 = p_2$; each pane represents a specific combination of c, $E[T]$ and expressions for $a_1(n)$ and $a_2(n)$
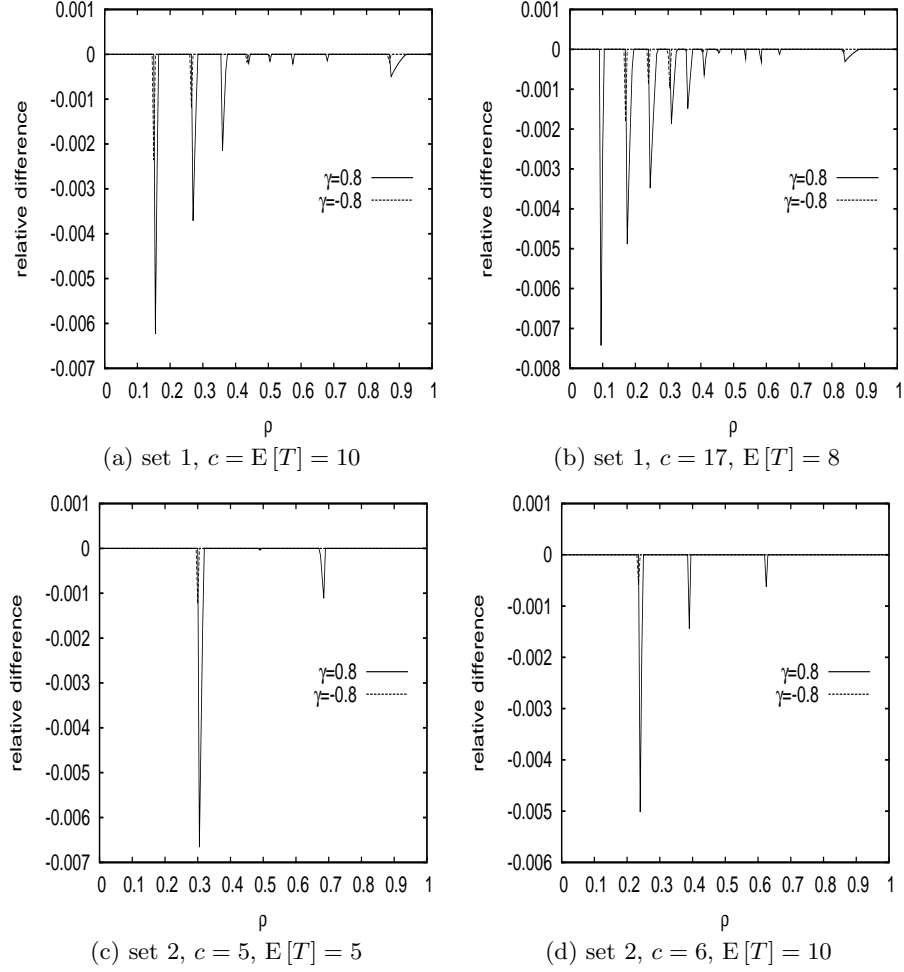
Fig. 4. Relative differences versus ρ for $\gamma = 0.8$ and $\gamma = -0.8$; $p_1 = p_2$; each pane represents a specific combination of c, $\mathrm{E}[T]$ and expressions for $a_1(n)$ and $a_2(n)$

5 Conclusions

In this paper, we have studied a batch-service queueing model with a server that waits to initiate service until at least l customers have accumulated in the system. The most prominent feature is that customers arrive according to a discrete batch Markovian arrival process. Such an arrival process has the advantage that it can capture most real-life characteristics up to any desired precision. We have deduced the probability generating function of the system content (i.e. the number of customers in the system) at arbitrary slot marks. In addition, we have pointed out that correlation merely has a small influence on the value of

l that minimizes the mean system content and consequently, that the existing results of the corresponding uncorrelated system can be applied to determine a near-optimal value of l. Finally, we have demonstrated that disregarding correlation in the arrival process can lead to a dramatic under- or overestimation of the mean system content and therefore cannot be ignored.

In the future, we will investigate whether these conclusions also hold in more general batch-service queueing models, such as models with general service times. We will also examine the influence of correlation on the customer delay.

A   Calculation of the unknowns $\mathbf{f}(n)$ in equation (6)

In this appendix, we demonstrate how the unknowns $\mathbf{f}(n)$ can be determined by means of the spectral decomposition method. Therefore, let $\lambda_i(z)$ denote the i-th eigenvalue of $\mathbf{A}(z)$ and let $\mathbf{r}_i(z)$ represent the corresponding $N \times 1$ right eigenvector of $\mathbf{A}(z)$, i.e.

$$\mathbf{A}(z)\mathbf{r}_i(z) = \lambda_i(z)\mathbf{r}_i(z) \ .$$

Multiplication of both sides of (6) by $\mathbf{r}_i(z)$ then produces

$$\mathbf{U}(z)\mathbf{r}_i(z) = \frac{T(\lambda_i(z))}{z^c - T(\lambda_i(z))}$$
$$\times \left[(z^c - 1)\sum_{n=0}^{l-1} \mathbf{f}(n)z^n + \frac{1}{1 - \alpha\lambda_i(z)}\sum_{n=l}^{c-1} \mathbf{f}(n)(z^c - z^n) \right] \mathbf{r}_i(z) \ , \qquad 1 \leq i \leq N \ . \quad (7)$$

Remark 1. We follow the commonly adopted convention that $\lambda_1(z)$ and $\mathbf{r}_1(z)$ denote the Perron-Frobenius eigenvalue and vector respectively, i.e. $\lambda_1(z)$ is the eigenvalue that satisfies $\lambda_1(1) = 1$.

Unlike the IID case, it is not possible to construct an irrefutable mathematical proof, based on Rouché's theorem, to show that each of the equations $z^c - T(\lambda_i(z)) = 0$, $1 \leq i \leq N$ necessarily has c solutions inside the closed complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$. Nevertheless, an example where this is not the case has not been encountered up to now, and, to the best of our knowledge, such an example, if it exists, has yet to be constructed. So for practical purposes, we can venture to state that the above equation has indeed c solutions inside the closed complex unit disk for each value of i, provided that the equilibrium condition $\rho < 1$ holds.

Let us characterise the k-th solution of the i-th equation by $z_{i,k}$. As $\lambda_1(1) = 1$, one of the solutions of $z^c - T(\lambda_1(z)) = 0$ equals one. Without loss of generality, we let $z_{1,1}$ be that zero. As $\mathbf{U}(z)$ is analytic inside the closed complex unit disk, the numerator of the right-hand-side of (7) must also vanish at these zeroes. This observation leads to $Nc - 1$ linear equations in the $1 \times N$ vectors $\mathbf{f}(n)$, $0 \leq n \leq c - 1$. The zero $z_{1,1}$ cannot be used as it produces the trivial equation $0 = 0$. Fortunately, we can resort to the normalisation condition to obtain another

equation. This condition is found by letting $z \to 1$ in (7) for $i = 1$ and taking into account that $\mathbf{r}_1(1) = \text{span}\{\mathbf{1}\} \triangleq \{\mathbf{x}|\mathbf{x} = K\mathbf{1}, K \in \mathbb{C}\}$, leading to

$$\mathbf{U}(1)\mathbf{1} = 1 = \frac{c \sum_{n=0}^{l-1} \mathbf{f}(n) + \frac{1}{1-\alpha} \sum_{n=l}^{c-1} \mathbf{f}(n)(c-n)}{c - T'(1)\lambda_1'(1)}\mathbf{1} \ .$$

One can easily show that $\lambda_1'(1) = \lambda$, so that the final equation reads:

$$\left[c \sum_{n=0}^{l-1} \mathbf{f}(n) + \frac{1}{1-\alpha} \sum_{n=l}^{c-1} \mathbf{f}(n)(c-n) \right] \mathbf{1} = c(1-\rho) \ .$$

References

1. Andersen, A.T., Nielsen B.F.: A Markovian approach for modeling packet traffic with long-range dependence. IEEE J. Selected Areas in Comm. 16(5), 719–732 (1998)
2. Arumuganathan, R., Jeyakumar, S.: Steady state analysis of a bulk queue with multiple vacations, setup times with N-policy and closedown times. Appl. Math. Model. 29, 972–986 (2005)
3. Bellalta, B.: A queueing model for the non-continuous frame assembly scheme in finite buffers. Proc. 16th Int. Conf. Anal. Stoch. Model. Tech. Appl. (ASMTA 2009), Madrid, June 9-12, pp.219–233 (2009)
4. Blondia, C.: A discrete-time batch Markovian arrival process as B-ISDN traffic model. Belgian J. Oper. Res., Stat. Comput. Sci. 32, 3–23 (1993).
5. Bruneel, H.: Queueing behavior of statistical multiplexers with correlated inputs. IEEE Trans. Commun. COM-36(12), 1339–1341 (1988)
6. Chakravarthy, S.: A finite-capacity GI/PH/1 queue with group services. Nav. Res. Log. 39(3), 345–357 (1992)
7. Chang, S.H., Choi, D.W.: Performance analysis of a finite-buffer discrete-time queue with bulk arrival, bulk service and vacations. Comp. Oper. Res. 32, 2213–2234 (2005)
8. Chang, S.H., Takine, T.: Factorization and stochastic decomposition properties in bulk queues with generalized vacations. Queueing Syst. 50, 165–183 (2005)
9. Chaudhry, M.L., Templeton, J.G.C.: A first course in bulk queues. John Wiley & Sons (1983)
10. Y. Chen, C. Qiao, X. Yu, Optical burst switching (OBS): a new area in optical networking research, IEEE Network 18(3) (2004) 16–23.
11. Claeys, D., Laevens, K., Walraevens, J., Bruneel, H.: Complete characterisation of the customer delay in a queueing system with batch arrivals and batch service. accepted in Math. Meth. Oper. Res.
12. Claeys, D., Walraevens, J., Laevens, K., Bruneel, H.: Delay analysis of two batch-service queueing models with batch arrivals: $Geo^X/Geo^c/1$. accepted in 4OR
13. De Turck, K., De Vuyst, S., Fiems, D., Wittevrongel, S.: Performance analysis of the IEEE 802.16e sleep mode for correlated downlink traffic. Telecomm. Syst. 39, 145–156 (2008)
14. Gail, H.R., Hantler, S.L., Taylor, B.A.: Spectral analysis of $M/G/1/$ and $G/M/1$ type Markov chains. Adv. Appl. Prob. 28(1), 114–165 (1996)
15. Gao, P., Wittevrongel, S., Bruneel, H.: On the behavior of multiserver buffers with geometric service times and bursty input traffic. IEICE Trans. Commun. E87-B(12), 3576–3583 (2004)

16. Goswami, V., Mohanty, J.R., Samanta, S.K.: Discrete-time bulk-service queues with accessible and non-accessible batches. Appl. Math. Comput. 182, 898–906 (2006)
17. Gupta, U.C., Goswami, V.: Performance analysis of finite buffer discrete-time queue with bulk service. Comp. Oper. Res. 29, 1331–1341 (2002)
18. Herrmann, C.: The complete analysis of the discrete time finite DBMAP/G/1/N queue. Perform. Eval. 43, 95-121 (2001)
19. Janssen, A.J.E.M., van Leeuwaarden, J.S.H.: Analytic computation schemes for the discrete-time bulk service queue. Queueing Syst. 50, 141-163 (2005)
20. Kim, B., Kim, J.: Queue size distribution in a discrete-time D-BMAP/G/1 retrial queue. Comp. Oper. Res. 37(7), 1220–1227 (2010)
21. Kim, N.K., Chaudhry, M.L.: Equivalences of batch-service queues and multi-server queues and their complete simple solutions in terms of roots. Stoch. Anal. Appl. 24, 753–766 (2006)
22. Lee, H.W., Moon, J.M., Kim, B.K., Park, J.G., Lee, S.W.: A simple eigenvalue method for low-order D-BMAP/G/1 queues. Appl. Math. Model. 29, 277-288 (2005)
23. Lu, K., Wu, D., Fang, Y., Qiu, R.C.: Performance analysis of a burst-frame-based MAC Protocol for ultra-wideband ad hoc networks. Proc. IEEE Int. Conf. Commun. 2005 (ICC 2005), Seoul, May 16-20, Vol.5, pp.2937–2941 (2005)
24. Powell, W.B., Humblet, P.: The bulk service queue with a general control strategy: theoretical analysis and a new computational procedure. Oper. Res. 34(2), 267–275 (1986)
25. Qiao, C.M., Yoo, M.S.: Optical burst switching (OBS) - a new paradigm for an optical Internet. J. High Speed Netw. 8(1), 69–84 (1999)
26. Samanta, S.K., Chaudhry, M.L., Gupta, U.C.: Discrete-time $Geo^X|G^{(a,b)}|1|N$ queues with single and multiple vacations. Math. Comp. Model. 45, 93–108 (2007)
27. Samanta, S.K., Gupta, U.C., Sharma, R. K.: Analyzing discrete-time D-BMAP/G/1/N queue with single and multiple vacations. Eur. J. Oper. Res. 182(1), 321–339 (2007)
28. Sikdar, K., Gupta, U.C.: Analytic and numerical aspects of batch service queues with single vacation. Comp. Oper. Res. 32, 943–966 (2005)
29. Yi, X.W., Kim, N.K., Yoon, B.K., Chae, K.C.: Analysis of the queue-length distribution for the discrete-time batch-service $Geo^X|G^{a,Y}|1|K$ queue. Eur. J. Oper. Res. 181, 787–792 (2007)
30. Zhang, Z.: Analysis of a discrete-time queue with integrated bursty inputs in ATM networks. Int. J. Digit. Analog Commun. Syst. 4, 191–203 (1991)
31. Zhao, Y.Q., Campbell, L.L.: Equilibrium probability calculations for a discrete-time bulk queue model. Queueing Syst. 22, 189–198 (1996)