Transient Analysis of M/M/1 Queues in Discrete Time with General Server Vacations

Böhm, Walter; Mohanty, Sri Gopal

Transient Analysis of M/M/1 Queues in Discrete Time with General Server Vacations

Walter Böhm, Sri Gopal Mohanty

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 42
September 1994

http://statmath.wu-wien.ac.at/

Transient analysis of M/M/1 queues in discrete time with general server vacations

W. Böhm S.G. Mohanty

Abstract

In this contribution we consider an M/M/1 queueing model with general server vacations. Transient and steady state analysis are carried out in discrete time by combinatorial methods. Using weak convergence of discrete parameter Markov chains we obtain also formulas for the corresponding continuous time queueing model. As a special case we discuss briefly a queueing system with T-policy operating.

1 Introduction

In 1961 with the appearance of his paper *The probability law of the busy period for two types of queueing processes* by L. Takács in *Operations Research* L. Takács set a new mile stone in the theory of queues. In this paper Takács used his beautiful generalization of the classical ballot theorem to give a very elegant and almost elementary derivation for the density of the busy period in M/G/1 and GI/M/1 queues. Takács was aware of the fact that the ballot theorem could solve even harder problems and in 1967 he published his book *Combinatorial Methods in the Theory of Stochastic Processes* in which he demonstrated by an ingenious combination of analytical and combinatorial methods how to solve problems arising in a broad class of stochastic processes, including queueing and dam processes.

Following the same spirit we consider the problem of transient analysis of a Markovian queueing model in discrete time with general server vacations. Generally speaking, vacation models are characterized by the fact that the server is unavailable over occasional intervals of time. During these time periods the server may be busy with maintenance work, e.g. working on other queues or be scanning for new work to do, a situation which we encounter typically in modern telecommunication systems. An excellent overview on this subject is given in Doshi (1985), but see also the new monograph of Takagi (1992).

Transient analysis will be carried out basically in discrete time, because it has been recognized that by no means all queueing phenomena can be modelled appropriately in continuous time. This is primarily due to the propagation of digital technologies in communication and data transmission and the emergence of powerful computer networks. In these systems measurement of time is essentially discrete in nature and changes in the system states can occur only at discrete epochs of time. Continuous time models for queueing problems in these systems are at best approximations, a discrete time analysis seems to be more appropriate.

However, adequateness is not the only motivation for using discrete time methods. They are also conceptually simple and rely very naturally on elementary combinatorial techniques as an alternative to analytical tools. The method of solution we propose to adopt (see Böhm and Mohanty (1991)) comes out very clearly in many of Takács earlier papers (e.g. see Takács (1975)). Whereas the classical procedure of problem solving in queueing theory is to set up a system of differential equations and to solve it in a top-to-bottom fashion (if possible) in terms of integral transforms, our method works in a bottom-to-top fashion. There we try to find decompositions of the original problem into simpler substructures and solve the problem at the substructural level so as to combine them finally into one. The procedure is not new, but somehow not very commonly used. While solving at the substructural level, one is able to apply either analytic or combinatorial methods, although the latter turns out to be elegant and often elementary. No doubt this fact was pointed out by Feller (1968), but it was Takács who really unravelled the power of combinatorial methods particularly in the application to stochastic processes.

Still there is another useful feature of discrete time methods. It is possible to derive in a surprisingly simple way formulas for continuous time models by a limiting procedure. In particular it can be shown that under certain (rather mild) regularity conditions Markovian queueing processes in discrete time converge in the weak sense to continuous time processes. So we are in the lucky situation of killing two birds in one shot: once we have a discrete time result, the corresponding continuous time formula may be derived almost without additional efforts in a fairly easy way. This approach has been introduced by Mohanty and Panny (1990a, 1990b) in the discrete time analogue of the M/M/1 model and further applied to such models with control policy in Böhm and Mohanty (1990) and to models involving batches in Böhm and Mohanty (1991). Incidentally, in the past the solution of a discrete time model was approximated by a solution to an appropriate continuous time model.

The plan of our paper is as follows: in section 2 we give a precise formulation of the model, introduce the basic notation and provide some technical prerequisites. In section 3 we study the transient distribution of the queue length by combinatorial methods and present also steady state results. Section 4 is devoted to a short discussion of the above mentioned limiting procedure and finally we discuss briefly an interesting special case, the T-policy queueing model.

2 Prerequisites

Let $Q(n)$ denote the number of customers waiting in a line queued up in front of a single server with the convention that the customer being in service at time n is included in $Q(n)$. Whenever the queue becomes empty the server goes on a vacation of random duration V, where V is the generic random variable of the sequence of vacation periods $V_1, V_2, \ldots$, which are i.i.d. with probability function $v(n) = P(V = n)$ and $v(0) = 0$. When the server returns from a vacation and finds at least one customer waiting he (she) resumes service immediately and continues to serve, until the queue becomes empty again. If there are no customers waiting at his (her) return, the server goes on another vacation without delay.

Let us assume that

- the probability that a customer arrives during the time slot $(n-1, n]$ is α,
- the probability that a customer finishes service during a given time slot equals γ, if the server is busy and zero if the server is on a vacation.
- The probability that the queue remains unchanged during a particular time slot equals $\beta = 1 - \alpha - \gamma > 0$, if the server is busy, and $1 - \alpha$ otherwise.
- events in different slots are independent and the probability that more than one arrival (departure) occurs during a given slot is zero.

Let us introduce an indicator function $\zeta(n)$, which we define in the following way:

$$\zeta(n) = \begin{cases} 1 & \text{if the server is busy at time } n \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

and let $R(n)$ the residual vacation time of the server at time n with the proviso that $R(n) \equiv 0$, if $\zeta(n) = 1$.

It is not difficult to verify that the trivariate process $\{Q(n), R(n), \zeta(n)\}$ is a Markov chain with state space $\mathcal{A} = \mathcal{I} \cup \mathcal{B}$, where

$$\mathcal{I} = \{(i, j, 0)\} \qquad i, j \geq 0,$$

and

$$\mathcal{B} = \{(k, 0, 1)\} \qquad k > 0.$$

It will be convenient to associate with this process a path in the (x, y)-plane which is constructed as follows: an arrival during a slot is represented by an upward step $(1, 1)$, a service completion by a downward step $(1, -1)$ and the event that the queue length remains unchanged during a given slot by a horizontal step $(1, 0)$. For instance, if we assume that the server is busy at time zero with $m > 0$ customers waiting and also busy at time n with $k > 0$ customers in the queue, a typical path may look like the following:

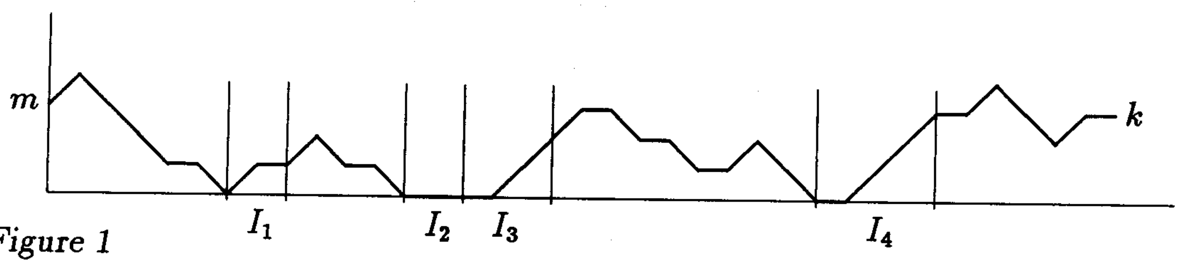

Figure 1

In this graph we have separated busy and vacation periods by vertical straight lines, the vacation periods are marked by I_k.

Figure 1 shows very clearly that the process $Q(n)$ behaves quite differently during busy and vacation periods. For vacation periods we immediately verify that

$$\begin{aligned} P(Q(n) = k, \zeta(n) = 0 | V = n) &= \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \\ &= h_k(n), \end{aligned}$$

a simple binomial distribution. To describe the behaviour of $Q(n)$ during busy periods we introduce a random walk process $Z(n) = \sum_{i=1}^{n} X_i$, with $Z(0) = 0$ and increments X_i being i.i.d with probability function

$$P(X_i = 1) = \alpha, \quad P(X_i = 0) = \beta, \quad P(X_i = -1) = \gamma.$$

Suppose now that a busy period starts with $a > 0$ customers, then it is not difficult to see, that the transitions of $Q(n)$ coincide with the transitions of the process $a + Z(n)$ as long as $Q(n) > 0$. More precisely, if

$$T_{a,0} = \inf\{n : a + Z(n) = 0\},$$

then the following holds

$$\begin{aligned} P(Q(n) = k, T_{a,0} > n, \zeta(n) = 1 | Q(0) = a, \zeta(0) = 1) &= P(a + Z(n) = k, T_{a,0} > n) \\ &= P_{a,k}^{(0)}(n), \end{aligned}$$

where $P^{(0)}_{a,k}(n)$ are the zero avoiding transition probabilities of the process $a+Z(n)$. Some basic results about the process $Z(n)$ and the functions $P^{(0)}_{a,k}(n)$ are summarized in the following Lemma:

Lemma 1 *Let $w_k(n) = P(Z(n) = k)$. Then for $-n \leq k \leq n$:*

$$w_k(n) = \sum_i \binom{2i-k}{i}\binom{n}{2i-k}\alpha^i\gamma^{i-k}\beta^{n-2i+k}. \tag{2}$$

The transition functions $w_k(n)$ satisfy the recurrence relations

$$kw_k(n) = n\alpha w_{k-1}(n-1) - n\gamma w_{k+1}(n-1), \tag{3}$$

valid for all k and $n > 0$, with initial condition $w_k(0) = \delta_{0k}$, and

$$\gamma(n+k+1)w_{k+1}(n) = \alpha(n-k+1)w_{k-1}(n) - \beta k w_k(n), \tag{4}$$

valid for $k \geq -n$ with the initial conditions

$$w_{-n}(n) = \gamma^n \quad \text{and} \quad w_k(n) = 0 \text{ for } k < -n \text{ and } k > n.$$

The zero avoiding transition probabilities are given by

$$\begin{aligned} P^{(0)}_{m,k}(n) &= w_{k-m}(n) - \rho^{-m}w_{k+m}(n) \qquad (5)\\ &= w_{k-m}(n) - \rho^k w_{-k-m}(n), \qquad (6) \end{aligned}$$

where $\rho = \alpha/\gamma$, the traffic intensity and $m, k > 0$.

Proof. Let $T(s) = \alpha s + \beta + \gamma/s$. Then clearly

$$w_k(n) = [s^k]T^n(s),$$

$[s^k]$ being the familiar coefficient operator, and (2) follows by trinomial expansion of $T^n(s)$. From (2) we also get the important symmetry relation

$$w_{-k}(n) = \rho^{-k}w_k(n). \tag{7}$$

Formulas (3) and (4) follow directly from (2). However, it is instructive to see how they can be derived using generating functions. To prove (3) we use the fact that the residue of the first derivative of a Laurent series is zero identically, thus

$$\begin{aligned} \text{Res}\left(\frac{d}{ds}\frac{T^n(s)}{s^k}\right) &= \text{Res}\left(-\frac{k}{s^{k+1}}T^n(s) + \frac{n}{s^k}T^{n-1}(s)(\alpha - \frac{\gamma}{s^2})\right) \\ &= 0, \end{aligned}$$

which yields (3) after some simplification. In a similar manner we prove (4). Consider

$$T\frac{d}{ds}T^n = n(\alpha - \frac{\gamma}{s^2})T^n. \tag{8}$$

Observing that

$$\frac{d}{ds}T^n = \sum_{k\geq -n} ks^{k-1}w_k(n),$$

we get (4) by comparing coefficients of like powers of s on both sides of (8). Note that (4) offers a very convenient way to compute the functions $w_k(n)$. In fact, the number of floating point operations required to calculate simultaneously all $w_i(n)$ for $-n \leq i \leq k$ is of order $O(n)$.

Formulas (5) and (6) follow from the reflection principle (see Mohanty and Panny (1990a)), and the proof is complete.

One final remark to Lemma 1. From (6) we obtain also

$$\begin{aligned} f_a(n) &= \gamma P_{a,1}^{(0)}(n) \\ &= \gamma w_{-a+1}(n) - \alpha w_{-a-1}(n). \end{aligned}$$

On the other hand we get from the recurrence relation (3) for $k = -a$:

$$-aw_{-a}(n) = n\alpha w_{-a-1}(n-1) - n\gamma w_{-a+1}(n),$$

which yields the well known formula

$$f_a(n) = \frac{a}{n}w_{-a}(n). \tag{9}$$

3 Transient and steady state solutions

For simplicity we will confine our discussion about the transient behaviour of the Markov chain $\{Q(n), R(n), \zeta(n)\}$ to the case $\zeta(0) = 1$, i.e. we assume that the server is busy initially. However, with a suitable adaptation the methods outlined below apply also to the case $\zeta(0) = 0$.

For $j = 0, 1$ let

$${}_jP_{m,k,i}(n) = P(Q(n) = k, \zeta(n) = j, R(n) = i | Q(0) = m, \zeta(0) = 1)$$

and define

$$ {}_0P_{m,k}(n) = \sum_{i \geq 1} {}_0P_{m,k,i}(n), $$

and

$$ {}_1P_{m,k}(n) = {}_1P_{m,k,0}(n). $$

Note that we have set $R(n) = 0$, if $\zeta(n) = 1$. Assume first that $\zeta(n) = 1$, i.e. the server is busy at time n, and consider the event

$$ \{Q(n) = k, \zeta(n) = 1 | Q(0) = m, \zeta(0) = 1\}, \quad m, k > 0 $$

A typical path corresponding to this event is given below in Figure 2. Vertical lines separate busy and vacation periods, vacation periods are marked again by I_i, and there is also a tail segment which starts when the last completed vacation begins. At the end of this last vacation there must be at least one customer in the queue, because otherwise the busy period, which is still in progress at time n, would not have started. It will be convenient to treat this tail segment separately, hence we consider first the path up to the beginning of the last vacation period.

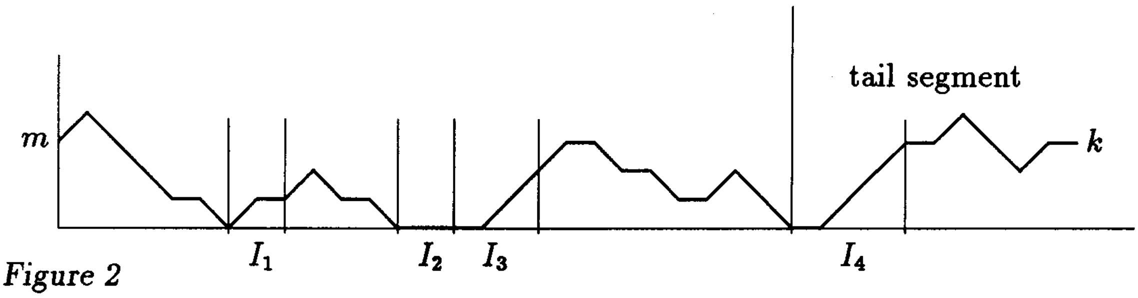

Figure 2

Now let us cut out those path segments which correspond to vacations. Suppose that there are $i > 0$ such vacations, not counting the vacation belonging to the tail segment and let us fix the number of customer arrivals during those vacations to $k_1, k_2, \ldots, k_i$. Some of these k's may be zero. For instance in the example above we have $k_1 = 1, k_2 = 0, k_3 = 2$. Next we join the remaining segments, which correspond to busy periods, to a new path. This path starts at height $m + K_i$, where $K_i = \sum_{\nu=1}^{i} k_\nu$ and terminates with a first passage on the line $y = 0$. The probability function of this path is $f_{m+K_i}(n)$ with pgf. $\xi^{m+K_i}(s)$.

Consider now the segments due to vacation periods and suppose that the vacations have lengths $a_1, \ldots, a_i$. Joining these segments leads to a non-decreasing path which has probability function

$$ v(a_1) \cdot v(a_2) \ldots \cdot v(a_i) \binom{a_1}{k_1} \alpha^{k_1} (1-\alpha)^{a_1 - k_1} \ldots \binom{a_i}{k_i} \alpha^{k_i} (1-\alpha)^{a_i - k_i}. $$

If we set

$$v^{i*}(n) = P(V_1 + V_2 + \ldots + V_i = n),$$

and sum over all tuples $a_1, \ldots, a_i$ and $k_1, \ldots, k_i$, such that

$$a_1 + \ldots + a_i = n, \quad k_1 + \ldots + k_i = K_i,$$

we obtain for the probability function of this path

$$P(V_1 + \ldots + V_i = n)\binom{n}{K_i}\alpha^{K_i}(1-\alpha)^{n-K_i} = v^{i*}(n)h_{K_i}(n).$$

Let us define a cycle as the succession of a vacation and a completed busy period and let $a_{i,m}(n)$ denote the probability that the initial busy period (which starts with $m > 0$ customers) and $i > 0$ cycles have in total length n. It follows that

$$a_{i,m}(n) = \sum_{\nu \geq 0} [v^{i*}(n)h_\nu(n)] * f_{m+\nu}(n), \tag{10}$$

and if $i = 0$, then

$$a_{0,m}(n) = f_m(n). \tag{11}$$

We note in passing that the series (10) is finite.

It remains to determine the probability of the tail segment. It consists of a completed vacation which terminates with $\nu > 0$ customers (if ν would be zero, the server would not resume service), followed by a path segment, which leads from height ν to terminal height k without touching or crossing the line $y = 0$. It has probability function

$$b_k(n) = \sum_{\nu > 0} [v(n)h_\nu(n)] * P^{(0)}_{\nu,k}(n). \tag{12}$$

If the server is on a vacation at time n, i.e. $\zeta(n) = 0$, only the probability of the tail segment changes. If at time n there are k customers waiting and the residual vacation time equals $i \geq 1$, then the tail segment has probability function

$$d_{k,i}(n) = h_k(n)v(n+i).$$

Let $d_k(n) = \sum_{i \geq 1} d_{k,i}(n)$. Then

$$d_k(n) = P(V > n)h_k(n), \quad k \geq 0. \tag{13}$$

We summarize our results in the following

Theorem 1 *For all $m, k > 0$*

$$ {}_1P_{m,k}(n) = P^{(0)}_{m,k}(n) + \sum_{i \geq 0} a_{i,m}(n) * b_k(n). \tag{14}$$

For $k \geq 0, m > 0$

$$ {}_0P_{m,k}(n) = \sum_{i \geq 0} a_{i,m}(n) * d_k(n). \tag{15}$$

Steady state results

Let T_{00} denote the length of a cycle which may also be degenerate, i.e. that at the end of a vacation there are no customers in the queue. The probability function of T_{00} is given by (10) for $i = 1$ and $m = 0$:

$$P(T_{00} = n) = a_{1,0}(n).$$

It is not difficult to determine the pgf. of T_{00}. If we set $\mathcal{C}(s) = \sum_{n \geq 0} s^n P(T_{00} = n)$, then

$$\begin{aligned}
\mathcal{C}(s) &= \sum_{a \geq 0} [\sum_{n \geq 0} s^n v(n) \binom{n}{a} \alpha^a (1-\alpha)^{n-a}] \xi^a(s) \\
&= \sum_{n \geq 0} s^n v(n) (\alpha \xi(s) + 1 - \alpha)^n \\
&= \mathcal{V}(s(\alpha \xi(s) + 1 - \alpha)),
\end{aligned}$$

where $\mathcal{V}(s)$ is the pgf. of $v(n)$. Thus

$$\begin{aligned}
E(T_{00}) &= \frac{d}{ds} \mathcal{V}(s(\alpha \xi(s) + 1 - \alpha))|_{s=1} \\
&= E(V) \frac{\gamma}{\gamma - \alpha},
\end{aligned}$$

which is finite if and only if $\rho < 1$ and vacations have finite expectation. By the Erdös-Feller-Pollard Theorem (Feller (1966, p. 365)) we have

$$\lim_{n \to \infty} {}_1P_{m,k}(n) = \frac{1}{E(T_{00})} \sum_{i \geq 1} b_k(i),$$

and

$$\lim_{n \to \infty} {}_0P_{m,k}(n) = \frac{1}{E(T_{00})} \sum_{i \geq 1} d_k(i).$$

Let $b_k = \sum_{i\geq 0} b_k(i)$ and $d_k = \sum_{i\geq 0} d_k(i)$. For b_k we find:

$$\begin{aligned} b_k &= \sum_{n\geq 0} b_k(n) \\ &= \sum_{a>0}\sum_{n\geq 0}[v(n)h_a(n)] * P^{(0)}_{a,k}(n). \end{aligned} \tag{16}$$

Since

$$P^{(0)}_{a,k}(n) = w_{k-a}(n) - \rho^k w_{-k-a}(n), \tag{17}$$

we are left with the problem of finding

$$\sum_{n\geq 0}[v(n)h_a(n)] * w_{k-a}(n) \quad k = 0, \pm 1, \pm 2, \ldots$$

After rearranging terms this sum reduces to

$$\sum_{n\geq 0} v(n)h_a(n)\sum_{i\geq 0} w_{k-a}(i) \quad k = 0, \pm 1, \pm 2, \ldots$$

It remains to determine $\sum_{i\geq 0} w_k(i)$. Using a beautiful theorem of Takács (1967, Theorem 3, pp 20) we find that for $\rho < 1$

$$\sum_{i\geq 0} w_k(i) = \begin{cases} \frac{\rho^k}{\gamma-\alpha} & k \geq 0 \\ \frac{1}{\gamma-\alpha} & k < 0 \end{cases}$$

Thus according to the Erdös-Feller-Pollard Theorem we have

Theorem 2 *If $E(V) < \infty$ and $\gamma > \alpha$, then*

$$\lim_{n\to\infty} {}_1P_{m,k}(n) = \frac{(\gamma-\alpha)b_k}{\gamma E(V)} \quad m, k > 0 \tag{18}$$

and

$$\lim_{n\to\infty} {}_0P_{m,k}(n) = \frac{(\gamma-\alpha)d_k}{\gamma E(V)} \quad m > 0, k \geq 0. \tag{19}$$

It can be checked that the steady state probability that the server is idle is independent of the distribution of the vacations and equals $1-\rho$ which is the steady state probability of the server being idle in simple M/M/1. With some algebra:

$$P(\zeta = 0) = \sum_{k\geq 0} d_k \frac{\gamma-\alpha}{\gamma E(V)}$$

$$
\begin{aligned}
&= \frac{1-\rho}{E(V)} \sum_{k\geq 0} \sum_{j\geq 0} P(V > j) h_k(j) \\
&= \frac{1-\rho}{E(V)} \sum_{j\geq 0} P(V > j) \\
&= 1-\rho.
\end{aligned}
$$

4 Continuous time results

As we have remarked in the introduction it is one of the major advantages of discrete time queueing models that we are killing two birds in one shot in the sense, that once we have a solution in discrete time, we may obtain results for continuous time queueing processes by an appropriate limiting procedure. In this section we will demonstrate how such a passage to the limit can be performed.

To begin with let us assume that the time interval $(0, t)$ has been split into n subintervals (slots) of equal length $\Delta = t/n$. Let $\alpha = \lambda t/n$, $\gamma = \mu t/n$, where $\lambda, \mu > 0$ are fixed real numbers. Thus we interpret $Q(n)$ as the number of customers in the system at time $n\Delta$. Similarly

$$
\begin{aligned}
v(n) &= P(V = n\Delta) \\
&= P(n(\Delta - 1) < V \leq n\Delta),
\end{aligned}
$$

and let us assume that

$$
\lim_{n\to\infty} \frac{1}{\Delta} v(n) = \psi(t) \tag{20}
$$

exists and is a nondefective probability density concentrated on R^+. Using these conventions we may regard the discrete time process $Q(n)$, which we have discussed above, as an approximation of the continuous time queueing process $Q(t)$, in which customers arrive according to a Poisson process with rate λ, having exponential service times with expectation $1/\mu$ and vacations with density $\psi(t)$. In fact it can be shown that under the above cited conditions the Markov chain $\{Q(n), R(n), \zeta(n)\}$ converges in the weak sense to the continuous time Markov process $\{Q(t), R(t), \zeta(t)\}$ (see Ethier and Kurtz, 1985, pp. 167, Theorem 2.5).

To derive the continuous time analogue of Theorem 1, we will need some technical prerequisites.

Lemma 2

$$\begin{aligned}\lim_{n\to\infty} w_k(n) &= \omega_k(t) \\ &= \rho^{k/2} e^{-(\lambda+\mu)t} I_k(2t\sqrt{\lambda\mu}),\end{aligned} \tag{21}$$

where $I_k(z)$ denote the modified Bessel functions.

A proof of this lemma may be found in Böhm and Mohanty (1991).

Lemma 2 implies also that

$$\begin{aligned}\lim_{n\to\infty} P^{(0)}_{m,k}(n) &= \Pi^{(0)}_{m,k}(t) \\ &= \omega_{k-m}(t) - \rho^{-m}\omega_{k+m}(t) \\ &= \omega_{k-m}(t) - \rho^{k}\omega_{-k-m}(t),\end{aligned} \tag{22}$$

and

$$\begin{aligned}\lim_{n\to\infty} f_m(n)\frac{1}{\Delta} &= \phi_m(t) \\ &= \lim_{\Delta\to 0}\frac{1}{\Delta}P(n(\Delta-1) < T_{m,0} \leq n\Delta) \\ &= \frac{m}{t}\omega_{-m}(t).\end{aligned} \tag{23}$$

Consider next

$$\lim_{n\to\infty} v^{i*}(n)h_\nu(n)\frac{1}{\Delta},$$

where $v^{i*}(n)$ denotes the i-fold convolution of $v(n)$. Now it is not difficult to verify that

$$\lim_{n\to\infty} h_\nu(n) = \frac{e^{-\lambda t}(\lambda t)^\nu}{\nu!} = \eta_\nu(t),$$

and in virtue of (20)

$$\lim_{\Delta\to 0} v^{i*}(n)h_\nu(n)\frac{1}{\Delta} = \psi^{i*}(t)\eta_\nu(t), \tag{24}$$

where $\psi^{i*}(t)$ is the i-fold convolution of the density $\psi(t)$. A bit more difficult is the determination of the limit of $a_{i,m}(n)$, which we will denote by $\alpha_{i,m}(t)$ (there will be no confusion with the jump probability α).

Let us set $\Delta = t/n$ and $j\Delta = s_j, 0 \leq s_j \leq t$, for $j = 0,\ldots,n$. Clearly, $s_0 = 0$ and $s_n = t$. Because of (28) and (29) it follows that

$$v^{i*}(j)h_\nu(j)\frac{n}{t} = \psi^{i*}(s_j)\eta_\nu(s_j) + O(n^{-\epsilon}), \quad \epsilon > 0,$$

and

$$f_{m+\nu}(j)\frac{n}{t} = \phi_{m+\nu}(s_j) + O(n^{-\tau}), \quad \tau > 0.$$

Therefore

$$\begin{aligned}
&\Big[[v^{i*}(n)h_\nu(n)] * f_{m+\nu}(n)\Big]\frac{n}{t} = \\
&= \sum_{j=0}^{n}(v^{i*}(j)h_\nu(j)\frac{n}{t})(f_{m+\nu}(n-j)\frac{n}{t})\frac{t}{n} \\
&= \sum_{j=0}^{n}[\psi^{i*}(s_j)\eta_\nu(s_j) + O(n^{-\epsilon})][\phi_{m+\nu}(t-s_j) + O(n^{-\tau})]\Delta \\
&= \sum_{j=0}^{n}\psi^{i*}(s_j)\eta_\nu(s_j)\phi_{m+\nu}(t-s_j)\Delta + O(n^{-\kappa}), \quad \kappa > 0 \\
&\to \int_0^t \psi^{i*}(s)\eta_\nu(s)\phi_{m+\nu}(t-s)\,ds \quad \text{for } n \to \infty.
\end{aligned}$$

In a similar way we find

$$\begin{aligned}
\lim_{n\to\infty} b_k(n) &= \beta_k(t) \\
&= \sum_{\nu>0}\int_0^t \psi(s)\eta_\nu(s)\Pi^{(0)}_{\nu,k}(t-s)\,ds,
\end{aligned}$$

and

$$\begin{aligned}
\lim_{n\to\infty} d_k(n) &= \sigma_k(t) \\
&= \eta_k(t)\int_t^\infty \psi(s)\,ds
\end{aligned}$$

Putting these pieces together, we have proved

Theorem 3 *Let ${}_1\Pi_{m,k}(t)$ and ${}_0\Pi_{m,k}(t)$ denote the limiting forms of (14) and (15) in the above defined sense, then*

$$ {}_1\Pi_{m,k}(t) = \Pi^{(0)}_{m,k}(t) + \sum_{i\geq 0}\int_0^t \alpha_{i,m}(s)\beta_k(t-s)\,ds, \tag{25}$$

and

$$ {}_0\Pi_{m,k}(t) = \sum_{i\geq 0}\int_0^t \alpha_{i,m}(s)\sigma_k(t-s)\,ds. \tag{26}$$

5 The special case T-policy

One of the principle motivations for considering server vacation models is to utilize the idle time of the server for additional tasks, like maintenance work or the service of other queues. A typical example has been reported by Heyman (1977): consider the operation of electronic telephone switching machines. They are controlled by computer programs and continuous monitoring of the queue reduces the amount of CPU time available for other tasks. To allow a more efficient use of CPU time one may introduce a scan interval of fixed length. Thus if the queue becomes empty, the server is virtually turned off and scans the queue for new customers only at fixed time points. If there are customers then service starts, otherwise the server is turned off again. This is T-policy, there are repeated vacations which have fixed length, say $T \geq 1$. Heyman has discussed the determination of an optimal value for the length of the scan interval. Let us see, how the formulas derived above look in the case of T-policy.

The probability function of the length of a vacation is given by

$$v(n) = \delta_{n,T}, \tag{27}$$

where δ_{ij} is the Kronecker delta. The specification (27) leads to considerable simplifications in our formulas: we have for $i > 0$

$$a_{i,m}(n) = \begin{cases} \sum_{a \geq 0} h_a(Ti) f_{m+a}(n - Ti) & n > Ti \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

and

$$\begin{aligned} b_k(n) &= \sum_{a>0} h_a(T)\, P^{(0)}_{a,k}(n - T) \\ d_k(n) &= \begin{cases} h_k(n) & n < T \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

For functions b_k and d_k, which we require for the steady state distribution, we find:

$$\begin{aligned} b_k &= \frac{\rho^k}{\gamma - \alpha} \sum_{a=1}^{k} \binom{T}{a} \gamma^a (1 - \alpha)^{T-a} + \frac{1}{\gamma - \alpha} \sum_{a>k} \binom{T}{a} \alpha^a (1 - \alpha)^{T-a} \\ &\quad - \frac{\rho^k}{\gamma - \alpha} (1 - (1 - \alpha)^T), \end{aligned}$$

which simplifies to

$$b_k = \frac{\rho^k}{\gamma - \alpha} [(1 + \gamma - \alpha)^T - 1],$$

if $k \geq T$. Similarly

$$d_k = \begin{cases} \sum_{j=1}^{T-1} h_k(j) & k < T \\ 0 & \text{otherwise} \end{cases}$$

Let us next determine the expected queue length in steady state. We obtain

$$\begin{aligned} E(Q|\zeta = 0) &= \frac{1}{1-\rho} \sum_{k \geq 0} k d_k \frac{(\gamma - \alpha)}{\gamma T} \\ &= \frac{1}{T} \sum_{j=1}^{T-1} \sum_{k \geq 0} k h_k(j) \\ &= \frac{\alpha}{2}(T-1). \end{aligned}$$

To find an expression for $E(Q|\zeta = 1)$ we have to evaluate $\sum_{k>0} k b_k$. Using

$$\sum_{k \geq a} k \rho^k = \frac{(1-\rho) a \rho^a + \rho^{a+1}}{(1-\rho)^2},$$

we find

$$\sum_{k>0} \frac{k\rho^k}{\gamma - \alpha} \sum_{a=1}^{k} \binom{T}{a} \gamma^a (1-\alpha)^{T-a} = \frac{1}{\gamma - \alpha} [\frac{\alpha T}{1-\rho} + \frac{\rho(1-(1-\alpha)^T)}{(1-\rho)^2}].$$

Furthermore

$$\begin{aligned} \frac{1}{\gamma - \alpha} \sum_{k>0} k \sum_{a>k} \binom{T}{a} \alpha^a (1-\alpha)^{T-a} &= \frac{1}{\gamma - \alpha} \sum_{k \geq 0} \frac{k(k-1)}{2} \binom{T}{k} \alpha^k (1-\alpha)^{T-k} \\ &= \frac{1}{\gamma - \alpha} \frac{\alpha^2 T(T-1)}{2}, \end{aligned}$$

and

$$\sum_{k>0} \frac{k\rho^k (1-(1-\alpha)^T)}{\gamma - \alpha} = \frac{\rho(1-(1-\alpha)^T)}{(\gamma-\alpha)(1-\rho)^2}.$$

Thus

$$\sum_{k>0} k b_k = \frac{\alpha T}{(\gamma - \alpha)(1-\rho)} + \frac{\alpha^2 T(T-1)}{2(\gamma - \alpha)},$$

and therefore

$$E(Q|\zeta = 1) = \frac{1}{1-\rho} + \frac{\alpha(T-1)}{2}.$$

For the unconditional expectation we get finally

$$E(Q) = \frac{\alpha(T-1)}{2} + \frac{\rho}{1-\rho}. \tag{29}$$

Let us now discuss T-policy in continuous time. For brevity we will confine our discussion to the limiting functions $\psi(t), \alpha_{i,m}(t), \beta_k(t)$ and $\sigma_k(t)$, which are the basic ingredients of Theorem 3.

Assume that $T = [nc]$, where $c > 0$ is a constant and let $tc = T^\star$. Then

$$\psi(t) = \lim_{n\to\infty} \frac{1}{\Delta}\delta_{n\Delta - T\Delta, 0} = \delta(t - T^\star),$$

where $\delta(t)$ is the Dirac delta function. Since $\psi^{i*}(t) = \delta(t - T^\star i)$, it follows that

$$\begin{aligned} \alpha_{i,m}(t) &= \sum_0^t \psi^{i*}(s)\eta_\nu(s)\phi_{m+\nu}(t-s)\,ds \\ &= U(t - T^\star i)\eta_\nu(T^\star i)\phi_{m+\nu}(t - T^\star i), \end{aligned}$$

$U(t)$ denoting Heavyside's unit step function. We have also

$$\begin{aligned} \beta_k(t) &= \sum_{\nu>0}\int_0^t \psi(s)\eta_\nu(s)\Pi^{(0)}_{\nu,k}(t-s)\,ds \\ &= U(t-T^\star)\sum_{\nu>0}\eta_\nu(T^\star)\Pi^{(0)}_{\nu,k}(t-T^\star), \end{aligned}$$

and

$$\begin{aligned} \sigma_k(t) &\doteq \eta_k(t)\int_t^\infty \psi(s)\,ds \\ &= \eta_k(t)U(T^\star - t). \end{aligned}$$

6 References

[1] Böhm W., Mohanty S.G.(1993) The transient solution of M/M/1 queues under (M,N)-policy. A combinatorial approach. *Journal of Statistical Planning and Inference*, **34**, 23-33.

[2] Böhm W., Mohanty S.G.(1991) On discrete time Markovian N-policy queues involving batches. To appear in *Sankyā, Ser. A*.

[3] Doshi B.T. (1985) Stochastic decomposition in a GI/G/1 queue with vacations. *J. Appl. Prob.*, **22**, 419-428.

[4] Ethier S.N., Kurtz T.G. (1986) *Markov Processes. Characterization and Convergence.*, New York, John Wiley and Sons.

[5] Feller W. (1968) *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., New York, John Wiley and Sons.

[6] Feller W. (1966) *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Ed., New York, John Wiley and Sons.

[7] Heyman D.P. (1977) The T-policy for the M/G/1 queue. *Management Science*, **23**, 7, 775-778.

[8] Mohanty S.G., Panny W. (1990a) A discrete time analogue of the M/M/1 queue and the transient solution: a geometric approach. *Sankyā, Ser. A*, **52**, 364-370.

[9] Mohanty S.G., Panny W. (1990b) A discrete time analogue of the M/M/1 queue and the transient solution. Proceedings of the 3rd. Hungarian Colloquium on Limit Theorems in Probability and Statistics, P. Révész ed., Amsterdam: North Holland (1990), 417-424.

[10] Takács L. (1961) The probability law of the busy period for two types of queueing processes. *Opns. Res.*, **9**, 402-407.

[11] Takács L. (1967) *Combinatorial Methods in the Theory of Stochastic Processes*, New York, John Wiley and Sons.

[12] Takács L. (1975) Combinatorial and analytic methods in the theory of queues. *Adv. Appl. Prob.*, **7**, 607-635.

[13] Takagi H. (1991) *Queueing Analysis. Vol 1, A Foundation of Performance Analysis*, Amsterdam, North Holland.

Walter Böhm,
Department of math. Statistics
University of Economics
Vienna, Austria

S. G. Mohanty
Mathematics and Statistics Department
McMaster University
Hamilton, Canada