Transient and stationary distributions for the GI/G/k queue with Lebesgue–dominated inter–arrival time distribution

Lothar Breuer
FB IV - Informatik, Universität Trier, 54286 Trier, Germany. breuer@info04.uni-trier.de

Abstract. In this paper, the multi–server queue with general service time distribution and Lebesgue–dominated iid inter–arival times is analyzed. This is done by introducing auxiliary variables for the remaining service times and then examining the embedded Markov chain at arrival instants. The concept of piecewise–deterministic Markov processes is applied to model the inter–arrival behaviour. It turns out that the transition probability kernel of the embedded Markov chain at arrival instants has the form of a lower Hessenberg matrix and hence admits an operator–geometric stationary distribution. Thus it is shown that matrix–analytical methods can be extended to provide a modeling tool even for the general multi-server queue.

1. Introduction

The study of general multi–server queues has proved to be much more difficult than the study of the single–server queue. While the classical paper by Kiefer and Wolfowitz (1955) examines the chain of the users' waiting times, a satisfying method of deriving the queueing process (giving the number of users in the system) has not been developed yet. There are good reasons for this lack of results. Models in terms of Markov jump processes such as $MAP/PH/k$ queues yield only approximations (even this has not been proven rigorously yet) while sometimes still requiring very large state spaces. The construction of an embedded Markov chain like in the analysis of $M/G/1$ or $GI/G/1$ queues (see Tweedie (1982) or Meyn, Tweedie (1993)) turns out to be much more difficult for multi-server queues, because the existence of more than one concurrent non–Markovian marginal process (namely the concurrent services) makes the behaviour of the queueing process between arrivals more complicated.

In the present paper, the concept of piecewise–deterministic Markov processes (see Davis (1984), Costa and Dufour (1990; 1999), or Davis (1993)) is exploited in order to describe the inter–arrival process in its full complexity. Having described this missing link between arrival instants, it is possible to conduct an analysis of the multi–server queue along the classical path of introducing auxiliary variables for the remaining service times and then examining the Markov chain at arrival instants. A nice feature of the analysis is then that the transition probability kernel of the embedded Markov chain can be arranged to have the form of a $G/M/1$–type matrix with kernel entries. Thus it admits operator–geometric solutions along the lines of Tweedie (1982).

This shows that the structures revealed in the matrix–analytical paradigm can even be found in the very general class of multi–server queues to be analyzed in this paper. By the way it shall be mentioned that an analysis of the $M/G/k$ queue can be performed in a much simpler way by modelling the queueing process as a piecewise–deterministic Markov process directly and then obtaining operator–geometric solutions, see Breuer (2003).

Consider the following $GI/G/k$ queue: Arrivals occur independently with iid inter–arrival times distributed by A. In order to avoid multiple events (like arrival and departure) occuring at the same time instant, we assume that A has a Lebesgue density a. For deterministic inter–arrival times (which are not Lebesgue–dominated), introduce a small perturbation which is Lebesgue–dominated. Then the present model applies and yields an approximation. Only single arrivals are allowed. The system shall have $k \in \mathbb{N}$ independent servers. Every user has identical service time distribution B. The service discipline is first come first served, and the capacity of the waiting room is infinite. Let $Q = (Q_t : t \in \mathbb{R}_0^+)$ denote the queueing process, which is defined as follows: The state space of Q shall be $E := \mathbb{N}_0 \times (\mathbb{R}_0^+)^k$ with the first dimension indicating the number of waiting users and the last k dimensions describing the remaining service times of the users in service. If the ith server is idle, then the state of the system is $(0, x)$ with $x \in (\mathbb{R}_0^+)^k$ and $x_i = 0$. The state space E contains the minimal information needed to give a full description of the system.

As already mentioned above, the method of analysis in the present paper will be the following: First, the queueing process is described between successive arrivals by means of a piecewise–deterministic Markov process. This representation can be used directly in order to determine the transient distribution of the queueing process at any time $t \in \mathbb{R}_0^+$. Furthermore, it yields the transition probability kernel of the embedded Markov chain immediately before arrival instants. This embedded chain will be used to derive a stability condition in terms of mean drift as well as an expression for the stationary distribution.

The rest of the paper is organized as follows: In section 2 the queueing process between successive arrival instants will be described. Section 3 contains an expression for the transient distribution of the queueing process. Section 4 gives a stability condition in terms of mean drift and section 5 contains an algorithm for determining the stationary distribution. A simple example is given in section 6.

2. The Inter–Arrival Process

Let I denote the queueing process between arrivals. Then I is a piecewise–deterministic Markov process with state space $E := \mathbb{N}_0 \times (\mathbb{R}_0^+)^k$ having the same interpretation as the state space for Q. A flow Φ on E shall be defined by

$$\Phi_t(n,x) := (n,(x_1-t)^+,\ldots,(x_k-t)^+)$$

for all $(n,x) = (n,x_1,\ldots,x_k) \in E$ and $t \in I\!R_0^+$, with $(s-t)^+ := \max(0,s-t)$ for all $s,t \in I\!R$. Obviously, this flow represents the proceeding service time. Furthermore, we define

$$t_*(x) := \begin{cases} \min\{x_i : 1 \le i \le k, x_i > 0\} & \text{for} \quad x \neq 0 \\ \infty & \text{for} \quad x = 0 \end{cases}$$

for all $x = (x_1,\ldots,x_k) \in (I\!R_0^+)^k$, denoting the time until the first server will become idle not taking into account future arrivals.

The transition measure Q_s describes the state changes of the system in the case of a server becoming idle. If there are any waiting users in the queue, it immediately will commence to serve a new user. Let $x = (x_1,\ldots,x_k) \in (I\!R_0^+)^k$ and $A = A_1 \times \ldots \times A_k \in \mathcal{B}^k$, with $\mathcal{B}^k$ denoting the σ–algebra of the Borel sets on $(I\!R_0^+)^k$. Then we have

$$Q_s((n,x),\{m\}\times A) := \delta_{m,n-1} \cdot \prod_{j=1,j\neq i}^{k} 1_{A_j}(x_j) \cdot B(A_i)$$

for $n \ge 1$ and $x_i = 0$, as well as

$$Q_s((n,x),\{m\}\times A) := \delta_{m,n} \cdot 1_A(x)$$

for $n = 0$, with $\delta_{m,n} = 1$ if $m = n$ and $\delta_{m,n} = 0$ if $m \neq n$ denoting the Kronecker function.

Note that for the case $n \ge 1$, only one server can become idle at a time. Since the queue has Lebesgue–dominated single arrival input and the servers work independently, the probability that two servers finish their work at the same time instant is zero. Furthermore, if one server had been idle before the other server and there had been any waiting users in the queue, it would have commenced serving one of them.

Denote $1_k \in I\!R^k$ as the vector with all entries being one and write

$$(x-y)^+ := ((x_1-y_1)^+,\ldots,(x_k-y_k)^+)$$

for all $x,y \in I\!R^k$. Let $P^I(t;(n,x),\{m\}\times A)$ denote the probability that at time $t \in I\!R_0^+$ after the last arrival, the inter–arrival process I is in state set $\{m\}\times A$ under the condition that it was in state (n,x) immediately after the last arrival. Further, let $P_I^{(l)}(t;(n,x),\{m\}\times A)$ denote the same probability, but restricted to the set of paths with $l \in I\!N_0$ jumps (i.e. service completions) until time t. Then the transition probability kernel and hence the transient distribution of the inter–arrival process X is given iteratively by

$$P^I(t; (n,x), \{m\} \times A) = \sum_{i=0}^{\infty} P_I^{(i)}(t; (n,x), \{m\} \times A) \qquad (1)$$

with

$$P_I^{(0)}(t; (n,x), \{m\} \times A) = \delta_{m,n} \cdot 1_A(x - t \cdot 1_k)$$

for $t < t_*(x)$ and $n \geq 1$,

$$P_I^{(0)}(t; (n,x), \{m\} \times A) = \delta_{m,0} \cdot 1_A((x - t \cdot 1_k)^+)$$

for $n = 0$, and $P_I^{(0)}(t; (n,x), \{m\} \times A) = 0$ else as start values and iterating by

$$P_I^{(i+1)}(t; (n,x), \{m\} \times A) = \\ = \int_E P_I^{(i)}(t - t_*(x); (l,y), \{m\} \times A)\ Q_s((n,x), d(l,y))$$

if $t > t_*(x)$ and $P_I^{(i+1)}(t; (n,x), \{m\} \times A) = 0$ else. Note that the sum in (1) can be truncated with arbitrarily small error as the service times are iid.

Remark 1 From the above definitions and the transition measure Q_s it is clear that $m > n$ implies $P^I(t; (n,x), \{m\} \times A) = 0$ for all $t \in \mathbb{R}_0^+, x \in \mathbb{R}^k$ and $A \in \mathcal{B}^k$. This expresses the obvious fact that during inter–arrival times the number of waiting users cannot increase.

Remark 2 A more efficient way to compute the transient distribution of the inter–arrival process is the following: Define the operator f by $f(P_I^{(i)}) := P_I^{(i+1)}$. Then the transition probability kernel P^I can be computed as the limit $P^I = \lim_{n\to\infty} P_n$ with $P_0 = P_I^{(0)}$ and $P_{n+1} = f(P_n) + P_0$ for all $n \in \mathbb{N}_0$.

3. Transient Distribution

Define a transition measure Q_a that describes the state changes in the queueing process induced by an arrival event. Remembering that B denotes the service time distribution of a new user, this is given for all $(n,x) \in E$ with $x = (x_1, \ldots, x_k)$ and $A = A_1 \times \ldots \times A_k$ by

$$Q_a((n,x), \{m\} \times A) := \delta_{m,n+1} \cdot 1_A(x)$$

if $\prod_{i=1}^{k} x_i > 0$ and

$$Q_a((n,x),\{m\}\times A) := \delta_{m,n}\cdot \prod_{j=1, j\neq i}^{k} 1_{A_j}(x_j)\cdot B(A_i)$$

if $i = \min\{l : x_l = 0\}$ exists. Note that the latter case in the definition of Q_a is possible only for $n = m = 0$. If we define further

$$P^{(0)}(t;(n,x),\{m\}\times A) := P^I(t;(n,x),\{m\}\times A)$$

and for $i \in I\!N_0$

$$\begin{aligned} P^{(i+1)}(t;(n,x),\{m\}\times A) := \int_0^t \int_{(h,z)\in E} P^{(i)}(t-u;(h,z),\{m\}\times A) \\ \times \int_{(l,y)\in E} P^I(u;(n,x),d(l,y))Q_a((l,y),d(h,z))\; a(u)du \end{aligned}$$

then the transient distribution is given as

$$P(t;(n,x),\{m\}\times A) = \sum_{i=0}^{\infty} P^{(i)}(t;(n,x),\{m\}\times A) \quad (2)$$

by conditioning on the number i of arrivals until time t. Computing the transient distribution will be more efficient by using the same scheme as in remark 2. Since the probability of i arrivals until time t tends to 0 as i tends to infinity, the sum in (2) can be truncated while keeping the error arbitrarily small.

4. Stability

Let $(T_n : n \in I\!N)$ denote the time instants of successive arrivals. Clearly, $(T_n : n \in I\!N)$ is a series of stopping times with respect to the canonical filtration of the queueing process Q. Further, define $X_n := Q_{T_n-}$ as the system state immediately before the nth arrival. Then $X = (X_n : n \in I\!N)$ is the embedded Markov chain immediately before arrival instants. Let $P^X((l,y),\{m\}\times A) := P(X_{n+1} \in \{m\}\times A | X_n = (l,y))$ denote the transition probabilities of the chain X. Using the results of section 2, these are given by

$$\begin{aligned} P^X((n,x),\{m\}\times A) \\ = \int_E \int_0^\infty P^I(t;(l,y),\{m\}\times A)\; a(t)\; dt\; Q_a((n,x),d(l,y)) \quad (3) \end{aligned}$$

for all $n, m \in \mathbb{N}_0$, $x \in (\mathbb{R}_0^+)^k$, and $A \in \mathcal{B}^k$. If the states are arranged according to their first dimension (i.e. the number of waiting users), then the transition probability kernel P^X assumes the form

$$P^X = \begin{pmatrix} B_0 & A_0 & 0 & 0 & 0 & \dots \\ B_1 & A_1 & A_0 & 0 & 0 & \dots \\ B_2 & A_2 & A_1 & A_0 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{4}$$

of a lower Hessenberg matrix with kernel entries. Here, the kernels are given by

$$A_i(x, A) = P^X((n, x), \{n - i + 1\} \times A)$$

and

$$B_i(x, A) = P^X((i, x), \{0\} \times A)$$

for all $n \geq i \in \mathbb{N}_0$, $x \in (\mathbb{R}_0^+)^k$ and $A \in \mathcal{B}^k$. By equation (3), remark 1 and the definition of Q_a we can write

$$A_i(x, A) = \int_0^\infty P^I(t; (n+1, x), \{n - i + 1\} \times A)\ a(t)\ dt \tag{5}$$

as well as

$$B_i(x, A) = \int_0^\infty P^I(t; (n+1, x), \{0\} \times A)\ a(t)\ dt$$

for $\prod_{i=1}^{k} x_i > 0$ and

$$B_i(x, A) = \int_0^\infty \int_0^\infty P^I(t; (0, y), \{0\} \times A)\ dB(y_i)\ a(t)\ dt$$

for $i = \min\{l : x_l = 0\}$, $y_l = x_l$ for all $l \neq i$. Furthermore, we have $B_{n+1}(x, (\mathbb{R}_0^+)^k) = B_n(x, (\mathbb{R}_0^+)^k) + A_n(x, (\mathbb{R}_0^+)^k)$ and thus

$$B_n(x, (\mathbb{R}_0^+)^k) = \sum_{i=n+1}^{\infty} A_i(x, (\mathbb{R}_0^+)^k)$$

for all $n \in \mathbb{N}_0$ and $x \in (\mathbb{R}_0^+)^k$.

Matrices of this kind have been analyzed in Tweedie (1982). Define the kernel $A = \sum_{i=0}^{\infty} A_i$. This kernel equals the transition probability kernel of the remaining service times immediately before arrival instants under the condition that there always is at least one waiting user. This condition implies that the servers behave independent from one another. Hence the stationary distribution π of the kernel A equals the k–fold product of the respective stationary distribution ν for one server. The latter is given in Tweedie (1982), p.389, as

$$\nu[0,t] = \mu \int_0^t (1 - B(u))\, du$$

for all $t \in I\!\!R_0^+$, with $\bar{\mu} = \int_0^\infty 1 - B(u)\, du < \infty$ denoting the mean service time and defining $\mu := (\bar{\mu})^{-1}$.

Define $\beta(x) := \sum_{n=1}^\infty nA_n(x, (I\!\!R_0^+)^k)$ for all $x \in (I\!\!R_0^+)^k$. Then the analogue to Neuts' mean drift condition (see Neuts (1978), theorems 2 and 3) is given by formula (3.1) in Tweedie (1982), namely

$$\int_{(I\!\!R_0^+)^k} \beta(x)\pi(dx) > 1 \tag{6}$$

meaning that in the mean, there will be more than one service during an inter–arrival time. This stability condition for the embedded chain X coincides with the classical form

$$\bar{\mu} < k \cdot \bar{\lambda} \tag{7}$$

of the stability condition (see Kiefer, Wolfowitz (1955)), with $\bar{\lambda}$ denoting the mean inter–arrival time, i.e. $\bar{\lambda} = \int_0^\infty t\ a(t)\ dt$ (see remark 3 below). Hence condition (6) is equivalent to the stability of the queueing process Q, implying ergodicity of the embedded chain X (because the embedding is at arrival instants, and thus reaching state $(0,0)$ under Q means remaining in state $(0,0)$ until the next arrival and hence reaching $(0,0)$ under X).

Remark 3 The equivalence of the two forms (6) and (7) is shown by the following arguments: The distribution π has Lebesgue–density

$$d\pi(x) = \prod_{i=1}^k \mu\,(1 - B(x_i))\, dx_i$$

for all $x = (x_1, \ldots, x_k) \in (I\!\!R_0^+)^k$. This means by definition of β and expressions (5) that

$$\int_{(I\!\!R_0^+)^k} \beta(x)\pi(dx) = \int_0^\infty \int_{(I\!\!R_0^+)^k} \sum_{n=1}^\infty nP^I(t; (n,x), \{1\} \times (I\!\!R_0^+)^k) \\ \times \prod_{i=1}^k \mu\,(1 - B(x_i))\ dx_1 \ldots dx_k\ a(t)\ dt$$

Now the inner integral equals the expected number of completed services during an inter–arrival time t under initial remaining service times $x_1, \ldots, x_k$ given that the number of waiting users will always remain positive. Under this condition the servers work independently and hence the above expectation equals k times the respective expectation for one single server. The latter is given by

$$E_t = \sum_{n=0}^{\infty} \int_0^t B^{*n}(t-x)\mu\,(1-B(x))\; dx$$

defining the iterated convolutions by $B^{*0}(t-x) := 1$, $B^{*1}(t-x) := B(t-x)$ and $B^{*n+1}(t-x) := (B * B^{*n})(t-x)$ for all $x \le t \in \mathbb{R}_0^+$. Now it is easy to see that

$$\begin{aligned} E_t &= \mu \int_0^t (1 - B(x) + B(t-x) - B(x)B(t-x) + B^{*2}(t-x) - \ldots)dx \\ &= \mu t \end{aligned}$$

since

$$\begin{aligned} \int_0^t B^{*n+1}(t-x)\; dx &= \int_{x=0}^t \int_{u=0}^{t-x} B(t-x-u)\; dB^{*n}(u)\; dx \\ &= \int_{u=0}^t \int_{x=0}^{t-u} B(t-x-u)\; dx\; dB^{*n}(u) \\ &= \int_{u=0}^t \int_{x=0}^{t-u} B(x)\; dx\; dB^{*n}(u) \\ &= \int_{x=0}^t \int_{u=0}^{t-x} B(x)\; dB^{*n}(u)\; dx \\ &= \int_0^t B(x)B^{*n}(t-x)\; dx \end{aligned}$$

for all $n \in \mathbb{N}_0$. Hence we have

$$\int_{(\mathbb{R}_0^+)^k} \beta(x)\pi(dx) = \int_0^\infty k \cdot \mu\; t\; a(t)\; dt = k \cdot \mu \cdot \bar{\lambda}$$

which yields the equivalence.

5. Stationary Distribution

Let us assume in this section that the queueing process Q and hence the embedded chain X are ergodic. Then the stationary distribution can be derived in a straightforward manner from Tweedie (1982) and standard renewal results. This shall be explained shortly in this section.

As in representation (4), partition the state space by $E = \bigcup_{l=0}^{\infty} E_l$ with $E_l = \{(l, x) : x \in (\mathbb{R}_0^+)^k\}$ called the lth level of E. Define the n–step taboo probabilities with taboo level $l \in \mathbb{N}_0$ recursively by

$${}_lP^1((n,x), \{m\} \times A) := P^X((n,x), \{m\} \times A)$$

and

$$
{}_lP^{n+1}((n,x),\{m\}\times A) := \\
= \sum_{k=l+1}^{\infty} \int_{(\mathbb{R}_0^+)^k} P^X((k,y),\{m\}\times A)\ {}_lP^n((n,x),\{k\}\times dy)
$$

for all $n, m \in \mathbb{N}_0$, $x \in (\mathbb{R}_0^+)^k$ and $A \in \mathcal{B}^k$.

Further define the kernel $L(x,A) := \sum_{n=1}^{\infty} {}_0P^n((0,x),\{0\}\times A)$. This kernel represents the transition probability kernel of a Markov chain Y which is embedded in the chain X at times of visiting level 0. If Q is ergodic, then L has an invariant probability measure ${}_0\Pi$ satisfying

$$
{}_0\Pi(A) = \int_{(\mathbb{R}_0^+)^k} L(y,A)\,{}_0\Pi(dy)
$$

for all $A \in \mathcal{B}^k$. The stationary distribution Π of the chain X is now given by

$$
c := \Pi(\{0\}\times E) = \left(\int_{(\mathbb{R}_0^+)^k} \sum_{n=1}^{\infty} n \cdot {}_0P^n((0,y),\{0\}\times(\mathbb{R}_0^+)^k)\ {}_0\Pi(dy) \right)
$$

and

$$
\Pi(\{0\}\times A) = c \cdot {}_0\Pi(A)
$$

at level 0, as well as

$$
\Pi(\{n\}\times A) = c \cdot \int_{(\mathbb{R}_0^+)^k} S^n(y,A)\ {}_0\Pi(dy)
$$

for $n \in \mathbb{N}$, with kernels S^n defined recursively by

$$
S^1(x,A) := \sum_{n=1}^{\infty} {}_0P^n((0,x),\{1\}\times A)
$$

and

$$
S^n(x,A) := \int_{(\mathbb{R}_0^+)^k} S(y,A)\ S^{n-1}(x,dy)
$$

for all $n > 1$. The kernel S^1 can be obtained as the limit $S^1 = \lim_{n\to\infty} S_n$ with $S_0 := 0$ and

$$
S_n(x,B) = \sum_{m=0}^{\infty} \int_{(\mathbb{R}_0^+)^k} A_m(y,B)\ S_{n-1}^m(x,dy) \tag{8}
$$

for $n \in \mathbb{N}$, $S_n^0 := I$ denoting the identity kernel, and S_{n-1}^m the mth iteration of the kernel S_{n-1}.

Having obtained the stationary distribution Π of X, the stationary distribution p of the queueing process Q is given as follows: Denote $A^c(t) := 1 - \int_0^t a(u)du$ as the complementary distribution function of the inter–arrival time. Further define the kernel

$$K_t((n,x), \{m\} \times A) := P(Q_t \in \{m\} \times A, H > t | Q_0 = (n,x)) \tag{9}$$
$$= A^c(t) \cdot P^I(t; (n,x), \{m\} \times A) \tag{10}$$

with H denoting the first inter–arrival time. Then we finally have

$$p(\{m\} \times A) = \bar{\lambda} \cdot \sum_{n=m}^{\infty} \int_{(\mathbb{R}_0^+)^k} \Pi(n, dx) \int_0^{\infty} K_t((n,x), \{m\} \times A)\ dt \tag{11}$$

for all $m \in \mathbb{N}_0$ and $A \in \mathcal{B}^k$. The sum may begin at $n = m$ because of representation (10) and remark 1.

6. An Example

In order to illustrate the algorithm given above, this section provides an application of it to the GI/M/2 queueing system. Let the inter–arrival times be iid with uniform distribution within the interval $[0, 1]$. The service time distributions shall be chosen as exponential ones with some rate $\mu \in \mathbb{R}^+$. This queue is stable under the condition $\mu > 1$ (see condition (7)).

The choice of exponential service times leads to great simplifications, as the memoryless behaviour of service facilities allows to neglect the service time already passed. Instead, it suffices to keep track of the number of users being served, which reduces the kernels in the above presentations to matrices. Indeed, numerical methods for the calculus of general kernels have not been developed as maturely yet as to provide calculation packages like MatLab or Octave. First steps in this direction have been taken by Nielsen and Ramaswami (1997), and this seems very promising as a start.

In the case of a GI/M/2 queue, the auxiliary state space can be reduced decisively from $(\mathbb{R}_0^+)^2$ to $\{0, 1, 2\}$. Hence, instead of kernels we have simple 3×3 - matrices as the blocks which build the transition matrix P^X (see equation (4) in section 4). They can be determined in a straightforward manner as

$$B_{0;0,0} = 1 - \frac{1}{\mu}\left(1 - e^{-\mu}\right), \quad B_{0;0,1} = \frac{1}{\mu}\left(1 - e^{-\mu}\right), \quad B_{0;0,2} = 0,$$
$$B_{0;1,0} = 1 - B_{0;1,1} - B_{0;1,2}, \quad B_{0;1,1} = \frac{1}{\mu}\left(1 - e^{-\mu}\right)^2$$

$$B_{0;1,2} = \frac{1}{2\mu}\left(1 - e^{-2\mu}\right)$$
$$B_{0;2,0} = 1 - B_{0;2,1} - B_{0;2,2} - A_{0;2,2}$$
$$B_{0;2,1} = 2 \cdot \frac{1}{\mu}\left(1 - e^{-\mu}\right)^2 - \frac{1}{\mu}\left(1 - e^{-2\mu}(1 + 2\mu)\right)$$
$$B_{0;2,2} = \frac{1}{2\mu}\left(1 - e^{-2\mu}(1 + 2\mu)\right)$$

for the matrix $B_0 = (B_{0;i,j})_{0 \le i,j \le 2}$.

By definition of the queue and the auxiliary variable, the matrices A_n with $n \in \mathbb{N}_0$ have positive entries only in position $(2, 2)$, since a positive number of waiting users implies that all service facilities are busy. These entries are easily determined as

$$A_{n;2,2} = \frac{1}{2\mu}\left(1 - e^{-2\mu}\sum_{m=0}^{n}\frac{(2\mu)^m}{m!}\right)$$

for all $n \in \mathbb{N}_0$. By the same argument, the entries $B_{n;i,j}$ vanish for $i \neq 2$. For the rest, we obtain

$$B_{n;2,0} = 1 - B_{n;2,1} - B_{n;2,2} - \sum_{m=0}^{n} A_{m;2,2}$$

with

$$B_{n;2,1} = 2 \cdot (B_{n-1;2,1} - A_{n+1;2,2}) \quad \text{and} \quad B_{n;2,2} = A_{n+1;2,2}$$

for all $n \in \mathbb{N}$.

Because of the structure of the matrices A_n, the kernel S reduces to a number which can be computed according to equation (8) without any problems. Denote by $(\pi_n^* : n \in \mathbb{N}_0)$ the distribution of the number of users in the system immediately before arrivals. Then we have $\pi_n^* = \pi_2^* \cdot S^{n-2}$ for all $n \geq 3$, and because of that

$$(\pi_0^*, \pi_1^*, \pi_2^*) = (\pi_0^*, \pi_1^*, \pi_2^*) \cdot \left(B_0 + \sum_{n=1}^{\infty} S^n B_n\right)$$

The entries of the matrix on the right can readily be computed.

The following results have been obtained from an implementation in Octave, a free MatLab version. The variation has been confined to the service rate, since any variation of the upper bound of the interval for inter–arrival times merely coincides with another time scaling. Because of $\pi_n^* = \pi_2^* \cdot S^{n-2}$ for all $n \geq 3$, it suffices to show the results for π_0^*, π_1^*, π_2^*, and S. These are given in the following table:

	π_0^*	π_1^*	π_2^*	S
$\mu = 4$	0.4487	0.4097	0.1417	0.1463
$\mu = 3$	0.4205	0.3828	0.1967	0.2088
$\mu = 2$	0.3580	0.3439	0.2981	0.3608
$\mu = 1.5$	0.2141	0.2253	0.2543	0.5464
$\mu = 1.1$	0.0722	0.0846	0.1123	0.8668
$\mu = 1.01$	0.0088	0.0106	0.0145	0.9852

Having computed the distribution $(\pi_n^* : n \in \mathbb{N}_0)$, the stationary distribution of the queueing process can easily be computed according to equation (11), which in the present case reduces to

$$p(m) = \frac{1}{2} \cdot \sum_{n=m}^{\infty} \pi_n^* \int_0^1 (1-t) P^I(t;n,m)\, dt$$

for all $m \in \mathbb{N}_0$. The values for $P^I(t;n,m)$ are determined in a straightforward manner, with the same arguments as used for the derivation of the matrices A_n and B_n.

References

Breuer, L.: 2003, 'Algorithmic Solutions for the M/G/k Queue'. In: D. Al-Dabass (ed.): *UKSIM2003, Proc. of the Sixth National Conference of the UK Simulation Society, Emmanuel College, Cambridge, 9–11 April 2003*. pp. 32–36, UK Simulation Society.

Costa, O.: 1990, 'Stationary distributions for piecewise-deterministic Markov processes.'. *J. Appl. Probab.* **27**(1), 60–73.

Davis, M.: 1984, 'Piecewise-deterministic Markov processes: A general class of non- diffusion stochastic models.'. *J. R. Stat. Soc., Ser. B* **46**, 353–388.

Davis, M.: 1993, *Markov models and optimization.* London: Chapman & Hall.

Dufour, F. and O. L. Costa: 1999, 'Stability of piecewise-deterministic Markov processes.'. *SIAM J. Control Optimization* **37**(5), 1483–1502.

Kiefer, J. and J. Wolfowitz: 1955, 'On the theory of queues with many servers.'. *Trans. Am. Math. Soc.* **78**, 1–18.

Meyn, S. and R. Tweedie: 1993, *Markov chains and stochastic stability.* Berlin: Springer-Verlag.

Neuts, M. F.: 1978, 'Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector.'. *Adv. Appl. Probab.* **10**, 185–212.

Nielsen, B. F. and V. Ramaswami: 1997, 'A Computational Framework for a Quasi Birth and Death Process with a Continuous Phase Variable'. In: V. Ramaswami and P. Wirth (eds.): *Teletraffic Contributions for the Information Age. Proceedings of the 15th International Teletraffic Congress - ITC 15*. Amsterdam, pp. 477–486, Elsevier, North–Holland.

Tweedie, R.: 1982, 'Operator-geometric stationary distributions for Markov chains, with application to queueing models.'. *Adv. Appl. Probab.* **14**, 368–391.