Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2007, Article ID 34631, 18 pages
doi:10.1155/2007/34631

Research Article

Buffer Overflow Period in a MAP Queue

Andrzej Chydzinski

Received 27 April 2006; Revised 23 October 2006; Accepted 20 November 2006

Recommended by Nahum Shimkin

The buffer overflow period in a queue with Markovian arrival process (MAP) and general service time distribution is investigated. The results include distribution of the overflow period in transient and stationary regimes and the distribution of the number of cells lost during the overflow interval. All theorems are illustrated via numerical calculations.

1. Introduction

One of the crucial performance issues of the single-server queue with finite buffer (waiting room) is losses, namely, customers (packets, cells, jobs) that were not allowed to enter the system due to the buffer overflow. This issue is especially important in the analysis of telecommunication networks.

Sample evolution of the queue length process in a finite-buffer system is depicted in Figure 1.1. In a time interval, where the number of customers in the system is equal to its capacity, all arrivals are blocked and lost. We call this interval *buffer overflow period* and it is equivalent to the remaining service time upon reaching a full buffer. Obviously, the duration of the buffer overflow period is responsible for the number of losses that occur consecutively.

Distribution of the buffer overflow period deserves our attention for many practical and theoretical reasons and these motivations can be found by the reader in the literature (e.g. [1, 2]). Herein, we add just one more argument that makes the overflow interval an interesting subject to study.

The overflow interval plays an important role in the design of forward error correction (FEC) algorithms in packet networks. The FEC technique is based on adding redundant packets to the original sequence in order to protect the transmitted data from the loss of

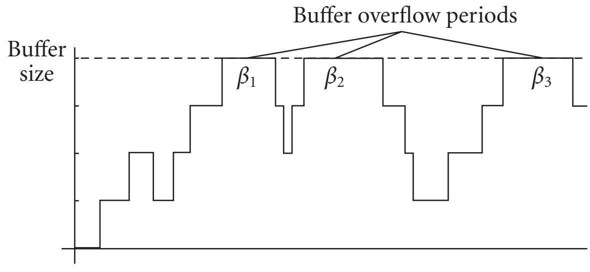

Figure 1.1. Sample path of the queue size process in a single-server system with finite buffer.

packets [3]. Precisely, to each original block of k consecutive packets, we add h redundancy packets, so that the loss of at most h packets in the resulting $k+h$ block can be recovered. FEC is particularly useful in real-time audiovisual transmission, in which the retransmission of lost packets would take too much time [4]. Naturally, distribution of the buffer overflow period has a direct impact on the number h that provides a low level of unrecoverable losses. Pay attention to the fact that the classic parameter, loss ratio, does not give sufficient information about the level of unrecoverable losses. This effect was also studied in [5]. Using a simple $M/G/1/b$ queue, it was shown there that by manipulating the model parameters we can keep the overall loss ratio on a constant level while obtaining drastically different distributions of consecutive losses.

To the best of the author's knowledge, there are no reported results on the duration of the buffer overflow period and the structure of consecutive losses in a queue with autocorrelated input rate. The previous works on the overflow period were concentrated on simple Poisson arrivals [1, 6–8], batch Poisson arrivals [2], or renewal arrivals [8, 9]. However, these processes are not able to mimic the complex autocorrelation structure of observed network traces, which can be bursty and self-similar [10, 11]. Naturally, these properties have a deep impact on the buffer overflow period. As the Markovian structure, up to some extent, can mimic the self-similar behaviour [12, 13], the MAP and its special cases (MMPP) have been widely used in computer networks traffic modeling [14–19]. Queues fed by a MAP are usually studied assuming an infinite buffer (see, for instance, [20–24] and the references given there). The number of papers dealing with characteristics of finite-buffer MAP systems is rather limited [25–29] and none of them is devoted to the overflow period.

The remaining part of the paper is organized in the following manner. In Section 2, a detailed description of the queueing model and notation used throughout the article are given. Section 3 presents the distribution of the first buffer overflow period with a proof, some auxiliary results, and comments. Distribution of the state of the underlying chain at the end of the overflow interval is computed in Section 4. Based on this result, the formulas for subsequent and stationary overflow periods are obtained. In Section 5, distribution of the number of consecutive losses that occur during the overflow period is presented. In Section 6, a set of results for the popular special case of the MAP arrival process, which is the Markov-modulated Poisson process (MMPP), is given. Two numerical examples, that illustrate the theorems proven, are presented in Section 7. In particular, in Example 7.2 a set of numerical results based on buffering of aggregated IP traffic is reported. Finally, remarks concluding the paper and propositions of future work are gathered in Section 8.

2. Queueing model and notation

Let $N(t)$ denote the number of arrivals in $(0,t]$ and $J(t)$, $1 \le J(t) \le m$, the state of the underlying Markov chain. Then the Markovian arrival process is defined as a 2-dimensional Markov process $(N(t),J(t))$ on the state space $\{(i,j) : i \ge 0, 1 \le j \le m\}$ with an infinitesimal generator R in the form

$$R = \begin{bmatrix} D_0 & D_1 & 0 & 0 & 0 & \cdot\cdot \\ 0 & D_0 & D_1 & 0 & 0 & \cdot\cdot \\ 0 & 0 & D_0 & D_1 & 0 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\cdot \end{bmatrix}, \tag{2.1}$$

where D_0 and D_1 are $m \times m$ matrices. D_1 is nonnegative, D_0 has nonnegative off-diagonal elements and negative diagonal elements, and $D = D_0 + D_1$ is an irreducible infinitesimal generator. It is assumed that $D \ne D_0$.

An alternative, constructive definition of a MAP is the following. Assume the underlying Markov process is in some state i, $1 \le i \le m$. The sojourn time in that state has exponential distribution with parameter μ_i. At the end of that time, there occurs a transition to another state and/or an arrival. Namely, with probability $p_i(0,j)$, $1 \le j \le m$, there will be a transition to state j without arrival and with probability $p_i(1,j)$ there will be a transition to state j with an arrival. It is assumed that

$$p_i(0,i) = 0, \qquad \sum_{k=0}^{1}\sum_{j=1}^{m} p_i(k,j) = 1, \quad 0 \le i \le m. \tag{2.2}$$

The following relations between parameters D_k and μ_i, $p_i(k,j)$ hold:

$$\begin{aligned} \mu_i &= -(D_0)_{ii}, \quad 1 \le i \le m, \\ p_i(0,j) &= \frac{1}{\mu_i}(D_0)_{ij}, \quad 1 \le i,\ j \le m,\ j \ne i, \\ p_i(1,j) &= \frac{1}{\mu_i}(D_1)_{ij}, \quad 1 \le i,\ j \le m. \end{aligned} \tag{2.3}$$

We consider herein a single-server queueing system fed by a MAP. The service time is distributed according to a distribution function $F(\cdot)$, which is not further specified, and the standard independence assumptions are made. The buffer size (system capacity) is finite and equal to b (including service position). This means that if a cell (customer) at its arrival finds the buffer overflowed, it is blocked and lost. We assume also that the time origin corresponds to a departure epoch. In Kendall's notation, the system described is denoted by MAP/G/1/b.

The following notation will be used throughout the article:

(i) $J(t)$ is the state of the underlying Markov chain at the moment t,

(ii) $X(t)$ is the queue size at the moment t (including service position),

(iii) $\mathbf{P}(\cdot)$ is the probability,

(iv) $P_{ij}(n,t) = \mathbf{P}(N(t) = n, J(t) = j \mid N(0) = 0, J(0) = i)$ is the counting function for the MAP,

(v) δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise),

(vi) $\mathbf{1}$ is the column vector of 1's,

(vii) the indicator function is

$$I(x > y) = \begin{cases} 1 & \text{if } x > y, \\ 0 & \text{if } x \le y. \end{cases} \tag{2.4}$$

Moreover, the following $m \times m$ matrices will be of use:

$$\begin{aligned} \mathbf{0} &= m \times m \text{ matrix of zeroes}, \\ I &= m \times m \text{ identity matrix}, \\ P(k,t) &= \left[P_{ij}(k,t)\right]_{i,j}, \\ Y_k &= \left[p_i(k,j)\right]_{i,j}, \quad k = 0,1, \end{aligned} \tag{2.5}$$

$$A_k = \left[\int_0^\infty P_{ij}(k,t)\,dF(t)\right]_{i,j}, \tag{2.6}$$

$$R_0 = \mathbf{0}, \qquad R_1 = A_0^{-1}, \qquad R_{k+1} = R_1\left(R_k - \sum_{i=0}^{k} A_{i+1}R_{k-i}\right), \quad k \ge 1, \tag{2.7}$$

$$G_n = (I - Y_0)\sum_{k=0}^{n} R_{n-k}A_k - Y_1\sum_{k=0}^{n-1} R_{n-1-k}A_k. \tag{2.8}$$

(In this notation $[z_{ij}]_{i,j}$ denotes an $m \times m$ matrix with elements z_{ij}.)

Finally, if $\{X_k\}_{k=1}^{\infty}$ is a sequence of $m \times m$ matrices (or column vectors of size m), we define an operator K_b as follows:

$$K_b(X) = (I - Y_0)\sum_{k=1}^{b} R_{b-k}X_k - Y_1\sum_{k=1}^{b-1} R_{b-1-k}X_k. \tag{2.9}$$

Throughout the paper, it is assumed that matrices A_0 and G_b are nonsingular. This assumption does not influence the applicability of the presented results—the author did not find any case with singular A_0 or G_b during his numerical experiments. However, it seems to be hard to prove the nonsingularity in general (see also [26, page 281] for some remarks on the singularity of A_0).

3. First overflow period

In this section, we are going to study the distribution of the first buffer overflow period (β_1 in Figure 1.1), depending on the initial queue length, $X(0)$, and the initial state of the underlying chain, $J(0)$. Formally, β_1 is defined as a difference $\zeta - \tau$, where τ is the first moment at which the buffer is full and ζ is the first departure moment after τ.

To present distribution of the first buffer overflow period, we will be using its tail:

$$\phi_{n,i}(t) = \mathbf{P}(\beta_1 > t \mid X(0) = n, J(0) = i), \tag{3.1}$$

usually in a column-vector form

$$\phi_n(t) = (\phi_{n,1}(t), \dots, \phi_{n,m}(t))^T, \tag{3.2}$$

and the distribution function (also in a column-vector form)

$$H_n(t) = \mathbf{1} - \phi_n(t). \tag{3.3}$$

Theorem 3.1. *Duration of the first buffer overflow period in the MAP/G/1/b system is distributed according to*

$$H_n(t) = \mathbf{1} - \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(g(t)) + \sum_{k=1}^{b-n} R_{b-n-k} g_k(t), \quad 0 \le n < b, \tag{3.4}$$

where

$$g_k(t) = \int_0^\infty (1 - F(u+t)) P(k-1, u) du \cdot D_1 \cdot \mathbf{1} \tag{3.5}$$

and $K_b(g(t))$ is defined in (2.9). (Naturally, replacing X_k by $g_k(t)$ in (2.9) is required.)

Before the proof is presented, let us make some observations. Firstly, it can be checked that if $m = 1$ then representation (3.4) reduces to the $M/G/1/b$-system formula (see [7, Theorem 3]). Secondly, vectors $g_k(t)$ can be computed by means of the uniformization technique (see [21, page 33]). Finally, to prove (3.4), the following lemma will be needed (see [30]).

Lemma 3.2. *Assume that A_0, A_1, $A_2, \dots$ and Ψ_1, $\Psi_2, \dots$ are known sequences of $m \times m$ matrices and A_0 is nonsingular. Then every solution of the system of equations*

$$\sum_{k=-1}^{n-1} A_{k+1} Y_{n-k} - Y_n = \Psi_n, \quad n \ge 1, \tag{3.6}$$

has the form

$$Y_n = R_n C + \sum_{k=1}^{n} R_{n-k} \Psi_k, \quad n \ge 1, \tag{3.7}$$

where C is an $m \times m$ matrix which does not depend on n and the sequence R_k is defined in (2.7).

Note that if $\Psi_1, \Psi_2, \dots$ is a sequence of column vectors of size m, then Y_n and C also become column vectors and Lemma 3.2 remains valid.

Proof of Theorem 3.1. Utilizing the total probability formula with respect to the first departure moment, we have for every $0 < n < b$, $1 \le i \le m$,

$$\begin{aligned}
&\mathbf{P}(\beta_1 > t \mid X(0) = n,\ J(0) = i) \\
&\quad = \sum_{j=1}^{m} \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\beta_1 > t \mid X(0) = n+k-1,\ J(0) = j) P_{ij}(k,\ u)dF(u) \\
&\qquad + \sum_{j=1}^{m} \sum_{k=1}^{m} \int_0^\infty dF(u) \int_0^u I(u-v > t) P_{ij}(b-n-1,\ v)(D_1)_{jk} dv,
\end{aligned} \tag{3.8}$$

where $(D_1)_{jk}$ denotes an element of the matrix D_1 in the position (j,k). The first term in (3.8) corresponds to the case where there is no buffer overflow before the first departure moment, u. Therefore the number of arrivals in $(0,u)$ must be less than $b-n$. The second term corresponds to the case where at some moment $v < u$ the buffer gets overflowed. In this case, we have $\beta = u - v$. If initially the system is empty, for every $1 \le i \le m$ we get

$$\begin{aligned}
\mathbf{P}(\beta_1 > t \mid X(0) = 0, J(0) = i) &= \sum_{j=1}^{m} p_i(0,j) \mathbf{P}(\beta_1 > t \mid X(0) = 0, J(0) = j) \\
&\quad + \sum_{j=1}^{m} p_i(1,j) \mathbf{P}(\beta_1 > t \mid X(0) = 1, J(0) = j).
\end{aligned} \tag{3.9}$$

Integrating the second term in (3.8) by parts and using matrix notation, we may rewrite (3.8) as

$$\phi_n(t) = \sum_{k=0}^{b-n-1} A_k \phi_{n+k-1}(t) + g_{b-n}(t), \tag{3.10}$$

while (3.9) gives

$$\phi_0(t) = Y_0 \phi_0(t) + Y_1 \phi_1(t). \tag{3.11}$$

Replacing $\varphi_n(t) = \phi_{b-n}(t)$ yields

$$\begin{gathered}
\sum_{k=-1}^{n-1} A_{k+1} \varphi_{n-k}(t) - \varphi_n(t) = \psi_n(t), \quad 0 < n < b, \\
\psi_n(t) = A_n \varphi_1(t) - g_n(t),
\end{gathered} \tag{3.12}$$

$$\varphi_b(t) = Y_0 \varphi_b(t) + Y_1 \varphi_{b-1}(t). \tag{3.13}$$

Now, applying Lemma 3.2, the system (3.12) has the following solution:

$$\varphi_n(t) = R_n c(t) + \sum_{k=1}^{n} R_{n-k} \psi_k(t), \quad n \ge 1, \tag{3.14}$$

where $c(t)$ is a column vector which does not depend on n. Putting $n = 1$ into (3.14), we get $c(t) = A_0\varphi_1(t)$ and

$$\varphi_n(t) = \sum_{k=0}^{n} R_{n-k}A_k\varphi_1(t) - \sum_{k=1}^{n} R_{n-k}g_k(t). \tag{3.15}$$

Replacing back $\phi_n(t) = \varphi_{b-n}(t)$ we obtain

$$\phi_n(t) = \sum_{k=0}^{b-n} R_{b-n-k}A_k\phi_{b-1}(t) - \sum_{k=1}^{b-n} R_{b-n-k}g_k(t). \tag{3.16}$$

Finally, using (3.16) and the boundary condition (3.11) yields

$$\phi_{b-1}(t) = G_b^{-1}K_b(g(t)), \tag{3.17}$$

which finishes the proof of Theorem 3.1. □

When applying formula (3.4), we usually store all matrices A_k, R_k and vectors g_k. Therefore, a memory space for $2b$ $m \times m$ matrices and b vectors of size m is required. Regarding the time complexity, if we assume that $m \times m$ matrix multiplication and inversion are of $O(m^3)$ order, then the number of floating-point operations needed to compute (3.4) grows as $O(m^3b^2)$. The main cost is connected with the computation of b matrices R_k using (2.7).

4. Subsequent overflows and stationary distribution

In this section, we are going to study distributions of the subsequent overflow periods ($\beta_2, \beta_3, \ldots$ in Figure 1.1) and the limiting distribution of β_k as $k \to \infty$.

If the queue is fed by a plain Poisson process, the solution is simple, as all of β_k except β_1 have the same distribution. This is due to the fact that after the overflow period ends, the number in the system is $b-1$, and this is the initial queue size for the next overflow period. Moreover, the distribution of β_k, $k \geq 2$ is equal to the distribution of β_1 assuming $X(0) = b-1$. Thus in the case of a Poisson input, the formula for the first overflow period immediately gives the subsequent distributions and the stationary distribution, just by putting $X(0) = n = b-1$.

In the case of a queue fed by a MAP, β_k has a different distribution for every k. Again, the initial queue size for every β_k, $k \geq 2$ is $b-1$, but now the distribution of β_k depends also on the state of the underlying Markov chain at the end of the previous overflow period. Therefore, we have to know how the state of this chain is distributed at the end of the overflow period.

For this purpose, let us consider a random sequence $\{\alpha_k\}_{k=0}^{\infty}$ defined as follows: $\alpha_0 = J(0)$ and α_k, $k \geq 1$ is equal to the state of the underlying Markov chain at the end of the kth overflow period. Therefore the distribution of β_1 depends on α_0 and $X(0)$, while the distribution of β_k, $k \geq 2$ depends only on α_{k-1}. It is easily seen that $\{\alpha_k\}_{k=0}^{\infty}$ is a discrete-time Markov chain.

Denoting

$$S_n = \left[\mathbf{P}(\alpha_1 = l \mid X(0) = n, J(0) = i)\right]_{i,l}, \tag{4.1}$$

the following will be shown.

THEOREM 4.1. *Distribution of the state of the underlying Markov chain at the end of the first buffer overflow period has the form*

$$S_n = \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(U) - \sum_{k=1}^{b-n} R_{b-n-k} U_k, \quad 0 \le n < b, \tag{4.2}$$

where

$$U_k = \sum_{i=k}^{\infty} A_i \tag{4.3}$$

and $K_b(U)$ is defined in (2.9). In particular, $S_{b-1} = G_b^{-1} K_b(U)$.

Proof of Theorem 4.1. Conditioning on the first departure epoch, we obtain for every $0 < n < b$, $1 \le i \le m$,

$$\begin{aligned}
&\mathbf{P}(\alpha_1 = l \mid X(0) = n, J(0) = i) \\
&\quad = \sum_{j=1}^{m} \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\alpha_1 = l \mid X(0) = n+k-1,\ J(0) = j) P_{ij}(k,u) dF(u) \\
&\qquad + \sum_{j=1}^{m} \sum_{k=1}^{m} \int_0^\infty dF(u) \int_0^u \mathbf{P}(J(u-v) = l \mid J(0) = k) P_{ij}(b-n-1, v) (D_1)_{jk} dv.
\end{aligned} \tag{4.4}$$

Again, the first term in (4.4) corresponds to the case where there is no buffer overflow before the first departure moment, u. The second term corresponds to the case where at some moment $v < u$, the buffer gets overflowed and in this case we have $\mathbf{P}(\alpha_1 = l) = \sum_k \mathbf{P}(J(u-v) = l \mid J(0) = k)$.

The boundary condition has the form

$$\begin{aligned}
\mathbf{P}(\alpha_1 = l \mid X(0) = 0, J(0) = i) &= \sum_{j=1}^{m} p_i(0,j) \mathbf{P}(\alpha_1 = l \mid X(0) = 0, J(0) = j) \\
&\quad + \sum_{j=1}^{m} p_i(1,j) \mathbf{P}(\alpha_1 = l \mid X(0) = 1, J(0) = j).
\end{aligned} \tag{4.5}$$

Using the matrix-exponential representation for transient distribution of a continuous-time Markov chain and applying matrix notation to (4.4) and (4.5), we have

$$\begin{aligned}
S_n &= \sum_{k=0}^{b-n-1} A_k S_{n+k-1} + U_{b-n}, \quad 0 < n < b, \\
S_0 &= Y_0 S_0 + Y_1 S_1,
\end{aligned} \tag{4.6}$$

with

$$U_k = \int_0^\infty dF(u) \int_0^u P(k-1,v) D_1 e^{D(u-v)} dv = \sum_{i=k}^\infty A_i. \tag{4.7}$$

We can easily finish the proof by proceeding in the same manner as in the case of Theorem 3.1. □

Now, using Theorems 3.1 and 4.1, it poses no problem to obtain the distribution of β_k for an arbitrary k. We only have to observe that S_n is a transition matrix for the chain α_k in the first step ($\alpha_0 \to \alpha_1$) and S_{b-1} is a transition matrix in every subsequent step ($\alpha_1 \to \alpha_2$, $\alpha_2 \to \alpha_3$, etc.). We get the following.

Corollary 4.2. *If $k \geq 2$, then the duration of the kth buffer overflow period in the MAP/G/1/b system is distributed according to*

$$\mathbf{P}(\beta_k < t) = v S_n (S_{b-1})^{k-2} H_{b-1}(t), \tag{4.8}$$

where vector v is a distribution of the underlying chain at $t = 0$, n is the queue size at $t = 0$, S_n and $H_n(t)$ are given by (4.2) and (3.4), respectively.

Finally, as a consequence of (4.8), we obtain the following.

Corollary 4.3. *Limiting distribution of β_k as $k \to \infty$ has the form*

$$\lim_{k\to\infty} \mathbf{P}(\beta_k < t) = w H_{b-1}(t), \tag{4.9}$$

where w is a stationary vector for the matrix S_{b-1} ($wS_{b-1} = w$, $w\mathbf{1} = 1$).

The existence and the uniqueness of the limiting distribution follows from the fact that the matrix S_{b-1} is irreducible and aperiodic.

5. Structure of consecutive losses

In this section, we are going to study the distribution of the number of losses that occur during the first, subsequent, and stationary overflow periods.

In the case of a Poisson arrival process, the probability that a group of i consecutive arrivals is lost during the overflow period is simply equal to

$$\int_0^\infty \frac{e^{-\lambda t}(\lambda t)^i}{i!} dH(t). \tag{5.1}$$

In the case of a MAP arrival process, we cannot use any formula similar to (5.1). This is due to the fact that we do not know the state of the underlying chain at the beginning of the overflow period. Moreover, this state may be correlated with the duration of the overflow period. Therefore, we have to proceed in similar way as in Section 4.

If we denote by γ_k the number of arrivals lost during the kth buffer overflow period and

$$q_{n,i}(l) = \mathbf{P}(\gamma_1 = l \mid X(0) = n, J(0) = i), \qquad q_n(l) = (q_{n,1}(l), \ldots, q_{n,m}(l))^T, \tag{5.2}$$

then the following theorem can be proven.

Theorem 5.1. *Distribution of the number of arrivals lost during the first buffer overflow period has the form*

$$q_n(l) = \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(v(l)) - \sum_{k=1}^{b-n} R_{b-n-k} v_k(l), \quad 0 \le n < b, \tag{5.3}$$

with $v_k(l) = A_{k+l}\mathbf{1}$ and $K_b(v(l))$ given in (2.9).

Proof of Theorem 5.1. Once again, conditioning on the first departure epoch, we obtain for an initially nonempty system

$$\begin{aligned} &\mathbf{P}(\gamma_1 = l \mid X(0) = n, J(0) = i) \\ &\quad = \sum_{j=1}^{m} \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\gamma_1 = l \mid X(0) = n+k-1, J(0) = j) P_{ij}(k,u) dF(u) \\ &\qquad + \sum_{j=1}^{m} \int_0^\infty P_{ij}(b-n+l,u) dF(u), \quad 0 < n < b,\ 1 \le i \le m. \end{aligned} \tag{5.4}$$

The boundary condition for an initially empty system is identical with (4.5), just replacing alpha's with gamma's is required. Thus we have

$$\begin{gathered} q_n(l) = \sum_{k=0}^{b-n-1} A_k q_{n+k-1}(l) + v_{b-n}(l), \quad 0 < n < b, \\ q_0(l) = Y_0 q_0(l) + Y_1 q_1(l), \end{gathered} \tag{5.5}$$

and the remaining part of the proof is the same as in the case of Theorem 3.1. □

Proceeding in the same manner as in the previous section, we may now obtain the number of losses in the kth overflow period as well as its limiting distribution.

Corollary 5.2. *If $k \ge 2$, then the distribution of the number of arrivals lost during the kth overflow period has the form*

$$\mathbf{P}(\gamma_k = l) = v S_n (S_{b-1})^{k-2} q_{b-1}(l), \tag{5.6}$$

where vector v is a distribution of the underlying Markov chain at $t = 0$, n is the queue size at $t = 0$, S_n and $q_n(l)$ are given by (4.2) and (5.3), respectively. Moreover, the number of arrivals lost during the overflow period in steady state is

$$\lim_{k\to\infty} \mathbf{P}(\gamma_k = l) = w q_{b-1}(l), \tag{5.7}$$

where w is a stationary vector for the matrix S_{b-1}.

6. Special case: MMPP

A Markov-modulated Poisson process is constructed by varying the arrival rate of the Poisson process according to an m-state continuous time Markov chain (see, for instance, [24]). Namely, when the Markov chain is in state i, arrivals occur according to a Poisson process of rate λ_i. Therefore, MMPP is parameterized by two $m \times m$ matrices:

(i) Q is infinitesimal generator of the continuous time Markov chain,

(ii) $\Lambda = \operatorname{diag}(\lambda_1,\dots,\lambda_m)$—its diagonal elements are equal to arrival rates, nondiagonal elements are zeroes.

It is easy to check that the MMPP is the special case of a MAP with the following parameters:

$$D_0 = Q - \Lambda, \qquad D_1 = \Lambda. \tag{6.1}$$

Therefore, all the theorems proven above are valid also for the MMPP, only a slight change in notation is required. In particular, we have now

$$Y_0 = [y_{ij}]_{i,j}, \quad y_{ij} = \begin{cases} 0 & \text{if } i = j, \\ \dfrac{Q_{ij}}{(\lambda_i - Q_{ii})} & \text{if } i \neq j, \end{cases}$$
$$Y_1 = \left[\frac{\Lambda_{ij}}{\lambda_i - Q_{ii}}\right]_{i,j}, \qquad g_k(t) = \int_0^\infty (1 - F(u+t))P(k-1,u)du \cdot \Lambda \cdot \mathbf{1}. \tag{6.2}$$

7. Numerical results

Example 7.1. Let us start with the same MMPP parameterization that was used in [25]. Namely, we assume the constant service time as a time unit and

$$\Lambda = \begin{bmatrix} 1.0722 & 0 \\ 0 & 0.48976 \end{bmatrix}, \qquad Q = \begin{bmatrix} -8.4733 \cdot 10^{-4} & 8.4733 \cdot 10^{-4} \\ 5.0201 \cdot 10^{-6} & -5.0201 \cdot 10^{-6} \end{bmatrix}. \tag{7.1}$$

The stationary distribution for Q is then $\pi = (0.00589, 0.99411)$ and the load offered to the queue is around 49%.

Basic steady-state parameters of this system, like distribution of the queue size and workload, can be found in [25]. Herein, we are going to examine overflow intervals and structure of consecutive losses of this system. For visualization, the probability density function

$$h_n(t) = (h_{n,1}(t),\dots,h_{n,m}(t))^T = \frac{dH_n(t)}{dt} \tag{7.2}$$

will be used rather than the distribution function $H_n(t)$.

Figure 7.1 shows the dependence of the density of the first overflow period on the buffer size. In part (a), the initial state of the underlying chain is 1, while in part (b) it is 2. In both parts, the queue is initially empty. We may observe that as b grows, the

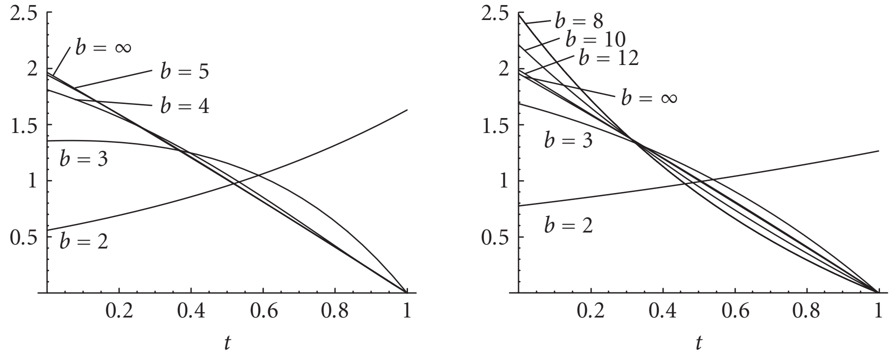

Figure 7.1. Shapes of the density function $h_{n,i}(t)$ for different buffer sizes. Part (a): $i = 1$, $n = 0$. Part (b): $i = 2$, $n = 0$.

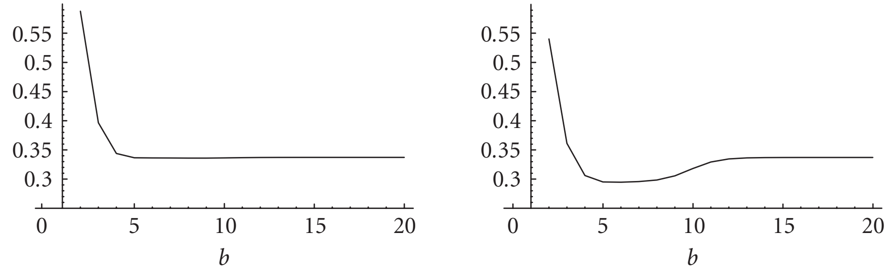

Figure 7.2. Mean duration of the first buffer overflow period versus the buffer size, b. Part (a): $i = 1$, $n = 0$. Part (b): $i = 2$, $n = 0$.

distribution of the first overflow period converges quickly to some limiting distribution. This is also confirmed in the plot representing average values (Figure 7.2.). The average duration of the buffer overflow stabilizes for b around 5 or 13, depending if the initial state of the underlying chain is 1 or 2. It is also interesting that the average value may not change monotonically (Figure 7.2(b)).

In Figure 7.3, the dependence of the density of the first overflow period on the initial queue size is depicted. For a wide range of n the density function does not change significantly, but it changes rapidly, as n closely approaches b.

Now we are going to examine the subsequent and stationary overflow periods assuming $b = 50$. First of all, we have to compute the transition matrix S_{b-1}. Using (4.2) gives

$$S_{b-1} = \begin{bmatrix} 0.999202 & 0.000798 \\ 0.509801 & 0.490199 \end{bmatrix}. \tag{7.3}$$

Table 7.1. Probability of losing l consecutive arrivals during the kth overflow period. System parameters: $J(0) = 2$, $X(0) = 49$, $b = 50$.

$l\backslash k$	1	2	3	4	∞
0	7.5377×10^{-1}	6.7843×10^{-1}	6.4156×10^{-1}	6.2352×10^{-1}	6.0622×10^{-1}
1	1.9772×10^{-1}	2.3715×10^{-1}	2.5644×10^{-1}	2.6589×10^{-1}	2.7494×10^{-1}
2	4.0314×10^{-2}	6.5763×10^{-2}	7.8217×10^{-2}	8.4313×10^{-2}	9.0155×10^{-2}
3	6.9565×10^{-3}	1.5120×10^{-2}	1.9116×10^{-2}	2.1072×10^{-2}	2.2946×10^{-2}
4	1.0618×10^{-3}	2.9472×10^{-3}	3.8699×10^{-3}	4.3215×10^{-3}	4.7544×10^{-3}
5	1.4607×10^{-4}	4.9485×10^{-4}	6.6554×10^{-4}	7.4908×10^{-4}	8.2915×10^{-4}
6	1.8237×10^{-5}	7.2606×10^{-5}	9.9214×10^{-5}	1.1223×10^{-4}	1.2471×10^{-4}
7	2.0731×10^{-6}	9.4304×10^{-6}	1.3031×10^{-5}	1.4793×10^{-5}	1.6482×10^{-5}
8	2.1530×10^{-7}	1.0968×10^{-7}	1.5282×10^{-6}	1.7394×10^{-6}	1.9417×10^{-6}

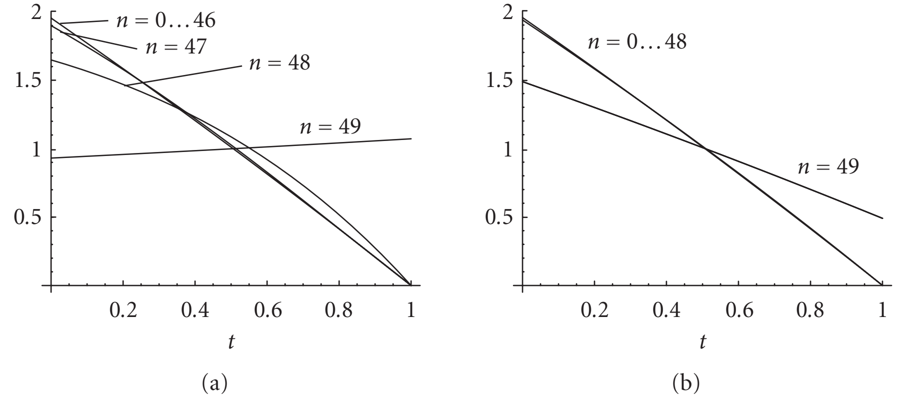

Figure 7.3. Shapes of the density function $h_{n,i}(t)$ for different initial queue sizes. Part (a): $i = 1$, $b = 50$. Part (b): $i = 2$, $b = 50$.

The stationary vector for S_{b-1} is then $w = (0.998437, 0.001563)$. Pay attention to the different values of π and w. While in general the underlying chain stays in state 2 for the majority of time, the state at the end of the overflow period is usually 1. This can be explained by the fact that the overflow usually occurs when the state of the underlying chain is 1 (λ_1 is higher than λ_2), and this state remains to the end of the overflow interval.

Figure 7.4 represents densities for the subsequent and stationary overflow periods, while Table 7.1 contains distributions of the number of losses during the subsequent and stationary overflow intervals. In both Figure 7.4 and Table 7.1, we may observe rather quick convergence to the limiting distribution.

Example 7.2. In the second example, we are going to use an MMPP fitted to aggregated IP traffic and show how the autocorrelation structure of the traffic influences the overflow period and the probability of consecutive packet losses. For this purpose, a publicly

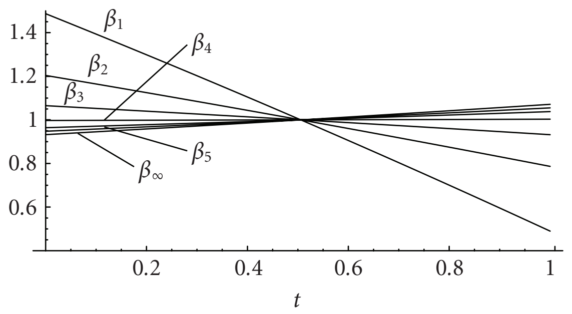

Figure 7.4. Densities of the duration of the first, subsequent, and stationary buffer overflow periods. System parameters: $J(0) = 2$, $X(0) = 49$, $b = 50$.

available trace file, recorded at the front range GigaPOP (FRG) aggregation point, which is run by PMA (passive measurement and analysis project, see http://pma.nlanr.net/) has been utilized. Precisely, one million packet headers from the file FRG-1137208198-1.tsh, recorded on Jan 14th, 2006, were used. The mean rate of the traffic is 58.11 MB/s, or 71732 packets/s with average packet size of 850 B. Using this traffic sample, the following MMPP parameters were fitted [29]:

$$Q = \begin{bmatrix} -172.53 & 38.80 & 30.85 & 0.88 & 102.00 \\ 16.76 & -883.26 & 97.52 & 398.9 & 370.08 \\ 281.48 & 445.97 & -1594.49 & 410.98 & 456.06 \\ 23.61 & 205.74 & 58.49 & -598.93 & 311.09 \\ 368.48 & 277.28 & 7.91 & 32.45 & -686.12 \end{bmatrix}, \tag{7.4}$$

$(\lambda_1,\dots,\lambda_5) = (59620.6, 113826.1, 7892.6, 123563.2, 55428.2)$.

The average interarrival times are almost the same in the traffic sample and the fitted MMPP (13.940 μs versus 13.941 μs) and, what is important, the MMPP preserves the autocorrelation structure of the original traffic up to three time scales (see Figure 7.5).

We assume that the queue is served at a rate of 80 MB/s (98690 pkts/s), which makes the service (transmission) time to be 10.133 μs and the offered load 73%. We assume also that the buffer can hold 120 packets (100 KB). These settings make the stationary full buffer probability to be 6.55×10^{-3}, which can be computed by means of the methodology presented in [25].

The transition matrix S_{b-1}, computed by means of Theorem 4.1, is now

$$S_{b-1} = \begin{bmatrix} 0.60242 & 0.09548 & 0.00020 & 0.29941 & 0.00247 \\ 0.00021 & 0.97342 & 0.00054 & 0.02204 & 0.00377 \\ 0.00260 & 0.24492 & 0.07887 & 0.66704 & 0.00657 \\ 0.00026 & 0.00677 & 0.00033 & 0.98956 & 0.00305 \\ 0.00398 & 0.13003 & 0.00014 & 0.30881 & 0.55703 \end{bmatrix}, \tag{7.5}$$

and its stationary vector $w = (0.00070, 0.23479, 0.00041, 0.75684, 0.00723)$.

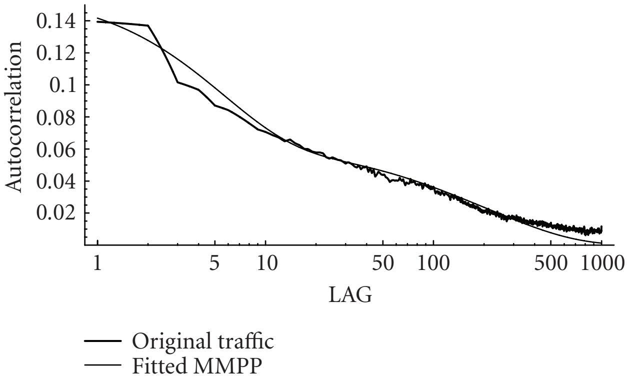

Figure 7.5. Autocorrelation between interarrival times in original traffic and fitted MMPP (Example 7.2). The thin curve represents MMPP.

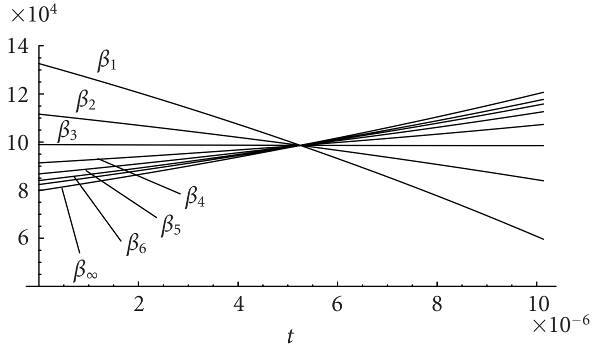

Figure 7.6. Densities of the duration of the first, subsequent, and stationary buffer overflow periods in Example 7.2. Initial parameters: $J(0) = 1$, $X(0) = b - 1$.

Figure 7.6 reports densities of the duration of the first, subsequent, and stationary buffer overflow periods for initial state of the underlying chain 1 and $X(0) = b - 1$. As in Example 7.1, we can observe rather quick convergence to the limiting distribution.

Now, let us check how the autocorrelation structure of the MMPP influences the duration of the overflow period. For this purpose, let us consider another queue, whose arrival process is simple Poisson and therefore the packet interarrival times are uncorrelated. We set for the same, as previously, arrival rate, service time and buffer size and thus we can compare results. They are shown in Figure 7.7 and Table 7.2. It is evident that in the MMPP case, the probability mass is concentrated around higher values and this causes longer overflow periods. As a consequence, we may observe that in the MMPP queue, the probability of losing a group of consecutive arrivals is much higher than in the Poisson-arrival case and this effect is stronger when the size of a group is larger (Table 7.2). For

Table 7.2. Probability of losing l consecutive packets during the overflow period for MMPP and Poisson arrival systems (both in steady state).

l	*MMPP*	*Poisson*
0	6.8230×10^{-1}	8.0571×10^{-1}
1	2.3465×10^{-1}	1.6359×10^{-1}
2	6.4755×10^{-2}	2.6638×10^{-2}
3	1.4826×10^{-2}	3.5958×10^{-3}
4	2.8914×10^{-3}	4.1307×10^{-4}
5	4.9020×10^{-4}	4.1217×10^{-5}
6	7.3432×10^{-5}	3.6307×10^{-6}
7	9.8470×10^{-6}	2.8608×10^{-7}
8	1.1947×10^{-6}	2.0379×10^{-8}
9	1.3233×10^{-7}	1.3242×10^{-9}
10	1.3480×10^{-8}	7.9087×10^{-11}

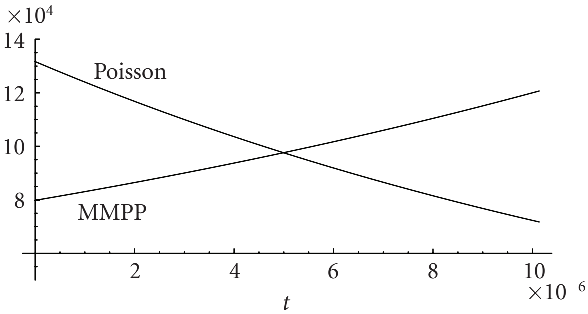

Figure 7.7. Stationary density of the buffer overflow period for MMPP and Poisson arrival systems (Example 7.2). In both systems service times, loads and buffer sizes are the same.

instance, losing 5 packets in row is 12 times more probable in the MMPP queue than in the Poisson one, while losing 10 packets in row is 170 times more probable.

8. Conclusions and future work

In this paper, the buffer overflow period in a finite-buffer queue fed by a MAP was studied. In particular, formulas for the distribution of the overflow period and the number of arrivals lost during the overflow period were developed. The transient and the stationary cases were both solved. Theoretical results were illustrated via numerical examples.

In addition, it was demonstrated that the autocorrelated structure of the arrival process may have a negative impact on the overflow period and the number of losses that occur consecutively.

As MAP is an analytically tractable model that can capture autocorrelation properties of observed arrival processes and because it is commonly used in network traffic modeling, the results presented herein are likely to be of practical importance.

As regards possible future work on this subject, the first interesting problem is the asymptotic behaviour of the overflow period in a large buffer queue. Based on numerical results, we may expect that as b grows, the duration of the overflow period converges to some limiting distribution. Probably, this limiting distribution has a much simpler form than (3.4), which potentially makes it of practical importance. For Poisson and batch Poisson arrivals, such formulas are already available, see [1, 2], respectively.

Acknowledgment

This material is based in part upon work supported by the Polish Ministry of Scientific Research and Information Technology under Grant no. 3 T11C 014 26.

References

[1] P.-T. de Boer, V. F. Nicola, and J.-K. C. W. van Ommeren, "The remaining service time upon reaching a high level in $M/G/1$ queues," *Queueing Systems*, vol. 39, no. 1, pp. 55–78, 2001.

[2] A. Chydzinski, "Duration of the buffer overflow period in a batch arrival queue," *Performance Evaluation*, vol. 63, no. 4-5, pp. 493–508, 2006.

[3] A. Jean-Marie, Y. Calas, and T. Alemu, "On the compromise between burstiness and frequency of events," *Performance Evaluation*, vol. 62, no. 1–4, pp. 382–399, 2005.

[4] G. Carle and E. Biersack, "Survey of error recovery techniques for IP-based audiovisual multicast applications," *IEEE Network*, vol. 11, no. 6, pp. 24–36, 1997.

[5] A. Chydzinski, "On the distribution of consecutive losses in a finite capacity queue," *WSEAS Transactions on Circuits and Systems*, vol. 4, no. 3, pp. 117–124, 2005.

[6] S. Asmussen, "Equilibrium properties of the $M/G/1$ queue," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 58, no. 2, pp. 267–281, 1981.

[7] A. Chydzinski, "On the remaining service time upon reaching a given level in $M/G/1$ queues," *Queueing Systems*, vol. 47, no. 1-2, pp. 71–80, 2004.

[8] A. Pacheco and H. Ribeiro, "Consecutive customer loss probabilities in $M/G/1/n$ and $GI/M(m)//n$ systems," in *Proceedings of Workshop on Tools for Solving Structured Markov Chains (SMCtools '06)*, Pisa, Italy, October 2006.

[9] D. Fakinos, "The expected remaining service time in a single server queue," *Operations Research*, vol. 30, no. 5, pp. 1014–1018, 1982.

[10] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[11] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, 1997.

[12] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," *Telecommunication Systems*, vol. 17, no. 1-2, pp. 185–211, 2001.

[13] P. Salvador, R. Valadas, and A. Pacheco, "Multiscale fitting procedure using Markov modulated Poisson processes," *Telecommunication Systems*, vol. 23, no. 1-2, pp. 123–148, 2003.

[14] S. H. Kang, Y. H. Kim, D. K. Sung, and B. D. Choi, "An application of Markovian arrival process (MAP) to modeling superposed ATM cell streams," *IEEE Transactions on Communications*, vol. 50, no. 4, pp. 633–642, 2002.

[15] S.-Q. Li and C.-L. Hwang, "On the convergence of traffic measurement and queueing analysis: a statistical-matching and queueing (SMAQ) tool," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 95–110, 1997.

[16] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, 1997.

[17] S. Shah-Heydari and T. Le-Ngoc, "MMPP models for multimedia traffic," *Telecommunication Systems*, vol. 15, no. 3-4, pp. 273–293, 2000.

[18] T. C. Wong, J. W. Mark, and K. C. Chua, "Delay performance of voice and MMPP video traffic in a cellular wireless ATM network," *IEE Proceedings - Communications*, vol. 148, no. 5, pp. 302–309, 2001.

[19] S. H. Kang and D. K. Sung, "Two-state MMPP modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '95)*, vol. 2, pp. 1422–1426, Singapore, November 1995.

[20] V. Ramaswami, "The $N/G/1$ queue and its detailed analysis," *Advances in Applied Probability*, vol. 12, no. 1, pp. 222–261, 1980.

[21] D. M. Lucantoni, "New results on the single server queue with a batch Markovian arrival process," *Communications in Statistics. Stochastic Models*, vol. 7, no. 1, pp. 1–46, 1991.

[22] D. M. Lucantoni, G. L. Choudhury, and W. Whitt, "The transient $BMAP/G/1$ queue," *Communications in Statistics. Stochastic Models*, vol. 10, no. 1, pp. 145–182, 1994.

[23] D. M. Lucantoni, "Further transient analysis of the $BMAP/G/1$ queue," *Communications in Statistics. Stochastic Models*, vol. 14, no. 1-2, pp. 461–478, 1998.

[24] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1993.

[25] A. Baiocchi and N. Blefari-Melazzi, "Steady-state analysis of the MMPP/G/1/K queue," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 531–534, 1992.

[26] C. Blondia, "The $N/G/1$ finite capacity queue," *Communications in Statistics: Stochastic Models*, vol. 5, no. 2, pp. 273–294, 1989.

[27] F. N. Gouweleeuw, "Calculating the loss probability in a $BMAP/G/1/N+1$ queue," *Communications in Statistics: Stochastic Models*, vol. 12, no. 3, pp. 473–492, 1996.

[28] S. Asmussen, M. Jobmann, and H.-P. Schwefel, "Exact buffer overflow calculations for queues via martingales," *Queueing Systems*, vol. 42, no. 1, pp. 63–90, 2002.

[29] A. Chydzinski, "Transient analysis of the MMPP/G/1/K queue," *Telecommunication Systems*, vol. 32, no. 4, pp. 247–262, 2006.

[30] A. Chydzinski, "Time to reach buffer capacity in a *BMAP* queue," to appear in *Stochastic Models*.

Andrzej Chydzinski: Institute of Computer Sciences, Silesian University of Technology,
Akademicka 16, 44-100 Gliwice, Poland
Email address: andrzej.chydzinski@polsl.pl