Single-server queue with Markov dependent inter-arrival and service times

TU/e technische universiteit eindhoven

SPOR-Report 2002-18

Single-server queue with Markov dependent inter-arrival and service times

I.J.B.F. Adan
V.G. Kulkarni

SPOR-Report
Reports in Statistics, Probability and Operations Research

Eindhoven, December 2002
The Netherlands

/department of mathematics and computing science

SPOR-Report
Reports in Statistics, Probability and Operations Research

Eindhoven University of Technology
Department of Mathematics and Computing Science
Probability theory, Statistics and Operations research
P.O. Box 513
5600 MB Eindhoven - The Netherlands

Secretariat: Main Building 9.10
Telephone: + 31 40 247 3130
E-mail: wscosor@win.tue.nl or wsbsecr@win.tue.nl
Internet: http://www.win.tue.nl/math/bs/cosor.html

ISSN 0926 4493

Single-server queue with Markov dependent inter-arrival and service times

I.J.B.F. Adan* and V.G. Kulkarni†

December 9, 2002

Abstract

In this paper we study a single-server queue where the inter-arrival times and the service times depend on a common discrete time Markov Chain. This model generalizes the well-known $MAP/G/1$ queue by allowing dependencies between inter-arrival and service times. The waiting time process is directly analyzed by solving Lindley's equation by transform methods. The Laplace Stieltjes transforms (LST) of the steady-state waiting time and queue length distribution are both derived, and used to obtain recursive equations for the calculation of the moments. Numerical examples are included to demonstrate the effect of the auto-correlation of and the cross-correlation between the inter-arrival and service times.

1 Introduction

In the literature much attention has been devoted to single-server queues with Markovian Arrival Processes (MAP), see, e.g., [7] and the references therein. These models typically assume that the service times are iid and independent of the arrival process. The present study concerns single-server queues where the inter-arrival times and the service times depend on a common discrete time Markov Chain. As such the model under consideration is a generalization of the $MAP/G/1$ queue.

A special case is the model with strictly periodic arrivals. These models arise, for example, in the modelling of inventory systems using periodic ordering policies. Queueing models with periodic arrival processes have been studied in, e.g., [10, 11, 6, 3].

Models with dependencies between inter-arrival and service time have been studied by several authors. The model in [2] can be interpreted as a queueing model with dependence

*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

†Department of Operations Research, University of North Carolina, CB 3180, Chapel Hill, N.C. 27599

between a service request and the subsequent inter-arrival time. In [1] the authors analyse a variant of the $M/G/1$ queue in which the service times of arriving customers depend on the length of the interval between their arrival and the previous arrival.

Starting point of the analysis is Lindley's equation for the waiting times. We solve this equation using the transform techniques and derive the LST of the steady-state waiting time distribution. By exploiting a well-known relation between the waiting time of a customer and the number of customers left behind by a departing customer we find the LST of the queue length distribution at departure epochs and at arbitrary time points. The LST's are then used to derive simple recursive equations for the moments of the waiting time and queue length.

The paper is organized as follows. In Section 2 we present the queueing model under consideration. The waiting time process is analyzed in Section 3. We first derive the LST of the steady-state waiting time distribution in subsection 3.1. This transform is used in subsection 3.2 to obtain a system of recursive equations for the moments of the waiting time. In the subsequent section we study the queue length process. Numerical examples are presented in Section 5.

2 Queueing model

We consider a single-server queue, where customers are served in order of arrival. Let $A_n, n \geq 1$, denote the time between the nth and $(n-1)$th arrival, and $S_n, n \geq 0$, the service time of the nth arrival. We assume that the sequences $\{A_n, n \geq 1\}$ and $\{S_n, n \geq 0\}$ are auto-correlated as well as cross-correlated. The nature of this dependence is described below.

The inter-arrival and service times are regulated by an irreducible discrete-time Markov chain $\{Z_n, n \geq 0\}$ with state space $\{1, 2, \ldots, N\}$ and transition probability matrix P. More precisely, the tri-variate process $\{(A_n, S_n, Z_n), n \geq 0\}$ has the following probabilistic structure:

$$\begin{aligned}
&P(A_{n+1} \leq x, S_n \leq y, Z_{n+1} = j \mid Z_n = i, A_n, (A_r, S_r, Z_r), 0 \leq r \leq n-1) \\
&\quad = P(A_1 \leq x, S_0 \leq y, Z_1 = j \mid Z_0 = i) \\
&\quad = G_i(y) p_{i,j} (1 - e^{-\lambda_j x}),
\end{aligned}$$

$x, y \geq 0; i, j = 1, 2, \ldots, N$. Thus random variable A_{n+1} has an exponential distribution and S_n has an arbitrary distribution, where both are independent of the past, given Z_n.

Let τ_i be the mean of the service time distribution G_i, and $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ be the stationary distribution of $\{Z_n, n \geq 0\}$. Then the system is stable if

$$\sum_{i=1}^{N} \pi_i \tau_i < \sum_{i=1}^{N} \pi_i \lambda_i^{-1},$$

since the left-hand side is the mean service time of a customer, and the right-hand side is the mean inter-arrival time between two consecutive customers in steady state. Let $\Gamma = \text{diag}(\tau_1, \ldots, \tau_N)$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$, and e be the column vector of ones. Then we can write the above stability condition in a matrix form as follows:

$$\pi(\Lambda^{-1} - \Gamma)e > 0. \tag{1}$$

Remark 2.1 The arrival process is a Markovian Arrival Process (MAP) (see [8]), where each transition of Z corresponds to an arrival. But transitions without arrivals can be mimicked be allowing customers with zero service times. If the distribution of S_n does not depend on Z_n, i.e., $G_i(\cdot) = G(\cdot)$ for all i, then the model reduces to the $MAP/G/1$ queue, which is a special case of the $BMAP/G/1$ queue (see [7]).

Remark 2.2 Periodic arrivals can be derived as a special case of the present model by setting $p_{i,i+1} = p_{N,1} = 1$ for $i = 1, 2, \ldots, N-1$. The waiting time process for the periodic model has also been studied in [3], where functional equations for the stationary distributions of waiting times are derived. These equations formulate a Hilbert Boundary Value problem, which can be solved if the LST's of all the interarrival time distributions or the LST's of all the service time distributions are rational.

Remark 2.3 In steady-state the cross-correlation between A_n and S_n is given by

$$\rho(A_n, S_n) = \rho(A_0, S_0) = \frac{\sum_{i=1}^{N} \pi_i(\lambda_i^{-1} - \lambda^{-1})(\tau_i - \tau)}{\left\{\sum_{i=1}^{N} \pi_i(\lambda_i^{-1} - \lambda^{-1})^2 \sum_{i=1}^{N} \pi_i(\tau_i - \tau)^2\right\}^{1/2}},$$

where $\lambda^{-1} = \sum_{i=1}^{N} \pi_i \lambda_i^{-1}$ and $\tau = \sum_{i=1}^{N} \pi_i \tau_i$.

3 The waiting time process

In this section we study the LST of the limiting distribution of the waiting time, and its moments.

3.1 Steady state LST

Let W_n denote the waiting time of the nth customer. Using the notation $E(X; A)$ to mean $E(X \cdot 1_A)$ for any event A, define

$$\phi_i^n(s) = E(e^{-sW_n}; Z_n = i), \qquad \text{Re}(s) \geq 0, n \geq 0, i = 1, 2, \ldots, N.$$

and, assuming the limit exists, define

$$\phi_i(s) = \lim_{n\to\infty} \phi_i^n(s), \quad i = 1, 2, \ldots, N.$$

The next theorem gives the equations satisfied by the transforms

$$\phi(s) = [\phi_1(s),\ \phi_2(s),\ \cdots,\ \phi_N(s)].$$

First we need the following notation:

$$\begin{aligned}
\tilde{G}_i(s) &= \int_0^\infty e^{-st} dG_i(t), \qquad 1 \le i \le N, \\
\tilde{G}(s) &= \mathrm{diag}(\tilde{G}_1(s),\ \tilde{G}_2(s),\ \cdots,\ \tilde{G}_N(s)), \\
H_{i,j}(s) &= \tilde{G}_i(s) p_{i,j} \lambda_j, \qquad 1 \le i, j \le N, \\
H(s) &= [H_{i,j}(s)] = \tilde{G}(s) P \Lambda. \qquad (2)
\end{aligned}$$

Theorem 3.1 *Provided condition (1) is satisfied, the transform vector $\phi(s)$ satisfies*

$$\begin{aligned}
\phi(s)[H(s) + sI - \Lambda] &= sv, \qquad (3) \\
\phi(0)e &= 1, \qquad (4)
\end{aligned}$$

where $v = [v_1,\ v_2,\ \cdots,\ v_N]$ is given by

$$v_j = \sum_{i=1}^N \phi_i(\lambda_j) \tilde{G}_i(\lambda_j) p_{i,j}.$$

Proof: Let T_n denote the sojourn time of the nth customer, i.e., $T_n = W_n + S_n$, $n \ge 0$. The waiting times W_n satisfy Lindley's equation (see [5]),

$$W_{n+1} = (W_n + S_n - A_{n+1})^+ = (T_n - A_{n+1})^+, \qquad n \ge 0,$$

where $(x)^+ = \max\{x, 0\}$. From Lindley's equation we obtain the following equation for the transforms $\phi_j^{n+1}(s)$, $j = 1, \ldots, N$,

$$\begin{aligned}
\phi_j^{n+1}(s) &= E(e^{-sW_{n+1}}; Z_{n+1} = j) \\
&= \sum_{i=1}^N P(Z_n = i) E(e^{-sW_{n+1}}; Z_{n+1} = j | Z_n = i) \\
&= \sum_{i=1}^N P(Z_n = i) p_{i,j} E(e^{-s(T_n - A_{n+1})^+} | Z_n = i, Z_{n+1} = j) \\
&= \sum_{i=1}^N P(Z_n = i) p_{i,j} \Bigg[E(\int_{x=0}^{T_n} e^{-s(T_n - x)} \lambda_j e^{-\lambda_j x} dx | Z_n = i) \\
&\qquad + E(\int_{x=T_n}^{\infty} \lambda_j e^{-\lambda_j x} dx | Z_n = i) \Bigg] \\
&= \sum_{i=1}^N P(Z_n = i) p_{i,j} E(e^{-sT_n} \frac{\lambda_j}{\lambda_j - s}(1 - e^{-(\lambda_j - s)T_n}) + e^{-\lambda_j T_n} | Z_n = i) \\
&= \sum_{i=1}^N p_{i,j} \left[\frac{\lambda_j}{\lambda_j - s} \phi_i^n(s) \tilde{G}_i(s) - \frac{s}{\lambda_j - s} \phi_i^n(\lambda_j) \tilde{G}_i(\lambda_j) \right], \qquad (5)
\end{aligned}$$

It is clear that $\phi_j^n(s)$ tends to $\phi_i(s)$ if the stability condition (1) is satisfied. Hence, letting $n \to \infty$ in (5), and rearranging, we get

$$\sum_{i=1}^{N} p_{i,j}\lambda_j\phi_i(s)\tilde{G}_i(s) - (\lambda_j - s)\phi_j(s) = s\sum_{i=1}^{N} p_{i,j}\phi_i(\lambda_j)\tilde{G}_i(\lambda_j), \qquad j = 1, \ldots, N. \tag{6}$$

Note that the sum on the right-hand side of the above equation is denoted by v_j. Then we can rewrite (6) in matrix form yielding (3). Equation (4) is just the normalization equation. This completes the proof.

Clearly, we need to determine the unknown vector v in equation (3). For that we need to study the solutions of

$$\det(H(s) + sI - \Lambda) = 0. \tag{7}$$

The following theorem gives the number and placement of the solutions to the above equation.

Theorem 3.2 *Equation (7) has exactly N solutions s_i, $1 \le i \le N$, with $s_1 = 0$ and $Re(s_i) > 0$ for $2 \le i \le N$.*

Proof: See Appendix.

With the above result we now give a method of determining the vector v in the following theorem.

Theorem 3.3 *Suppose the condition of stability (1) is satisfied, and the $N - 1$ solutions s_i, $2 \le i \le N$ with $Re(s_i) > 0$ to equation (3) are distinct. Let a_i be a non-zero column vector satisfying*

$$(H(s_i) + s_iI - \Lambda)a_i = 0, \quad 2 \le i \le N.$$

Then v is given by the unique solution to the following N linear equations:

$$va_i = 0, \quad 2 \le i \le N, \tag{8}$$

$$v\Lambda^{-1}e = \pi(\Lambda^{-1} - \Gamma)e. \tag{9}$$

Proof: Since s_i satisfies equation (7), it follows that there is a non-zero column vector a_i such that

$$(H(s_i) + s_iI - \Lambda)a_i = 0, \qquad i = 1, 2, \ldots, N.$$

In particular $a_1 = e$, the column vector of ones. Post-multiplying equation (3) with $s = s_i$ by a_i, we get

$$\phi(s_i)[H(s_i) + s_iI - \Lambda]a_i = s_iva_i = 0, \qquad i = 1, 2, \ldots, N.$$

Since $s_i \neq 0$, for $2 \leq i \leq N$, v must satisfy equation (8). To derive the remaining equation, we take the derivative of equation (3) with respect to s, yielding

$$\phi(s)[H'(s) + I] + \phi'(s)(H(s) + sI - \Lambda) = v.$$

Setting $s = 0$ we get

$$\phi(0)(H'(0) + I) + \phi'(0)(P - I)\Lambda = v.$$

Post-multiplying by $\Lambda^{-1}e$ gives

$$\phi(0)(H'(0) + I)\Lambda^{-1}e + \phi'(0)(P - I)\Lambda\Lambda^{-1}e = b\Lambda^{-1}e.$$

Using $(P - I)e = 0$, $H'(0) = -\text{diag}(\tau)P\lambda$ and $\phi(0) = \pi$ (where the latter follows from (3) with $s = 0$ and the normalization equation (4)), the above can be simplified to

$$\pi(\Lambda^{-1} - \Gamma)e = v\Lambda^{-1}e.$$

The uniqueness of the solution follows from the general theory of Markov chains that under the condition of stability, there is a unique stationary distribution and thus also a unique solution $\phi(s)$ to the equations (3) and (4). This completes the proof.

3.2 Steady state moments

Once v is known, the entire transform vector $\phi(s)$ is known. We can use it to compute the moments of the waiting time in steady state. They are given in the following theorem. First some notation:

$$\begin{aligned}
m_{r,i} &= \lim_{n\to\infty} E(W_n^r; Z_n = i), \qquad r = 0, 1, 2, \ldots, \; 1 \leq i \leq N, \\
m_r &= [m_{r,1},\ m_{r,2},\ \cdots,\ m_{r,N}], \qquad r = 0, 1, 2, \ldots, \\
\tau_{r,i} &= \int_0^\infty x^r dG_i(x), \qquad r = 0, 1, 2, \ldots, \; 1 \leq i \leq N, \\
\Gamma_r &= \text{diag}[\tau_{r,1},\ \tau_{r,2},\ \cdots,\ \tau_{r,N}], \qquad r = 0, 1, 2, \ldots
\end{aligned}$$

Note that $\tau_{1,i} = \tau_i$ and $\Gamma_1 = \Gamma$. We assume that the above moments exist.

Theorem 3.4 *The moment-vectors m_r satisfy the following recursive equations:*

$$m_0 = \pi, \tag{10}$$

$$m_1(I - P) = m_0(\Gamma_1 P - \Lambda^{-1}) + v\Lambda^{-1}, \tag{11}$$

$$m_1(\Lambda^{-1} - \Gamma_1)e = \frac{1}{2} m_0 \Gamma_2 e, \tag{12}$$

$$m_r(I - P) = r m_{r-1}(\Gamma_1 P - \Lambda^{-1}) + \sum_{k=2}^{r} \binom{r}{k} m_{r-k}\Gamma_k P, \tag{13}$$

$$m_r(\Lambda^{-1}e - \Gamma_1 e) = \frac{1}{r+1}\left[\sum_{k=2}^{r+1} \binom{r+1}{k} m_{r+1-k}\Gamma_k e\right]. \tag{14}$$

Proof: We have

$$\begin{aligned}\phi(s) &= \sum_{r=0}^{\infty}(-1)^r m_r \frac{s^r}{r!},\\ \tilde{G}(s) &= \sum_{r=0}^{\infty}(-1)^r \Gamma_r \frac{s^r}{r!}.\end{aligned}$$

From equation (2) it follows that

$$H(s) = \sum_{r=0}^{\infty}(-1)^r \Gamma_r P \Lambda \frac{s^r}{r!}.$$

Substituting in equation (3) we get

$$\sum_{r=0}^{\infty}(-1)^r m_r \frac{s^r}{r!}\left(\sum_{r=0}^{\infty}(-1)^r \Gamma_r P \Lambda \frac{s^r}{r!} + sI - \Lambda\right) = sv.$$

Equating the coefficients of s^0 we get:

$$m_0(\Gamma_0 P \Lambda - \Lambda) = 0.$$

Since $\Gamma_0 = I$, and Λ is invertible this simplifies to

$$m_0(I - P) = 0.$$

We also know that $m_0 e = 1$. But these equations have a unique solution π, and hence we get equation (10). Next, equating the coefficients of s^1 and simplifying, we get (11). Multiplying this equation by e yields equation (9). Now $(I-P)$ is non-invertible with rank $N-1$, hence we need an additional equation. We get that by equating the coefficients of s^2, which, after simplification, yields

$$m_2(I - P) = m_0\Gamma_2 P + 2m_1(\Gamma_1 P - \Lambda^{-1}).$$

Multiplying by e, we get equation (12). This gives the required additional equation for m_1. Equations (11) and (12) uniquely determine m_1. Note that m_1 depends in Γ_2, the diagonal matrix of second moments of the service times, as expected. Proceeding in this fashion, equating coefficients of s^r, $r \geq 2$, we get equation (13). Using the same equation for $r+1$ and multiplying it by e yields equation (14). Equations (13) and (14) uniquely determine m_r. This completes the proof.

Remark 3.5 It is not essential to assume the existence of all the moments of all the service times. If $k+1$ is the first integer for which at least one entry of G_k becomes infinite, the above equations can still be used to determine m_r for $0 \leq r \leq k-1$.

4 Queue Length Distribution

Now we will analyze the queue length distribution in the queueing system described in Section 2.

4.1 Steady state LST

In this subsection we study the LST of the queue length distribution at departures and at arbitrary times in steady state. Toward this end, first define $N(t)$ to be the number of arrivals up to time t, with $N(0) = 0$, and the continuous time Markov chain $\{Z(t), t \geq 0\}$ to be

$$Z(t) = Z_{N(t)+1}, \qquad t \geq 0.$$

We are interested in

$$\psi_{i,j}(z,t) = E(z^{N(t)}; Z(t) = j \mid Z(0) = i), \qquad t \geq 0,\ 1 \leq i, j \leq N.$$

Let

$$\psi(z,t) = [\psi_{i,j}(z,t)]$$

be the matrix of the generating functions. The following theorem gives a method of computing $\psi(z,t)$. First we need some notation. Let $-\mu_i(z)$, $1 \leq i \leq N$, be the N eigenvalues of $\Lambda P z - \Lambda$, assumed to be distinct, and let $y_i(z)$ and $x_i(z)$ be the orthonormal left and right eigenvectors corresponding to $-\mu_i(z)$. The matrices $A_i(z)$ are defined as

$$A_i(z) = x_i(z) y_i(z), \qquad 1 \leq i \leq N.$$

Then we have the following theorem.

Theorem 4.1 *The generating function matrix $\psi(z,t)$ is given by*

$$\psi(z,t) = e^{(\Lambda P z - \Lambda)t} = \sum_{i=1}^{N} e^{-\mu_i(z)t} A_i(z).$$

Proof: It is easy to show that ψ satisfies the following differential equation (cf. Section 2.1 in [7]):

$$\frac{\partial}{\partial t}\psi(z,t) = (\Lambda P z - \Lambda)\psi(z,t), \qquad t \geq 0, \tag{15}$$

with the initial condition $\psi(z,0) = I$. Hence the solution is given by

$$\psi(z,t) = e^{(\Lambda P z - \Lambda)t}.$$

This proves the first equality. The second one holds, since we can write (see, e.g., [13])

$$(\Lambda P z - \Lambda)^n == \sum_{i=1}^{N} (-\mu_i(z))^n A_i(z).$$

Using the above theorem we first derive the generating function of the queue length seen by departures in steady state. Let the random variable L^d denote the queue length and Z^d the state of $\{Z(t), t \geq 0\}$ seen by a departure in steady state. We use the following notation:

$$\begin{aligned} g_j(z) &= E(z^{L^d}; Z^d = j), \qquad 1 \leq j \leq N, \\ g(z) &= (g_1(z), \ldots, g_N(z)), \\ \theta_{i,j}(z) &= \phi_i(z)\tilde{G}_i(z)p_{i,j}, \qquad 1 \leq i, j, k \leq N, \\ \Theta(z) &= [\theta_{i,j}(z)]. \end{aligned}$$

The result is given in the next theorem:

Theorem 4.2 *The generating function $g(z)$ is given by*

$$g(z) = \sum_{i=1}^{N} e'\Theta(\mu_i(z))A_i(z). \tag{16}$$

Proof: Suppose that the process $\{(A_n, S_n, Z_n, W_n), n \geq 0\}$ is in steady state. Then the zeroeth customer finds the system in steady state, and finds $Z_0 = i$ with probability π_i. Given that he finds $Z_0 = i$, his service time S_0 has distribution $G_i(\cdot)$, and the LST of his sojourn time $T_0 = S_0 + W_0$ is given by $\phi_i(s)\tilde{G}_i(s)/\pi_i$. Upon arrival of the zeroeth customer the process $\{Z_n, n \geq 0\}$ jumps from $Z_0 = i$ to state $Z_1 = k$ with probability $p_{i,k}$; hence, $Z(0) = Z_1 = k$ with probability $p_{i,k}$. The new customers that arrive during the sojourn time of this customer are exactly the ones that are left behind by the zeroeth customer, so their number is equal to L^d (cf. [4]). Thus, conditional upon $Z_0 = i$ and $T_0 = t$, we have

$$E(z^{L^d}; Z^d = j \mid Z_0 = i; T_0 = t) = \sum_{k=1}^{N} p_{i,k}\psi_{k,j}(z,t)$$

Hence we obtain

$$\begin{aligned} g_j(z) &= \sum_{i=1}^{N} \int_0^\infty E(z^{L^d}; Z^d = j \mid Z_0 = i; T_0 = t) dP(T_0 \leq t; Z_0 = i) \\ &= \sum_{i=1}^{N} \int_0^\infty dP(T_0 \leq t; Z_0 = i) \sum_{k=1}^{N} p_{i,k}\psi_{k,j}(z,t) \\ &= \sum_{i=1}^{N} \int_0^\infty dP(T_0 \leq t; Z_0 = i) \sum_{k=1}^{N} p_{i,k} \sum_{l=1}^{N} e^{-\mu_l(z)}[A_l(z)]_{k,j} \\ &= \sum_{l=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{N} \phi_i(\mu_l(z))\tilde{G}_i(\mu_l(z))p_{i,k}[A_l(z)]_{k,j} \end{aligned} \tag{17}$$

$$= \sum_{l=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\Theta_{i,k}(\mu_l(z))[A_l(z)]_{k,j}$$
$$= \sum_{l=1}^{N}[e'\Theta(\mu_l(z))A_l(z)]_j.$$

We now connect the queue length as seen by a departure in steady state to the queue length at an arbitrary time. Let the random variable L denote the queue length at an arbitrary time, and Z the state of $\{Z(t), t \geq 0\}$ at an arbitrary time. Define

$$\begin{aligned} h_j(z) &= E(z^L; Z = j), \qquad 1 \leq j \leq N, \\ h(z) &= (h_1(z), \ldots, h_N(z)), \\ \lambda^{-1} &= \sum_{i=1}^{N} \pi_i \lambda_i^{-1}. \end{aligned}$$

Note that λ is the mean overall arrival rate. The connection between $h(z)$ and $g(z)$ is formulated in the following theorem (cf. Section 3.3 in [7]):

Theorem 4.3 *The generating function $h(z)$ satisfies*

$$h(z)(\Lambda P z - \Lambda) = \lambda(z-1)g(z). \tag{18}$$

Proof: Let $L(t)$ denote the queue length at time t; the process $\{(L(t), Z(t)), t \geq 0\}$ has state space $\{(n,i), n \geq 0, i = 1, \ldots, N\}$. In steady state the average number of transitions per unit time out of state (n,i) is equal to the number of transitions into state (n,i); hence

$$\begin{aligned} &P(L^d = n-1, Z^d = i)\lambda + P(L = n, Z = i)\lambda_i \\ &\quad = P(L^d = n, Z^d = i)\lambda + \sum_{j=1}^{N} P(L = n-1, Z = j)\lambda_j p_{j,i}. \end{aligned}$$

Taking the transforms gives equation (18).

Remark 4.4 The classical approach for $M/G/1$-type models is to consider the embedded Markov chain at departure epochs first, and then to determine the waiting time distribution by using the connection between this distribution and the departure distribution (cf. [7]). For the present model, the type of customer to be served next does not only depend on the customer type of the departing customer, but also on the sojourn time of the departing customer. This feature essentially complicates an embedded Markov chain approach.

4.2 Steady state moments

The results of the previous subsection can be used to determine the factorial moments of the queue length distribution. Define

$$\begin{aligned}
d_{r,i} &= E(L^d(L^d-1)\cdots(L^d-r+1); Z^d=i), \qquad r=0,1,2,\ldots,\ 1\le i\le N,\\
d_r &= [d_{r,1}, d_{r,2}, \ldots, d_{r,N}], \qquad r=0,1,2,\ldots,\\
a_{r,i} &= E(L(L-1)\cdots(L-r+1); Z=i), \qquad r=0,1,2,\ldots,\ 1\le i\le N,\\
a_r &= [a_{r,1}, a_{r,2}, \ldots, a_{r,N}], \qquad r=0,1,2,\ldots.
\end{aligned}$$

The moment vectors d_r may be obtained by differentiating (16) and then using the relation $d_r = g^{(r)}(1)$. However, the derivatives of $\mu_i(z)$ and $A_i(z)$ at $z=1$ may be hard to determine. It is easier to proceed as follows. Let

$$M_r(t) = \frac{\partial^r}{\partial z^r}\psi(1,t), \qquad t\ge 0.$$

The moment matrices $M_r(t)$ satisfy the differential equations (see (15)

$$\frac{d}{dt}M_r(t) = r\Lambda P M_{r-1}(t) + (\Lambda P - \Lambda)M_r(t), \qquad t\ge 0, \tag{19}$$

with the initial condition $M_r(0)=I$ if $r=0$, and $M_r(0)=0$ if $r>0$. Multiplying both sides of (19) by $e^{(\Lambda-\Lambda P)t}$, it is readily seen that these equations can be rewritten as

$$\frac{d}{dt}\left(e^{(\Lambda-\Lambda P)t}M_r(t)\right) = re^{(\Lambda-\Lambda P)t}\Lambda P M_{r-1}(t), \qquad t\ge 0.$$

Hence,

$$M_r(t) = re^{-(\Lambda-\Lambda P)t}\int_0^t e^{(\Lambda-\Lambda P)x}\Lambda P M_{r-1}(x)dx, \qquad t\ge 0, \tag{20}$$

from which it is obvious that $M_r(t)$ can be determined recursively, starting with $r=0$. The following lemma gives the solutions for $r=0$ and $r=1$; the solutions for $r>1$ can be obtained similarly. In the lemma we abbreviated $\mu_i(1)$ and $A_i(1)$ simply by μ_i and A_i.

Lemma 4.5 *The moment matrices $M_0(t)$ and $M_1(t)$ satisfy*

$$\begin{aligned}
M_0(t) &= \sum_{i=1}^N e^{-\mu_i t}A_i;\\
M_1(t) &= \sum_{k=1}^N \sum_{\substack{l=1\\ l\ne k}} \frac{e^{-\mu_k t}-e^{-\mu_l t}}{\mu_l-\mu_k}A_k\Lambda P A_l + \sum_{k=1}^N te^{-\mu_k t}A_k\Lambda P A_k.
\end{aligned}$$

Proof: The expression for $M_0(t)$ immediately follows from Theorem 4.1. Substituting

$$e^{\pm(\Lambda-\Lambda P)t} = \sum_{k=1}^{N} e^{\pm\mu_k t} A_k$$

into (20) for $r = 1$ and using that $A_k^2 = A_k$ and $A_k A_l = 0$ if $k \neq l$, yields

$$\begin{aligned} M_1(t) &= e^{-(\Lambda-\Lambda P)t} \left\{ \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l\neq k}}^{N} \frac{e^{(\mu_k-\mu_l)t} - 1}{\mu_k - \mu_l} A_k \Lambda P A_l + \sum_{k=1}^{N} t A_k \Lambda P A_k \right\} \\ &= \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l\neq k}}^{N} \frac{e^{-\mu_l t} - e^{-\mu_k t}}{\mu_k - \mu_l} A_k \Lambda P A_l + \sum_{k=1}^{N} t e^{\mu_k t} A_k \Lambda P A_k, \end{aligned}$$

which completes the proof of the lemma.

Once $M_r(t)$ is known, the moment vectors d_r follow from (cf. (17))

$$d_{r,j} = \frac{d^r}{dz^r} g_j(1) = \sum_{i=1}^{N} \int_0^{\infty} dP(T_0 \leq t; Z_0 = i) \sum_{k=1}^{N} p_{i,k} [M_r(t)]_{k,j}.$$

The results for d_0 and d_1 are formulated in the following theorem.

Theorem 4.6 *The moment vectors d_0 and d_1 are given by*

$$\begin{aligned} d_0 &= \sum_{i=1}^{N} e' \Theta(\mu_i) A_i; \\ d_1 &= \sum_{i=1}^{N} e' \left\{ \sum_{k=1}^{N} \sum_{\substack{l=1 \\ l\neq k}}^{N} \frac{\Theta(\mu_k) - \Theta(\mu_l)}{\mu_l - \mu_k} A_k \Lambda P A_l - \sum_{k=1}^{N} \Theta'(\mu_k) A_k \Lambda P A_k \right\}. \end{aligned}$$

The moment vectors a_r can be found by exploiting the relationship formulated in Theorem 4.3.

Theorem 4.7 *The moment-vectors a_r satisfy the recursive equations:*

$$a_r \Lambda (P - I) = r(\lambda d_{r-1} - a_{r-1} \Lambda P), \tag{21}$$

$$a_r \Lambda e = \lambda d_r e, \tag{22}$$

where $a_{-1} = d_{-1} = 0$ by convention.

Proof: We have

$$\begin{aligned} g(z) &= \sum_{r=0}^{\infty} d_r \frac{(z-1)^r}{r!}, \\ h(z) &= \sum_{r=0}^{\infty} a_r \frac{(z-1)^r}{r!}. \end{aligned}$$

Substituting in equation (18) we get

$$\sum_{r=0}^{\infty} a_r \frac{(z-1)^r}{r!} \Lambda(P - I + (z-1)P) = \lambda(z-1) \sum_{r=0}^{\infty} d_r \frac{(z-1)^r}{r!},$$

which can be rearranged as

$$\sum_{r=0}^{\infty} \left(\frac{a_r}{r!} \Lambda(P-I) + \frac{a_{r-1}}{(r-1)!} \Lambda P \right) (z-1)^r = \sum_{r=0}^{\infty} \lambda d_{r-1} \frac{(z-1)^r}{(r-1)!}.$$

Equating the coefficients of $(z-1)^r$ gives

$$a_r \Lambda(P - I) = r(\lambda d_{r-1} - a_{r-1} \Lambda P). \tag{23}$$

Since $P - I$ is non-invertible with rank $N - 1$ we need an extra equation. This one is obtained by post-multiplying (23) with e yielding

$$\lambda d_{r-1} e - a_{r-1} \Lambda e = 0.$$

This completes the proof of the theorem.

Remark 4.8 For $r = 0$ the equations (21)-(22) reduce to (note that $d_0 e = 1$)

$$\begin{aligned} a_0 \Lambda(P - I) &= 0, \\ a_0 \Lambda e &= \lambda. \end{aligned}$$

Hence, we get

$$a_0 = \lambda \pi \Lambda^{-1}.$$

The mean queue length $a_1 e$ can also be obtained by application of Little's law, i.e.,

$$a_1 e = \lambda(m_1 + e' \Gamma_1) e.$$

5 Numerical Examples

In this section we present some examples to demonstrate the effects of auto-correlation and cross-correlation of the inter-arrival and service times. In each example we set $N = 4$ and we assume exponential service times.

Example 1: Positively correlated inter-arrival and service times.
The arrival rates and mean service times are given by

$$[\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] = [.1\ 10\ .1\ 10], \quad [\tau_1\ \tau_2\ \tau_3\ \tau_4] = u[5\ .05\ 5\ .05],$$

where u is a parameter, greater or equal to 0. It is readily verified that the cross-correlation between the inter-arrival and service time is equal to 1 (see Remark 2.3). We study two cases. In case (a), the transition probability matrix is set to

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

So the inter-arrival and service times are auto-correlated. This situation will be compared with case (b), where P satisfies

$$P = \begin{pmatrix} .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \end{pmatrix}.$$

In this case the successive inter-arrival times and service times are iid, i.e., there is no auto-correlation. In both cases, the traffic intensity is given by

$$\rho = \frac{\sum \pi_i \tau_i}{\pi_i \lambda_i^{-1}} = .5u.$$

In Figure 1 we present the mean waiting times as a function of the traffic intensity ρ. The results show that there is not much difference between the two cases; in fact, the case with no auto-correlation performs slightly better.

Example 2: Negatively correlated inter-arrival and service times.
The arrival rates and mean service times are given by

$$[\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] = [.1\ 10\ .1\ 10], \quad [\tau_1\ \tau_2\ \tau_3\ \tau_4] = u[.05\ 5\ .05\ 5],$$

where $u \geq 0$. Now the cross-correlation between the inter-arrival and service time is equal to -1. We will again evaluate the mean waiting times for the two cases (a) and (b) as in Example 1, using the transition probability matrices stated there. The results are

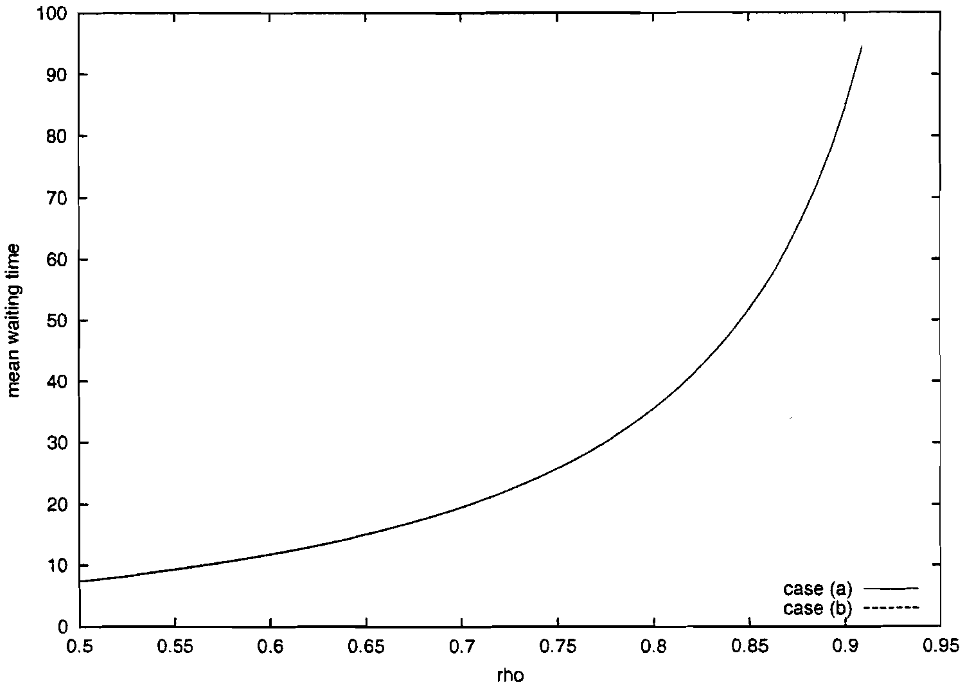

Figure 1: Positive cross-correlation: Mean waiting times as a function of ρ for the case with and without auto-correlation.

presented in Figure 2, as a function of the traffic intensity $\rho = .5u$. Clearly, in this case, auto-correlation is able to exploit the big differences between the mean inter-arrival and service times. The mean waiting times for auto-correlated inter-arrival and service times are substantially less than the ones for no auto-correlation.

Example 3: Independent inter-arrival and service times.
Now the arrival rates and mean service times are given by

$$[\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] = [.1\ 10\ .1\ 10], \quad [\tau_1\ \tau_2\ \tau_3\ \tau_4] = u[5\ 5\ .05\ .05],$$

where $u \geq 0$. Again we consider two cases. In case (a), we use

$$P = \begin{pmatrix} 0 & .5 & 0 & .5 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ .5 & 0 & .5 & 0 \end{pmatrix},$$

so that the inter-arrival times are auto-correlated. In case (b), P is given by

$$P = \begin{pmatrix} .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \end{pmatrix}.$$

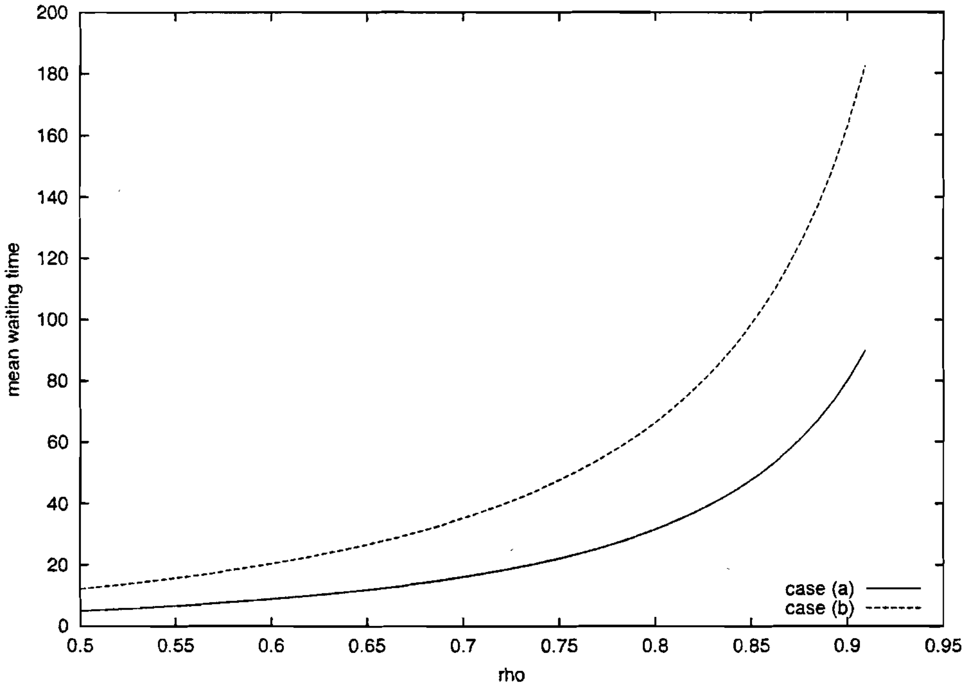

Figure 2: Negative cross-correlation: Mean waiting times as a function of ρ for the case with and without auto-correlation.

so that the inter-arrival times are independent (and hyper-exponentially distributed). In both cases the service times are iid and hyper-exponentially distributed. However, there is no cross correlation between the inter-arrival times and the service times. The traffic intensity is given by $\rho = .5u$ as before. In Figure 3 the mean waiting times are shown for both cases; the results illustrate that the mean waiting times are substantially less for auto-correlated inter-arrival times.

These examples clearly indicate the impact of auto-correlation and cross-correlation on the waiting times.

6 Appendix: Proof of Theorem 3.2

We first assume that for some $\epsilon > 0$ the transforms $\tilde{G}_i(s)$ are analytic for all s with $\mathrm{Re}(s) > -\epsilon$. This holds, e.g., for service time distributions with an exponential tail or for distributions with a finite support.

Let us consider the determinant

$$\det(\tilde{G}(s)P + s\Lambda^{-1} - I) = \det(H(s) + sI - \Lambda)/\det(\Lambda),$$

and let C_δ denote the circle with its center located at $\max_i \lambda_i$ and radius $\delta + \max_i \lambda_i$, with $0 < \delta < \epsilon$. We will prove that the determinant has exactly N zeros inside the circle C_δ for

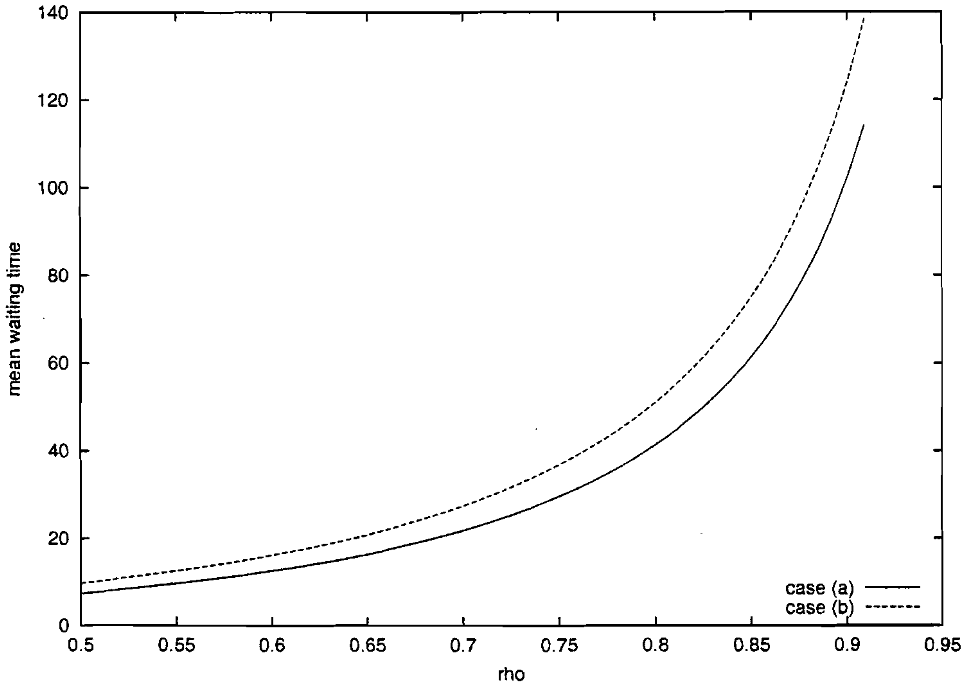

Figure 3: No cross-correlation: Mean waiting times as a function of ρ for the auto-correlated and independent inter-arrival times.

all δ sufficiently small. We follow the main idea in the proof of a similar theorem in the Appendix 2 of [12]. We first prove the following lemma.

Lemma 6.1 *For* $0 \leq u \leq 1$ *and small* $\delta > 0$,

$$\det(u\tilde{G}(s)P + s\Lambda^{-1} - I) \neq 0, \qquad s \in C_\delta. \tag{24}$$

Proof: For $0 \leq u \leq 1$ and $s \in C_\delta$ with $\mathrm{Re}(s) \geq 0$, the matrix $u\tilde{G}(s)P + s\Lambda^{-1} - I$ is diagonally dominant, since

$$\begin{aligned} \mid u\tilde{G}_i(s)p_{i,i} + s/\lambda_i - 1 \mid &\geq \mid s/\lambda_i - 1 \mid - u\tilde{G}_i(0)p_{i,i} \geq 1 + \delta/\lambda_i - u\tilde{G}_i(0)p_{i,i} \\ &> \lambda\tilde{G}_i(0) - u\tilde{G}_i(0)p_{i,i} = \sum_{j\neq i} u\tilde{G}_i(0)p_{i,j} \geq \sum_{j\neq i} \mid u\tilde{G}_i(s)p_{i,j} \mid . \end{aligned} \tag{25}$$

Hence, its determinant is nonzero (see, e.g., pp. 146-147 in [9]). To prove this for $s \in C_\delta$ with $\mathrm{Re}(s) < 0$ we first note that the determinant is nonzero if and only if 0 is not an eigenvalue. So we proceed by studying the eigenvalues of $u\tilde{G}(s)P + s\Lambda^{-1} - I$ near $s = 0$. If we write

$$u\tilde{G}(s)P + s\Lambda^{-1} - I = P - I + s\Lambda^{-1} + ((u-1)\tilde{G}(s) + \tilde{G}(s) - I)P, \tag{26}$$

we see that for (s, u) close to $(0, 1)$, the matrix above is a perturbation of $P - I$. Since P is irreducible, $P - I$ has a simple eigenvalue 0. Then in a neighborhood of $(0, 1)$, there exist differentiable $x(s,u)$ and $\mu(s,u)$ such that

$$(u\tilde{G}(s)P + s\Lambda^{-1} - I)x(s,u) = \mu(s,u)x(s,u), \qquad e'x(s,u) = 1,$$

and such that $\mu(0,1) = 0$ and $x(0,1) = e$. Differentiating this equation with respect to s and setting $s = 0$ and $\lambda = 1$ in the result, we obtain (the subscript s indicates the derivative with respect to s)

$$(P - I)x_s(0,1) + (\Lambda^{-1} - \Gamma)e = \mu_s(0,1)e.$$

Pre-multiplying both sides with π gives

$$\mu_s(0,1) = \pi(\Lambda^{-1} - \Gamma)e. \tag{27}$$

Similarly, by differentiating with respect to u we get

$$\mu_\lambda(0,1) = 1.$$

Hence, for (s,λ) close to $(0,1)$, it holds that

$$\mu(s,u) \approx s\pi(\Lambda^{-1} - \Gamma)e + u - 1.$$

Since $\pi(\Lambda^{-1} - \Gamma)e > 0$ by virtue of (1), we can conclude that $\mu(s,u) \neq 0$ for $s \in C_\delta$ with $\mathrm{Re}(s) < 0$, for small $\delta > 0$ and u close to 1, say $1 - \hat{\delta} \leq u \leq 1$. Finally, for $0 \leq u < 1 - \hat{\delta}$, it can be shown, similarly to (25), that $u\tilde{G}(s)P + s\Lambda^{-1} - I$ is diagonally dominant for $s \in C_\delta$ with $\mathrm{Re}(s) < 0$, provided δ is small enough such that

$$1 - \delta/\lambda_i > (1 - \hat{\delta})\tilde{G}_i(-\delta).$$

Let $f(u)$ be the number of zeros of $\det(u\tilde{G}(s)P + s\Lambda^{-1} - I)$ inside C_δ. Then we have

$$f(u) = \frac{1}{2\pi i}\int_{C_\delta} \frac{\frac{\partial}{\partial s}\det(u\tilde{G}(s)P + s\Lambda^{-1} - I)}{\det(u\tilde{G}(s)P + s\Lambda^{-1} - I)}ds.$$

So $f(u)$ is a continuous function on $[0,1]$, integer-valued, and hence constant. Since $f(0) = N$, this implies that also $f(1) = N$. Letting δ tend to 0, we can conclude that $\det(\tilde{G}(s)P + s\Lambda^{-1} - I)$ has N zeros inside or on C_0.

Clearly, $s = 0$ satisfies

$$\det(\tilde{G}(s)P + s\Lambda^{-1} - I) = 0.$$

That $s = 0$ is a simple solution of the above equation is a consequence of P being irreducible and the stability condition (1); the arguments are presented below.

We evaluate the derivative of the determinant in $s = 0$. Let $D = diag(d_1, \ldots, d_N)$ be the diagonal matrix of eigenvalues of $P - I$, with $d_1 = 0$; since P is irreducible, 0 is simple eigenvalue of $P - I$. Let Y and X denote the matrices of corresponding left and right eigenvectors, with $YX = I$. Hence,

$$P - I = YDX, \qquad YX = I.$$

Now in a neighborhood of the origin, there exist differentiable $D(s)$, $Y(s)$ and $X(s)$ such that

$$\tilde{G}(s)P + s\Lambda^{-1} - I = Y(s)D(s)X(s)$$

and such that $D(0) = D$, $Y(0) = Y$ and $X(0) = X$. Hence,

$$\det(\tilde{G}(s)P + s\Lambda^{-1} - I) = \det(Y(s))\det(D(s))\det(X(s)).$$

By differentiating this equation with respect to s, setting $s = 0$ in the result and using that $d_1 = 0$, we obtain that the derivative of $\det(\tilde{G}(s)P + s\Lambda^{-1} - I)$ in $s = 0$ is equal to $d_1'(0)d_2 \cdots d_N$. This is nonzero, since (see (27))

$$d_1'(0) = \pi(\Lambda^{-1} - \Gamma)e > 0.$$

To finally complete the proof of Theorem 3.2 we have to remove the initial assumption that for some $\epsilon > 0$ the transforms $\tilde{G}_i(s)$ are analytic for all s with $\text{Re}(s) > -\epsilon$. To this end, first consider the 'truncated' service time distributions $G_i^K(x)$ defined as $G_i^K(x) = G_i(x)$ for $0 \leq x < K$ and $G_i^K(x) = 1$ for $x \geq K$. Then Theorem 3.2 holds for the distributions $G_i^K(x)$; by letting K tend to infinity, the result also follows for the original service time distributions.

Remark 6.2 We have not only proved the existence of N solutions of Equation (7), but also that they are located inside or on the circle with its center at $\max_i \lambda_i$ and radius $\max_i \lambda_i$. This is useful in numerical procedures for finding these zeros.

References

[1] S.C. Borst, O.J. Boxma and M. B. Combé (1993) An $M/G/1$ queue with customer collection. *Comm. Statist. Stochastic Models* **9** 341–371.

[2] O.J. Boxma and D. Perry (2001) A queueing model with dependence between service and interarrival times. *European J. Oper. Res.* **128** 611–624.

[3] J.W. Cohen (1996) On periodic Pollaczek waiting time processes. *Athens Conference on Applied Probability and Time Series Analysis, Vol. I (1995)*, 361–378, Lecture Notes in Statist. **114**, Springer, New York.

[4] R. B. Cooper (1981) *Introduction to queueing theory.* North Holland, NY.

[5] L. KLEINROCK (1975) *Queieuing Systems, Vol. I: Theory.* Wiley-Interscience, NY.

[6] A. J. LEMOINE (1989) Waiting time and workload in queues with periodic Poisson input. *J. Appl. Probab.* **26** 390–397.

[7] D. M. LUCANTONI (1991) New results on the single-server queue with a batch Markovian arrival process. *Stochastic Models* **7**, 1–46.

[8] D.M. LUCANTONI, K.S. MEIER-HELLSTERN AND M.F. NEUTS (1990) A single-server queue with server vacations and a class of nonrenewal arrival processes. *Adv. in Appl. Probab.* **22**, 676–705.

[9] M. MARCUS AND H. MINC (1964) *A survey of Matrix theory and matrix inequalities.* Allyn and Bacon, Boston.

[10] T. ROLSKI (1987) Approximation of periodic queues. *Adv. in Appl. Probab.* **19** 691–707.

[11] T. ROLSKI (1989) Relationships between characteristics in periodic Poisson queues. *Queueing Systems Theory Appl.* **4** 17–26.

[12] J.H.A. DE SMIT (1983) The queue $GI/M/s$ with customers of different types or the queue $GI/H_m/s$. *Adv. in Appl. Probab.* **15**, 392–419.

[13] D.P HEYMAN AND M. J. SOBEL (1982) *Stochastic Models in Operations Research, Vol I.* McGraw Hill, NY.