

CBM
R

entER

for
Economic Research

8414
1993
60

Discussion paper



Center
for
Economic Research

No. 9360

**THE POWER-SERIES ALGORITHM
EXTENDED TO THE *BMAP/PH/1* QUEUE**

by W.B. van den Hout and
J.P.C. Blanc

October 1993

ISSN 0924-7815

THE POWER-SERIES ALGORITHM EXTENDED TO THE *BMAP/PH/1* QUEUE

W.B. van den Hout* and J.P.C. Blanc

Tilburg University, Faculty of Economics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

The power-series algorithm is developed for a single server queue with a Batch Markovian Arrival Process and independent general phase-type service time distribution. The service time distribution of the first service after an idle period is allowed to differ from the distribution of other service times. It is proved that the steady-state probabilities as functions of the load of the system are analytic at the origin and recursive expressions are derived to calculate the coefficients of the power-series expansions. These power series are used to study the queue length and the waiting time distribution. The present paper is a preliminary to extensions of the algorithm to multi-queue systems with non-Poissonian arrival process and service distributions.

Keywords: Single server queue, Batch Markovian Arrival Process, power-series expansion.

1. Introduction

If customers arrive one at a time at a queue and future arrivals are independent of the arrivals in the past, a Poisson process is usually a good description of the arrival process. However, these conditions may not be satisfied. Consider for example a central computer with several terminals where the offered data packets consist of several jobs and the number of active terminals varies with time. Here the Poisson process would be a very inadequate approximation. Also in the study of ATM systems, the Poisson process is considered to be unsuitable to model the bursty nature of the arrival process. A far less limited class of arrival processes is the class of Batch Markovian Arrival Processes (*BMAP*), which was introduced by Lucantoni [8] and is equivalent to the versatile Markovian point process (*N*) introduced earlier by Neuts [9]. This class of arrival processes contains many well-known special cases. Examples of *BMAPs* with maximal batch size equal to 1 are Markov-modulated Poisson processes, processes with general phase-type (*PH*) interarrival times (not necessarily independent) and overflow processes from finite Markovian queues. Also processes of which the subsequent batch sizes depend

* The investigations were supported in part by the Netherlands Foundation for Mathematics SMC with financial aid from the Netherlands Organization for the Advancement of Scientific Research (NWO).



K.U.B.
BIBLIOTHEEK
TILBURG

on each other or on the interarrival times are included in this class. A more detailed list of special cases is given by Lucantoni [8].

The power-series algorithm (PSA) is a device to compute performance measures for multi-queue systems that can be described as a continuous time Markov process. The basic idea is to transform the infinite set of non-recursively solvable balance-equations into a set of recursively solvable equations. This is done by multiplying all transitions in the arrival process by a scalar ρ . For low values of ρ the system will be relatively empty, for high values it will be full, so ρ is a measure of the load of the system and the steady-state probabilities are clearly functions of ρ . It will be shown for the *BMAP/PH/1* that they are analytical functions of ρ , at least for small values of ρ , so that they can be written as power series in the load ρ , and that the coefficients of these power series can be calculated recursively.

The PSA has been applied to queues in parallel [1], coupled processor models [7], the shortest-queue model [4] and various polling models [3]. Recently, the PSA has been extended to calculate partial derivatives of performance measures with respect to system parameters [5]. All previous models use Poisson arrival streams and exponential or Coxian service times. The aim of the present paper is to provide a better theoretical justification of the PSA and to extend the PSA to models with a Batch Markovian Arrival Process (with finite maximal batch size) and general phase-type service time distribution (possibly different for the first service after each idle period). The discussion is restricted to the single server model to keep the notation simple and to provide a basis for the analysis of multi-queue systems. The *BMAP/G/1* queue was already analysed by Lucantoni [8], using the matrix-analytic approach. However, this method seems to be unsuitable for multi-queue systems.

In section 2 the *BMAP/PH/1* model and its global balance equations will be described. In section 3 the algorithm to calculate the coefficients of the power-series expansions of the steady-state probabilities is derived and it is proved that the power-series expansions converge on a disc around the origin. In section 4 it is shown how these power series can be used to compute the queue-length distribution and moments of the waiting time distribution. In section 5 some examples are given and in the final section 6 conclusions are drawn.

2. The BMAP/PH/1 model

The behaviour of a Batch Markovian Arrival Process depends on an underlying continuous time Markov chain. Transitions in this chain may trigger batch arrivals. Let the number of states of this chain be I ($< \infty$). In state i , the transition rate is $\rho\alpha_i$ and a transition to state h occurs with probability π_{ih} ($i, h=1, \dots, I$). When such a transition is made, a batch of size m arrives with probability q_{ih}^m ($m=0, \dots, M$). The α_i are strictly positive and finite and normalized in such a way that the queueing system is stable for $0 \leq \rho < 1$. The factor ρ is used to set the time scale of the arrival process, while keeping the time scale of the service process constant. It will be called the load of the system. The steady-state probabilities will be considered as functions of this load. It is assumed that the Markov chain is irreducible, which is no restriction because only the steady-state behaviour will be studied. The maximal batch size M ($< \infty$) is assumed to be such that arrivals of batches with size M do occur: $\sum_{ih} \pi_{ih} q_{ih}^M > 0$. This again is no restriction.

The service times are mutually independent random variables and also independent of the arrival process. They have a general phase-type distribution with J ($< \infty$) phases. The transition rate in phase j is β_j and a transition to phase h occurs with probability ϕ_{jh} ($h, j=1, \dots, J$). Service is ended after phase j with probability $\phi_{j0} = 1 - \sum_{h=1, \dots, J} \phi_{jh}$. The service time is zero with probability ϕ_{00} . If the service time is positive, then the initial phase is phase j with probability ϕ_{0j} ($j=1, \dots, J$; $\sum_{j=1, \dots, J} \phi_{0j} = 1$). The rates β_j and the mean service time are assumed to be strictly positive and finite.

Type-1 customers are customers who are served after an idle period, up to and including the first customer with non-zero service time, that is all customers with zero waiting time. A type-2 customer is any other customer. The service times of both types of customers are allowed to differ in the initial distribution: $\tilde{\phi}_{00}$ and $\tilde{\phi}_{0j}$ for type-1 customers, ϕ_{00} and ϕ_{0j} for type-2 customers ($j=1, \dots, J$). That differences are not modelled with different β_j and ϕ_{jh} ($h, j=1, \dots, J$) may at first sight seem more restrictive than it actually is: any pair of phase-type distributions can be modelled by taking $\{\phi_{jh}\}_{j,h=1, \dots, J}$ block-diagonal, with the blocks corresponding to the different distributions. The queueing system is stable if the mean interarrival time is larger than the mean of the type-2 service time distribution. If the type-1 and type-2 distributions have equal means, the factor ρ is equal to the overall mean service time divided by the mean interarrival time (the usual definition of the load).

The queue-length distribution is determined under the assumption that the service discipline is non-preemptive, workconserving and service time independent. This means

that services are not interrupted, the service requirements are not affected by the service discipline, the server is idle only if the system is empty and the service order and service times are independent. Examples are first-come-first-served, last-come-first-served and service in random order. For this class of service disciplines the order of service does not influence the queue-length distribution. The waiting time distribution is studied under the assumption that the service discipline is first-come-first-served, so that the waiting time distribution can be determined by conditioning on the situation at arrivals.

Consider the continuous time Markov process $\{[N_t, I_t, J_t]; t \geq 0\}$, where N_t denotes the number of customers in the system (waiting or being served), I_t the state of the *BMAP*, and J_t the service phase, all at time $t \geq 0$. For $N_t = 0$, let J_t be the initial phase of the next customer (with non-zero service time) to be served, so that the initial phase of the service of any customer is determined right after the departure of the preceding customer. The steady-state probabilities are defined as

$$p(\rho; n, i, j) = \lim_{t \rightarrow \infty} \Pr\{N_t = n, I_t = i, J_t = j \mid N_0 = n_0, I_0 = i_0, J_0 = j_0, \text{ at load } \rho\}, \quad (2.1)$$

with $(n, i, j) \in \Omega = \mathbb{N}_0 \times \{1, \dots, I\} \times \{1, \dots, J\}$. The steady-state probabilities do not depend on the initial conditions (n_0, i_0, j_0) . The balance equations are

$$\begin{aligned} \rho \alpha_i \quad p(\rho; 0, i, j) &= \sum_{m=0}^M \sum_{h=1}^I \rho \alpha_h \pi_{hi} q_{hi}^m \bar{\phi}_{00}^m \quad p(\rho; 0, h, j) \\ &+ \sum_{\ell=0}^{\infty} \sum_{h=1}^J \beta_h \phi_{h0} \phi_{00}^{\ell} \bar{\phi}_{0j} \quad p(\rho; 1+\ell, i, h), \\ (\rho \alpha_i + \beta_j) \quad p(\rho; n, i, j) &= \sum_{m=n}^M \sum_{h=1}^I \rho \alpha_h \pi_{hi} q_{hi}^m \bar{\phi}_{00}^{m-n} (1 - \bar{\phi}_{00}) \quad p(\rho; 0, h, j) \\ &+ \sum_{m=0}^{M \wedge (n-1)} \sum_{h=1}^I \rho \alpha_h \pi_{hi} q_{hi}^m \quad p(\rho; n-m, h, j) \\ &+ \sum_{h=1}^J \beta_h \phi_{hj} \quad p(\rho; n, i, h) \\ &+ \sum_{\ell=0}^{\infty} \sum_{h=1}^J \beta_h \phi_{h0} \phi_{00}^{\ell} (1 - \phi_{00}) \phi_{0j} \quad p(\rho; n+1+\ell, i, h), \end{aligned} \quad (2.2)$$

for $n \geq 1$, $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$, and where $x \wedge y$ denotes the minimum of x and y . In the right-hand side (RHS) of the first equation, the first term corresponds to a batch arrival in an empty system. All customers in this batch have a zero service time (of type-1), so that the system is immediately empty again. The second term corresponds to a service completion, followed by enough zero service times (of type-2) to empty the system. In the RHS of the second equation, the first term vanishes if n is larger than the maximal batch size M . It corresponds to a batch arrival in an empty system, followed by a number of zero services. The second term corresponds to a batch arrival in a non-empty system. Since no new service is started, there can be no zero service times here. The last two terms correspond to changes in the service phase: the first without completion, the second with completion and followed by a number of zero service times. Together with the normalization

$$\sum_{n=0}^{\infty} \sum_{i=1}^I \sum_{j=1}^J p(\rho; n, i, j) = 1, \quad (2.3)$$

this set of equations determines the steady-state probabilities. However, since the number of states is not finite it can only be solved in special cases. Approximations can be obtained by truncating the state space. Alternatives are the matrix-geometric approach and the PSA. The latter will be described in the next section. If the buffer size of the queue is finite, the state space is also finite. Hence, solving the balance equations directly seems more natural than using the PSA. However, for large buffer sizes the PSA may still be more efficient.

3. The power-series algorithm to calculate the steady-state probabilities

In this section it is proved that the steady-state probabilities, as functions of ρ , are analytic at $\rho=0$ and recursive expressions are derived to calculate the coefficients of the power-series expansions at $\rho=0$. This is done in three steps. In Theorem 1 it is proved that the state probabilities satisfy $p(\rho; n, i, j) = O(\rho^{\lceil n \rceil_M})$ for $\rho \downarrow 0$, where

$$\lceil n \rceil_M = \text{Min} \left\{ k \in \mathbb{N}_0 \mid k \geq \frac{n}{M} \right\}, \quad \text{for } n \in \mathbb{N}_0 \quad (3.1)$$

(so $\lceil n \rceil_M$ denotes n/M rounded upward). This does not imply that the steady-state probabilities are analytic at the origin. (For example the function $F(x) = \sqrt{x}$ is $O(1)$ for $x \downarrow 0$, but it is not analytic at $x=0$.) The order found in Theorem 1 is used in

Theorem 2 to derive recursive expressions for the coefficients of the power series, basically by determining all derivatives at $\rho=0$. That these coefficients exist still does not prove that the state probabilities are analytic at the origin, since the power series may not converge. (For example, the series $\sum_{n \geq 0} n! x^n$ has finite coefficients, but it does not converge for $x \neq 0$, so it is not analytic at $x=0$.) Finally, in Theorem 3 it is proved that the power series found in Theorem 2 do converge in a disk around the origin. Of course, one would like to have convergence for all values of ρ in $[0,1)$, but in general this is not the case. Examples can be given with singularities close to the origin. To obtain convergence in these cases, a conformal mapping is used, resulting in different power series.

Theorem 1. *The steady-state probabilities of a stable BMAP/PH/1 queue satisfy*

$$p(\rho; n, i, j) = O\left(\rho^{\lceil n \rceil_M}\right), \quad \rho \downarrow 0, \quad \text{for } (n, i, j) \in \Omega. \quad (3.2)$$

Proof. Define the following subset of Ω :

$$S(n, \Gamma) = \left\{ (\ell, i, j) \in \Omega \mid (\ell \leq n) \text{ or } (\ell = n+1 \text{ and } j \in \Gamma) \right\}, \quad (3.3)$$

for $n \geq 0$, $\Gamma \subseteq \{1, \dots, J\}$. In steady state, the rates at which the process leaves and enters $S(n, \Gamma)$ are equal:

$$\begin{aligned} & \sum_{i=1}^I \sum_{j=1}^J \sum_{m=n+1+I(j \in \Gamma)}^M \sum_{h=1}^I \rho \alpha_i \pi_{ih} q_{ih}^m \left(1 - \tilde{\phi}_{00}^{m-n-I(j \in \Gamma)}\right) p(\rho; 0, i, j) \\ & + \sum_{\ell=1}^n \sum_{i=1}^I \sum_{j=1}^J \sum_{m=n-\ell+1+I(j \in \Gamma)}^M \sum_{h=1}^I \rho \alpha_i \pi_{ih} q_{ih}^m p(\rho; \ell, i, j) \\ & + \sum_{i=1}^I \sum_{j \notin \Gamma} \left[\sum_{m=1}^M \sum_{h=1}^I \rho \alpha_i \pi_{ih} q_{ih}^m + \sum_{h \notin \Gamma} \beta_j \phi_{jh} \right] p(\rho; n+1, i, j) \quad (3.4) \\ & = \sum_{i=1}^I \sum_{j \notin \Gamma} \left[\beta_j \phi_{j0} + \sum_{h \in \Gamma} \beta_j \phi_{jh} \right] p(\rho; n+1, i, j) \\ & + \sum_{\ell=1}^{\infty} \sum_{i=1}^I \sum_{j=1}^J \left[\beta_j \phi_{j0} \phi_{00}^{\ell} + \sum_{h \in \Gamma} \beta_j \phi_{j0} \phi_{00}^{\ell-1} (1 - \phi_{00}) \phi_{0h} \right] p(\rho; n+1+\ell, i, j). \end{aligned}$$

The function $I(E)$ denotes the indicator function of event E . In the first two terms, the summation over m is zero if the upper index is smaller than the lower index. The first term of the LHS corresponds to a batch arrival in an empty system, followed by a limited number of zero service times. The second term corresponds to a batch arrival in a non-empty system. The states with $n+1$ customers in the system are only in $S(n, \Gamma)$ if the service phase is in Γ . If so, then after a batch arrival or a transition to a service phase not in Γ the system will leave $S(n, \Gamma)$. If not, then after a service completion or a transition to a service phase in Γ the system will enter $S(n, \Gamma)$. Finally, if the number of customers in the system is larger than $n+1$, then a service completion followed by a sufficiently large number of zero service times will bring the system back into $S(n, \Gamma)$. A distinction has to be made between whether the new number of customers is smaller than or exactly equal to $n+1$, since in the second case the system will only enter $S(n, \Gamma)$ if the new service phase is in Γ .

Let $\Gamma_k \subseteq \{1, \dots, J\}$, $k \geq 0$, be the set of all phases from which service can be ended within k transitions:

$$\Gamma_k = \begin{cases} \emptyset, & k=0, \\ \{j \in \{1, \dots, J\} \mid \phi_{j0} > 0\}, & k=1, \\ \Gamma_{k-1} \cup \{j \in \{1, \dots, J\} \mid \sum_{h \in \Gamma_{k-1}} \phi_{jh} > 0\}, & k \geq 2. \end{cases} \quad (3.5)$$

Because the mean service time is finite, there exists a $K \leq J$ such that $\Gamma_K = \{1, \dots, J\}$.

The theorem is proved by induction over k and n . Suppose that for some $n \geq 0$ and $k \in \{0, \dots, K-1\}$

$$p(\rho; \ell, i, j) = O(\rho^{\lceil \ell \rceil_M}), \quad \rho \downarrow 0, \quad \text{for } (\ell, i, j) \in S(n, \Gamma_k). \quad (3.6)$$

Because probabilities are bounded, (3.6) is true for $n=k=0$: $p(\rho; 0, i, j) = O(1)$ for $\rho \downarrow 0$, for all $(0, i, j) \in S(0, \emptyset)$. Consider equation (3.4) for the set $S(n, \Gamma_k)$. In the second term of the LHS, the lower index of the summation over m is larger than the upper index if $\ell < n+1-M$. Because arrivals of batches with size M do occur and because of (3.6), this second term is of order $O(\rho^{1+\lceil n+1-M \rceil_M}) = O(\rho^{\lceil n+1 \rceil_M})$ for $\rho \downarrow 0$. The first and third term of the LHS are of the same or higher order. Since all coefficients in (3.4) are non-negative, this implies that all probabilities in the RHS of (3.4) with positive coefficients are also of this order, especially those in the first term:

$$p(\rho; n+1, i, j) = O(\rho^{\lceil n+1 \rceil_M}), \rho \neq 0, \quad \text{for } i \in \{1, \dots, I\}, j \in \Gamma_{k+1} \setminus \Gamma_k. \quad (3.7)$$

Hence, (3.6) is true for $(\ell, i, j) \in S(n, \Gamma_{k+1})$ and, by induction over k , also for $(\ell, i, j) \in S(n, \Gamma_K)$. The fact that $S(n, \Gamma_K) = S(n+1, \Gamma_0)$ finishes the proof of Theorem 1, by induction over n . \square

Theorem 1 shows that the order of the steady-state probability of a certain state is equal to the minimal number of transitions in the arrival process to reach that state from an empty system. For single arrivals ($M=1$), $p(\rho; n, i, j) = O(\rho^n)$ for $\rho \neq 0$, which is indeed what was used in the previous applications of the PSA. The theorem also shows that, if the power-series expansion of $p(\rho; n, i, j)$ exists, the coefficients of all powers ρ^k with $k < \lceil n \rceil_M$ are zero. Theorem 2 describes how this can be used to calculate the remaining coefficients. To formulate the balance equations in matrix notation, define the following matrices and column vectors:

$$\begin{aligned} P_n(\rho) &= \{ p(\rho; n, i, j) \}_{i=1, \dots, I, j=1, \dots, J}, & \text{for } n \geq 0, \\ A &= \text{diag}(\{ \alpha_i \}_{i=1, \dots, I}), \\ \Pi &= \{ \pi_{ih} \}_{i,h=1, \dots, I}, \\ Q_m &= \{ q_{ih}^m \}_{i,h=1, \dots, I}, & \text{for } m=0, \dots, M, \\ \Psi_m &= \{ \pi_{ih} q_{ih}^m \}_{i,h=1, \dots, I}, & \text{for } m=0, \dots, M, \\ B &= \text{diag}(\{ \beta_j \}_{j=1, \dots, J}), \\ \Phi &= \{ \phi_{jh} \}_{j,h=1, \dots, J}, \\ \phi_{\cdot 0} &= \{ \phi_{j0} \}_{j=1, \dots, J}, \\ \phi_0 &= \{ \phi_{0j} \}_{j=1, \dots, J}, \\ \tilde{\phi}_0 &= \{ \tilde{\phi}_{0j} \}_{j=1, \dots, J}. \end{aligned} \quad (3.8)$$

Let further e_H (0_H) be a column vector with all its H components equal to 1 (0), and O_{HK} an H by K matrix with all its components equal to 0. Notice that the matrices Ψ_m , $m=0, \dots, M$, are entry-wise products of Π with Q_m , and that $\Sigma_m \Psi_m = \Pi$. Notice also that $(I - \Phi) e_J = \phi_{\cdot 0}$. The set of equations (2.2) and (2.3) can be rewritten as:

$$\begin{aligned} \rho A P_0(\rho) &= \sum_{m=0}^M \rho \tilde{\phi}_{00}^m \Psi_m^T A P_0(\rho) \\ &+ \sum_{\ell=0}^{\infty} \phi_{00}^\ell P_{1+\ell}(\rho) B \phi_{\cdot 0} \tilde{\phi}_0^T, \end{aligned} \quad (3.9)$$

$$\begin{aligned}
\rho A P_n(\rho) + P_n(\rho) B = & \sum_{m=n}^M \rho \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \Psi_m^T A P_0(\rho) \\
& + \sum_{m=0}^{M \wedge (n-1)} \rho \Psi_m^T A P_{n-m}(\rho) \\
& + P_n(\rho) B \tilde{\Phi} \\
& + \sum_{\ell=0}^{\infty} \phi_{00}^{\ell} (1 - \phi_{00}) P_{n+1+\ell}(\rho) B \phi_{\cdot 0} \phi_0^T,
\end{aligned} \tag{3.10}$$

for $n \geq 1$, and

$$\sum_{n=0}^{\infty} e_J^T P_n(\rho) e_J = 1. \tag{3.11}$$

If zero service times do not occur, then most summations in (3.9) and (3.10) reduce to a single term, since in that case $\tilde{\phi}_{00}^h = \phi_{00}^h = 0$ unless $h=0$.

Summing (3.9) and (3.10) over all $n \geq 0$ and post-multiplying with a column of ones shows that

$$(I - \Pi)^T A \sum_{n=0}^{\infty} P_n(\rho) e_J = 0_J. \tag{3.12}$$

Together with (3.11) this implies that the I -vector $\sum_n P_n(\rho) e_J$ equals the steady-state distribution ν of the Markov chain underlying the $BMAP$, determined by $(I - \Pi)^T A \nu = 0_I$ and $e_J^T \nu = 1$. The solution to these equations does not depend on ρ and is unique, because Π is irreducible.

Rearranging (3.9) shows that $P_0(\rho)$ is an outer product of an I -vector with $\tilde{\phi}_0$:

$$P_0(\rho) = \frac{1}{\rho} A^{-1} \left[I - \sum_{m=0}^M \tilde{\phi}_{00}^m \Psi_m^T \right]^{-1} \left[\sum_{\ell=0}^{\infty} \phi_{00}^{\ell} P_{1+\ell}(\rho) \right] B \phi_{\cdot 0} \tilde{\phi}_0^T. \tag{3.13}$$

Post-multiplying this equality with e_J shows that the I -vector must equal $P_0(\rho) e_J$, since $\tilde{\phi}_0^T e_J = 1$. Hence,

$$P_0(\rho) = P_0(\rho) e_J \tilde{\phi}_0^T. \tag{3.14}$$

This is a consequence of the definition of J_t at times when $N_t=0$ (see section 2).

Theorem 2. *The steady-state probabilities of a stable BMAP/PH/1 queue can formally be expanded as power series in terms of the load ρ of the system:*

$$P_n(\rho) = \sum_{k=\lceil n \rceil_M}^{\infty} \rho^k U_{k,n}, \quad (3.15)$$

where the $U_{k,n}$ are I by J matrices determined by

$$U_{0,0} = \nu \tilde{\phi}_0^T, \quad (I - \Pi)^T \wedge \nu = 0_I, \quad e_J^T \nu = 1, \quad (3.16^a)$$

$$U_{k,0} = - \sum_{n=1}^{kM} U_{k,n} e_J \tilde{\phi}_0^T, \quad k \geq 1, \quad (3.16^b)$$

$$\begin{aligned} U_{k,n} B (I - \Phi) &= \sum_{m=n}^M \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \Psi_m^T A U_{k-1,0} \\ &+ (\Psi_0 - I)^T A U_{k-1,n} + \sum_{m=1}^{M \wedge (n-1)} \Psi_m^T A U_{k-1,n-m} \\ &+ \sum_{\ell=0}^{kM-n-1} \phi_{00}^{\ell} (1 - \phi_{00}) U_{k,n+1+\ell} B \phi_{00} \phi_0^T, \quad 1 \leq n \leq kM, \end{aligned} \quad (3.16^c)$$

with $U_{k,n} = 0_{IJ}$ if $kM < n$.

Proof. Define for $\rho \in [0,1)$ and $n \geq 0$:

$$\begin{aligned} R_{k,n}(\rho) &= \begin{cases} P_n(\rho), & k = \lceil n \rceil_M, \\ R_{k-1,n}(\rho) - \rho^{k-1} \hat{R}_{k-1,n}, & k > \lceil n \rceil_M, \end{cases} \\ \hat{R}_{k,n} &= \lim_{\rho \downarrow 0} \rho^{-k} R_{k,n}(\rho), \quad k \geq \lceil n \rceil_M. \end{aligned} \quad (3.17)$$

It will be shown that all the $\hat{R}_{k,n}$ exist and that they satisfy the equalities determining the corresponding $U_{k,n}$, which implies that they are identical. Notice from the definition that if $\hat{R}_{k-1,n}$ exists, then $R_{k,n}(\rho) = o(\rho^{k-1})$ for $\rho \downarrow 0$.

Substituting (3.17) into (3.11), (3.12) and (3.14) leads to

$$\begin{aligned}
e_I^T \sum_{n=0}^{\infty} R_{\lceil n \rceil_M, n}(\rho) e_J &= 1, \\
(I - \Pi)^T A \sum_{n=0}^{\infty} R_{\lceil n \rceil_M, n}(\rho) e_J &= 0_I, \\
R_{0,0}(\rho) &= R_{0,0}(\rho) e_J \tilde{\phi}_0^T.
\end{aligned} \tag{3.18}$$

By theorem 1, the $R_{\lceil n \rceil_M, n}(\rho)$ are $O(\rho^{\lceil n \rceil_M})$ for $\rho \downarrow 0$. Hence, letting $\rho \downarrow 0$ in (3.18) renders

$$e_I^T \hat{R}_{0,0} e_J = 1, \quad (I - \Pi)^T A \hat{R}_{0,0} e_J = 0_I, \quad \hat{R}_{0,0} = \hat{R}_{0,0} e_J \tilde{\phi}_0^T. \tag{3.19}$$

This shows that $\hat{R}_{0,0}$ exists and satisfies the same equalities as $U_{0,0}$ in (3.16^a).

Substituting (3.17) into (3.10) for $1 \leq n \leq M$ leads to

$$\begin{aligned}
R_{1,n}(\rho) B (I - \Phi) &= \sum_{m=n}^M \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \rho \Psi_m^T A R_{0,0}(\rho) \\
&\quad + \rho (\Psi_0 - I)^T A R_{1,n}(\rho) + \sum_{m=1}^{M \wedge (n-1)} \rho \Psi_m^T A R_{1,n-m}(\rho) \\
&\quad + \sum_{\ell=0}^{\infty} \phi_{00}^{\ell} (1 - \phi_{00}) R_{\lceil n+1+\ell \rceil_M, n+1+\ell}(\rho) B \phi_0 \phi_0^T.
\end{aligned} \tag{3.20}$$

Dividing by ρ and letting $\rho \downarrow 0$ renders

$$\begin{aligned}
\hat{R}_{1,n} B (I - \Phi) &= \sum_{m=n}^M \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \Psi_m^T A \hat{R}_{0,0} \\
&\quad + \sum_{\ell=0}^{M-n-1} \phi_{00}^{\ell} (1 - \phi_{00}) \hat{R}_{1,n+1+\ell} B \phi_0 \phi_0^T.
\end{aligned} \tag{3.21}$$

Considering $n=M$ down to $n=1$ shows that, since $\hat{R}_{0,0}$ exists, so do $\hat{R}_{1,n}$ ($1 \leq n \leq M$) and they satisfy the same equalities as the $U_{1,n}$ in (3.16^c).

Substituting (3.17) into (3.18) and using (3.19) leads to

$$\begin{aligned}
& e_I^T \left[R_{1,0}(\rho) + \sum_{n=1}^{\infty} \left\{ R_{[n]_M+1,n}(\rho) + \rho^{[n]_M} \hat{R}_{[n]_M,n} \right\} \right] e_J = 0, \\
& (I - \Pi)^T A \left[R_{1,0}(\rho) + \sum_{n=1}^{\infty} \left\{ R_{[n]_M+1,n}(\rho) + \rho^{[n]_M} \hat{R}_{[n]_M,n} \right\} \right] e_J = 0_I, \quad (3.22) \\
& R_{1,0}(\rho) = R_{1,0}(\rho) e_J \tilde{\phi}_0^T,
\end{aligned}$$

which implies

$$R_{1,0}(\rho) = - \sum_{n=1}^{\infty} \left\{ R_{[n]_M+1,n}(\rho) + \rho^{[n]_M} \hat{R}_{[n]_M,n} \right\} e_J \tilde{\phi}_0^T. \quad (3.23)$$

Dividing by ρ and letting $\rho \rightarrow 0$, now renders

$$R_{1,0}(\rho) = - \sum_{n=1}^M \bar{R}_{[n]_M,n} e_J \tilde{\phi}_0^T. \quad (3.24)$$

Hence, also $\hat{R}_{1,0}$ exists, and satisfies the same equality as $U_{1,0}$ in (3.16^b).

Next, suppose that for some $K \geq 1$ it has been shown that, for $1 \leq k \leq K$,

$$R_{k,0}(\rho) = - \sum_{n=1}^{\infty} \left\{ R_{[n]_M+k,n}(\rho) + \rho^{[n]_M+k-1} \sum_{h=0}^{k-1} \hat{R}_{[n]_M+k-1,n+hM} \right\} e_J \tilde{\phi}_0^T, \quad (3.25)$$

and that for $1 \leq n \leq kM$,

$$\begin{aligned}
& R_{k,n}(\rho) B (I - \Phi) \\
&= \sum_{m=n}^M \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \rho \Psi_m^T A R_{k-[n]_M,0}(\rho) \\
&+ \rho (\Psi_0 - I)^T A R_{k,n}(\rho) + \sum_{m=1}^{M \wedge (n-1)} \rho \Psi_m^T A R_{k-I(m \geq n-M [n]_M+M-1),n-m}(\rho) \\
&+ \rho^k (\Psi_0 - I)^T A \hat{R}_{k-1,n} + \sum_{m=1}^{n-M [n]_M+M-1} \rho^k \Psi_m^T A \hat{R}_{k-1,n-m} \quad (3.26) \\
&+ \sum_{\ell=0}^{\infty} \phi_{00}^{\ell} (1 - \phi_{00}) R_{[n+1+\ell]_M - [n]_M+k,n+1+\ell}(\rho) B \phi_{,0} \phi_0^T \\
&+ \sum_{h=1}^{k-[n]_M} \sum_{\ell=(h+[n]_M-1)M-n}^{\infty} \phi_{00}^{\ell} (1 - \phi_{00}) \rho^{[n+1+\ell]_M - [n]_M+k-h} \\
&\quad \times \hat{R}_{[n+1+\ell]_M - [n]_M+k-h,n+1+\ell} B \phi_{,0} \phi_0^T,
\end{aligned}$$

and that the $\hat{R}_{k,n}$ ($1 \leq k \leq K$) exist and satisfy the same equalities as the corresponding $U_{k,n}$ in (3.16). By (3.20) to (3.24), this is true for $K=1$. It can be shown, by substituting (3.17) and some more tedious calculations, that then (3.25) and (3.26) are also true for $k=K+1$, and that the $\hat{R}_{K+1,n}$ exist and satisfy the same equalities as the corresponding $U_{K+1,n}$ in (3.16). By induction over K , this proves theorem 2. \square

Because the mean service time is finite, $B(I-\Phi)$ is non-singular so that the matrices $U_{k,n}$ can be found from (3.16^c). Notice that, if the service time distribution is Coxian, as in previous applications of the PSA, $B(I-\Phi)$ consists of zeros, except for the main diagonal and an adjacent diagonal so that no extra work needs to be done to compute the LU-decomposition.

A more intuitive way to find the recursive relations (3.16) would be to assume beforehand that the steady-state probabilities are analytic at $\rho=0$, so that the power-series expansions (3.15) exist. Two analytical functions are only equal if all coefficients of the power-series expansions are equal. Therefore, substitution of these expansions into (3.10), (3.11), (3.12) and (3.14), and equating coefficients of corresponding powers of ρ on either side of the equality signs, also leads to (3.16).

From formula (3.16) it can be seen that each matrix $U_{k,n}$ is a function of matrices of which either the first index is smaller or of which the first index is equal and the second index is larger. Therefore, the coefficients can be recursively calculated for increasing values of k , and for each fixed k for decreasing values of n , starting with $n=kM$. For $n > kM$ the coefficients are zero. The power-series algorithm to compute all coefficients up to and including the coefficients of the K^{th} power of ρ is as follows:

Solve the set of equations $(I-\Pi)^T A \nu = 0_I$ and $e_I^T \nu = 1$;

$U_{0,0} \leftarrow \nu \tilde{\phi}_0^T$;

$k \leftarrow 1$;

while $k \leq K$ do

 Calculate $U_{k,n}$ from (3.16^c) for $n=kM, kM-1, \dots, 1$;

 Calculate $U_{k,0}$ from (3.16^b) ;

$k \leftarrow k+1$.

The memory requirements to store the coefficients approximately equals $\frac{1}{2}K^2M$ times the memory requirements of an I by J matrix of reals. However, if one is not interested in the complete queue-length distribution, but only in some characteristics of the distribution

(like moments and the probability of an empty system), the memory requirements can be significantly reduced. If the memory space of matrices that are no longer needed for the recursion is used again, the required number of matrices is only KM , which is equal to the number of considered steady-state probabilities. The number of multiplications to compute the coefficients is of the order $I^2 J^4 M^2 K^3$, but if the arrival and/or service process have some special structure, the computation time can be considerably reduced. Usually, the memory space requirements are more restrictive than the computation time requirements.

In the following theorem it is proved that the steady-state probabilities are analytic in a disk around the origin, which justifies writing the steady-state probabilities as power series in ρ . This is proved by showing that the power series found in Theorem 2 converge in a neighbourhood of $\rho=0$ and a lower bound on the radius of convergence is obtained. Up till now, analyticity at the origin was only proved for a specific coupled processor model [7], using the special $M/M/1$ structure underlying the model.

Theorem 3. *The steady-state probabilities of a stable BMAP/PH/1 queue are analytic functions of the load ρ in a disk around $\rho=0$.*

Proof. For a vector x and a (not necessarily square) matrix A , consider the following norms:

$$\begin{aligned} \|x\|_p &= \left(\sum_h |x_h|^p \right)^{\frac{1}{p}}, & 1 \leq p \leq \infty, \\ \|A\|_{p,q} &= \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_q}, & 1 \leq p, q \leq \infty. \end{aligned} \quad (3.27)$$

The norm $\|\cdot\|_{1,1}$ is the maximal absolute column sum, $\|\cdot\|_{\infty,\infty}$ the maximal absolute row sum, $\|\cdot\|_{1,\infty}$ the total absolute sum and $\|\cdot\|_{\infty,1}$ the maximal absolute value. The following inequalities hold:

$$\begin{aligned} \|A+B\|_{p,q} &\leq \|A\|_{p,q} + \|B\|_{p,q}, \\ \|AB\|_{p,q} &\leq \|A\|_{p,r} \|B\|_{r,q}, & 1 \leq r \leq \infty, \end{aligned} \quad (3.28)$$

known as the triangle inequality and consistency. Applying this to equations (3.16^b) and (3.16^c) shows that, for $k \geq 1$ and $1 \leq n \leq kM$,

$$\begin{aligned}
\|U_{k,0}\|_{\infty,1} &\leq a \sum_{n=1}^{kM} \|U_{k,n}\|_{\infty,1}, \\
\|U_{k,n}\|_{\infty,1} &\leq b \max \left\{ \max_{0 \leq m \leq M \wedge n} \|U_{k-1,n-m}\|_{\infty,1}, \right. \\
&\quad \left. \max_{0 \leq \ell \leq kM-n-1} \|U_{k,n+1+\ell}\|_{\infty,1} \right\},
\end{aligned} \tag{3.29}$$

where

$$\begin{aligned}
a &= \|e_J \tilde{\phi}_0^T\|_{1,1} = J \max_{j=1,\dots,J} \phi_{0j} \geq 1, \\
b &= \|(\Psi_0 - I)^T A\|_{\infty,\infty} \|(I - \Phi)^{-1} B^{-1}\|_{1,1} \\
&\quad + \sum_{m=1}^M \|\Psi_m^T A\|_{\infty,\infty} \|(I - \Phi)^{-1} B^{-1}\|_{1,1} \\
&\quad + \|B \phi_0 \phi_0^T (I - \Phi)^{-1} B^{-1}\|_{1,1}.
\end{aligned} \tag{3.30}$$

So if there are numbers $u_{k,n}$, such that

$$\begin{aligned}
u_{0,0} &= \|U_{0,0}\|_{\infty,1}, \\
u_{k,0} &\geq a \sum_{n=1}^{kM} u_{k,n}, \\
u_{k,n} &\geq b \max \left\{ \max_{0 \leq m \leq M \wedge n} u_{k-1,n-m}, \right. \\
&\quad \left. \max_{0 \leq \ell \leq kM-n-1} u_{k,n+1+\ell} \right\},
\end{aligned} \tag{3.31}$$

then $\|U_{k,n}\|_{\infty,1} \leq u_{k,n}$, which implies that the absolute value of each element of $U_{k,n}$ is at most $u_{k,n}$. Of course, equality in (3.31) would give better bounds, but then (3.31) would be far more difficult to solve. If $b \leq 1$, a solution is

$$u_{k,n} = \begin{cases} b^{\lceil n \rceil_M} \|U_{0,0}\|_{\infty,1}, & k = \lceil n \rceil_M, \\ c b^{\lceil n \rceil_M} (b+c)^{k-\lceil n \rceil_M-1} u_{0,0}, & k > \lceil n \rceil_M, \end{cases} \tag{3.32}$$

with $c=abM$. If $b > 1$ a solution is

$$u_{k,n} = \begin{cases} b^{(M+1)\lceil n \rceil_M - n} \|U_{0,0}\|_{\infty,1}, & k = \lceil n \rceil_M, \\ c (b+c)^{k-\lceil n \rceil_M-1} u_{\lceil n \rceil_M, n}, & k > \lceil n \rceil_M, \end{cases} \tag{3.33}$$

with $c = \max\{a(b+b^2+\dots+b^M), b^{M+1}\}$. The proof that (3.32) and (3.33) satisfy (3.31) is elementary and uses the fact that for both solutions, $u_{k,n}$ is non-increasing in n for each fixed k . In both cases the geometric series $\sum_{k \geq [n]_M} \rho^k u_{k,n}$, $n \geq 0$, converges for $|\rho| < (b+c)^{-1}$. But then also the power-series expansions of the steady-state probabilities (3.15) converge and are analytic for $|\rho| < (b+c)^{-1}$, which proves Theorem 3. \square

An upper bound on the error in $p(\rho; n, i, j)$ that is made, when only the first K coefficients are computed, is given by

$$\left| p(\rho; n, i, j) - \sum_{k=[n]_M}^K \rho^k U_{k,n,i,j} \right| \leq \sum_{k=K+1}^{\infty} \rho^k u_{k,n}. \quad (3.34)$$

Unfortunately, $(b+c)^{-1}$ is usually too small to be of any practical use. Notice that coefficients are only calculated for probabilities $P_n(\rho)$ with $n \leq KM$, and that the number of calculated coefficients decreases with n . Nevertheless, if the power-series expansions converge, any degree of accuracy can be obtained by increasing K .

There are systems for which (3.15) converges for all $0 \leq \rho < 1$. For example, for an $M/M/1$ queue, the steady-state probabilities are equal to $\rho^n(1-\rho)$. The coefficients of the power-series expansion are zero for all $k \geq n+2$, so that the analytic continuations of the steady-state probabilities are entire functions of ρ . More generally, for an $M^X/GE_J/1$ queue, it can be shown by studying the balance equations that $U_{k,n} = 0$ for all $k \geq Jn+2$. The arrival process M^X has exponential interarrival times and batch arrivals with finite maximal batch size. The GE_J service distribution is a generalized Erlang distribution, i.e. the convolution of J independent, not necessarily similar, exponential distributions.

On the other hand, there are also systems with singularities very close to the origin. For example, the steady-state probabilities of the $GE_2/M/1$ queue with $1/\alpha_1 + 1/\alpha_2 = \beta_1 = 1$ and $\alpha_1 \neq \alpha_2$ have singularities at

$$\rho = \frac{-1}{\alpha_1 \alpha_2 - 4} \left[1 \pm \frac{2}{\sqrt{\alpha_1 \alpha_2}} \right]. \quad (3.35)$$

Because $\alpha_1 \alpha_2 > 4$, these singularities are negative and can lie arbitrarily close to $\rho=0$ if $\alpha_1 \alpha_2$ is large enough, that is if one of α_1 and α_2 is large and the other is close to one. This seems to be a typical example: no singularities with positive real part were found so far, and the PSA behaves worse when the system parameters are of different orders of magnitude.

If the radius of convergence is smaller than one, the following bilinear conformal

mapping can be used [1]:

$$\theta = \frac{(1+G)\rho}{1+G\rho}, \quad \rho = \frac{\theta}{1+G(1-\theta)}, \quad G \geq 0. \quad (3.36)$$

This transformation maps the interval $[0,1]$ onto itself, and the disk in \mathbb{C} with centre $\hat{\rho} = G(1+2G)^{-1}$ and radius $1-\hat{\rho}$ onto the unit disk. The steady-state probabilities can be expanded as power series in terms of θ :

$$\tilde{P}_n(\theta) = P_n\left(\frac{\theta}{1+G(1-\theta)}\right) = \theta^{[n]_M} \sum_{k=0}^{\infty} \theta^k V_{k,n}, \quad (3.37)$$

and the coefficients are now determined by

$$V_{0,0} = \nu \tilde{\phi}_0^T, \quad (I - \Pi)^T A \nu = 0_I, \quad e_I^T \nu = 1, \quad (3.38^a)$$

$$V_{k,0} = - \sum_{n=1}^{kM} V_{k,n} e_J \tilde{\phi}_0^T, \quad k \geq 1, \quad (3.38^b)$$

$$\begin{aligned} & \left\{ (1+G) V_{k,n} - G V_{k-1,n} \right\} B (I - \Phi) \\ &= \sum_{m=n}^M \tilde{\phi}_{00}^{m-n} (1 - \tilde{\phi}_{00}) \Psi_m^T A V_{k-1,0} \\ &+ (\Psi_0 - I)^T A V_{k-1,n} + \sum_{m=1}^{M \wedge (n-1)} \Psi_m^T A V_{k-1,n-m} \\ &+ (1+G) \sum_{\ell=0}^{kM-n-1} \phi_{00}^{\ell} (1 - \phi_{00}) V_{k,n+1+\ell} B \phi_{,0} \phi_0^T \\ &- G \sum_{\ell=0}^{(k-1)M-n-1} \phi_{00}^{\ell} (1 - \phi_{00}) V_{k-1,n+1+\ell} B \phi_{,0} \phi_0^T, \quad 1 \leq n \leq kM, \end{aligned} \quad (3.38^c)$$

with $V_{k,n} = 0_{IJ}$ if $kM < n$.

Lower bounds on the radius of convergence are similar to those in Theorem 3, with b replaced by

$$\begin{aligned} b_G &= \frac{1}{1+G} \|(\Psi_0 - I)^T A\|_{\infty, \infty} \|(I - \Phi)^{-1} B^{-1}\|_{1,1} \\ &+ \frac{1}{1+G} \sum_{m=1}^M \|\Psi_m^T A\|_{\infty, \infty} \|(I - \Phi)^{-1} B^{-1}\|_{1,1} \\ &+ \frac{G}{1+G} + \frac{1+2G}{1+G} \|B \phi_{,0} \phi_0^T (I - \Phi)^{-1} B^{-1}\|_{1,1}. \end{aligned} \quad (3.39)$$

The corresponding lower bound on the radius of convergence is usually still to small. Nevertheless, the mapping does serve its purpose. For $G \rightarrow \infty$, the mapping (3.36) maps the disk $|\rho - 1/2| \leq 1/2$ onto the unit disk. So far, only singularities with negative real part were found. If this is generally true, then the analyticity at the origin of the steady-state probabilities ensures that convergence can always be obtained by choosing G large enough, since all singularities can be mapped outside the unit disk, while keeping the unit interval inside the unit disk. Unfortunately, convergence is usually slow for large G .

To store all coefficients when using the algorithm with mapping, the memory requirements are still about $1/2 K^2 M$ times the memory requirements of an I by J matrix of reals. If the memory space of matrices that are no longer needed for the recursion is used again, the required number of matrices is now about $2KM$, about doubled as compared to computation without the mapping. The number of multiplications to compute the coefficients is also roughly doubled.

4. The queue-length and waiting time distribution

When the coefficients of the steady-state probabilities have been calculated, the probability $p_n(\rho)$ of n customers in the system can be approximated in the obvious way:

$$p_n(\rho) = e_I^T P_n(\rho) e_J \approx \sum_{k=\lceil n \rceil_M}^K \rho^k e_I^T U_{k,n} e_J, \quad n \geq 0. \quad (4.1)$$

The epsilon algorithm can be used to accelerate the convergence of these series. For a description of this algorithm see Wynn [11] or Blanc [2,3]. If the conformal mapping is used, then ρ and U should be replaced by θ and V . The same is true for all other formulas in this section.

After the $p_n(\rho)$ have been computed, the ℓ^{th} moment of number of customers in the system, $L_\ell(\rho)$, can easily be calculated from them. However, to accelerate the convergence, it is better to compute the coefficients of the power-series expansions of these moments:

$$L_\ell(\rho) = \sum_{n=1}^{\infty} n^\ell p_n(\rho) = \sum_{n=1}^{\infty} n^\ell \sum_{k=\lceil n \rceil_M}^{\infty} \rho^k e_I^T U_{k,n} e_J = \sum_{k=1}^{\infty} \rho^k d_{\ell,k}, \quad (4.2)$$

with

$$d_{\ell,k} = \sum_{n=1}^{kM} n^{\ell} \mathbf{e}_I^T \mathbf{U}_{k,n} \mathbf{e}_J. \quad (4.3)$$

Since $L_{\ell}(\rho)$ has a pole of order ℓ at $\rho=1$, the power series will converge slowly for heavy traffic, but the rate of convergence can be strongly improved by using the fact that the series $\{d_{\ell,k}\}_{k \geq 0}$ will tend to a polynomial in k of order $\ell-1$ as $k \rightarrow \infty$. For $\ell=1$ and $\ell=2$, this means that there are constants a , b and c such that $\lim_{k \rightarrow \infty} (d_{1,k} - a) = 0$ and $\lim_{k \rightarrow \infty} (d_{2,k} - b - ck) = 0$, which leads to the extrapolations

$$\begin{aligned} L_1(\rho) &\approx \sum_{k=1}^K \rho^k d_{1,k} + \sum_{k=K+1}^{\infty} \rho^k d_{1,K} \\ &= \sum_{k=1}^K \rho^k d_{1,k} + d_{1,K} \frac{\rho^{K+1}}{1-\rho}, \end{aligned} \quad (4.4)$$

$$\begin{aligned} L_2(\rho) &\approx \sum_{k=1}^K \rho^k d_{2,k} + \sum_{k=K+1}^{\infty} \rho^k \left[d_{2,K} + (d_{2,K} - d_{2,K-1})(k-K) \right] \\ &= \sum_{k=1}^K \rho^k d_{2,k} + \left[d_{2,K} + \frac{d_{2,K} - d_{2,K-1}}{1-\rho} \right] \frac{\rho^{K+1}}{1-\rho}. \end{aligned} \quad (4.5)$$

For higher order moments similar extrapolations can be used. From the generating function of the queue-length distribution of an $M/G/1$ queue [6, page 238], it can be shown that these approximations of $L_{\ell}(\rho)$ are exact for $M/PH/1$ systems if $K \geq \ell + 1$, all the same whether the approximations of the steady-state probabilities are good or not. This underlines the advantage of evaluating the power-series expansions of the moments instead of calculating the moments from the steady-state probabilities.

Characteristics of the waiting time distribution are found by conditioning on the state of the system at arrival moments. Define

$$\Lambda_{n,m}(\rho) = \rho \Psi_m^T \mathbf{A} \mathbf{P}_n(\rho). \quad (4.6)$$

Then $\Lambda_{n,m}(\rho)$ is an I by J matrix whose $(i,j)^{\text{th}}$ element equals the mean number of transitions per unit time, caused by arrivals of batches of size m which result in a transition to state $(n+m, i, j)$. The mean customer arrival rate equals

$$\lambda(\rho) = \sum_{n=0}^{\infty} \sum_{m=1}^M m \mathbf{e}_I^T \Lambda_{n,m}(\rho) \mathbf{e}_J = \rho \mathbf{e}_I^T \sum_{m=1}^M m \Psi_m^T \mathbf{A} \boldsymbol{\nu}, \quad (4.7)$$

where $\boldsymbol{\nu}$ is again the steady-state distribution of the Markov chain underlying the *BMAP*.

Clearly $\lambda(\rho)$ is linear in ρ . The probability that an arbitrary customer is served without delay, is equal to the arrival rate of type-1 customers divided by the total arrival rate, which can be found by conditioning on the batch size:

$$\begin{aligned} \Pr\{\text{No Delay}\} &= \frac{1}{\lambda(\rho)} \sum_{m=1}^M \frac{1-\tilde{\phi}_{00}^m}{1-\tilde{\phi}_{00}} e_I^T \Lambda_{0,m}(\rho) e_J \\ &\approx \frac{\rho}{\lambda(\rho)} \sum_{k=0}^K \rho^k \sum_{m=1}^M \frac{1-\tilde{\phi}_{00}^m}{1-\tilde{\phi}_{00}} e_I^T \Psi_m^T A U_{k,0} e_J. \end{aligned} \quad (4.8)$$

Moments of the waiting time distribution can be calculated under the assumption that the service discipline is FCFS. Let $\mu_{\ell,n,m}$ be the J -vector whose j^{th} element equals the ℓ^{th} moment of the waiting time of a customer, conditioned on the fact that the customer arrives in a batch of size m while just before arrival there were n customers in the system and the service process was in phase j . Then $W_\ell(\rho)$, the ℓ^{th} moment of the waiting time distribution, is given by

$$\begin{aligned} W_\ell(\rho) &= \frac{1}{\lambda(\rho)} \sum_{n=0}^{\infty} \sum_{m=1}^M m e_I^T \Lambda_{nm}(\rho) \mu_{\ell,n,m} \\ &\approx \frac{\rho}{\lambda(\rho)} \sum_{k=0}^K \rho^k \sum_{n=0}^{kM} \sum_{m=1}^M m e_I^T \Psi_m^T A U_{k,n} \mu_{\ell,n,m}. \end{aligned} \quad (4.9)$$

If the maximal batch size M is one, the first coefficient of $W_\ell(\rho)$ is zero because customers who arrive in an empty system in a batch of size 1 have zero waiting time ($\mu_{\ell,0,1}=0_J$). The use of extrapolations like in (4.4) and (4.5) and the use of the epsilon algorithm again strongly accelerates the convergence. As before, the approximations of moments of the waiting time are exact for $M/PH/1$ systems if $K \geq \ell + 1$ [6, page 256].

To calculate $\mu_{\ell,n,m}$, first define μ_ℓ as the J -vector of which the j^{th} element equals the ℓ^{th} moment of the residual service time, conditioned on the fact that service is in phase j . Since the type-1 and the type-2 service time distributions differ only in the initial distribution, μ_ℓ is valid for both. Define the scalar $\hat{\mu}_\ell$ as the ℓ^{th} moment of a complete type-2 service time. Conform Neuts [10, page 46], μ_ℓ and $\hat{\mu}_\ell$ satisfy

$$\mu_\ell = \ell! [B(I-\Phi)]^{-\ell} e_J, \quad \hat{\mu}_\ell = (1-\phi_{00}) \phi_0^T \mu_\ell, \quad \ell \geq 1. \quad (4.10)$$

Now, the $\mu_{\ell,n,m}$ can be calculated by conditioning on the place in the batch and, for $n=0$, on the number of type-1 zero service times:

$$\begin{aligned}
\mu_{1,n,m} &= \begin{cases} \frac{1}{m} \sum_{h=1}^m \sum_{\ell=0}^{h-2} \tilde{\phi}_{00}^{\ell} (1 - \tilde{\phi}_{00}) [\mu_1 + (h - \ell - 2) \hat{\mu}_1 e_J] , & n=0 , \\ \frac{1}{m} \sum_{h=1}^m [\mu_1 + (n + h - 2) \hat{\mu}_1 e_J] , & n \geq 1 , \end{cases} \\
\mu_{2,n,m} &= \begin{cases} \frac{1}{m} \sum_{h=1}^m \sum_{\ell=0}^{h-2} \tilde{\phi}_{00}^{\ell} (1 - \tilde{\phi}_{00}) [\mu_2 + (h - \ell - 2) \{ \hat{\mu}_2 e_J + 2 \hat{\mu}_1 \mu_1 + (h - \ell - 3) \hat{\mu}_1^2 e_J \}] , & n=0 , \\ \frac{1}{m} \sum_{h=1}^m [\mu_2 + (n + h - 2) \{ \hat{\mu}_2 e_J + 2 \hat{\mu}_1 \mu_1 + (n + h - 3) \hat{\mu}_1^2 e_J \}] , & n \geq 1 . \end{cases} \quad (4.11)
\end{aligned}$$

Moments of the waiting times of the first or last customer in a batch (instead of an arbitrary customer), and of sojourn times can be found by using appropriate definitions of $\mu_{\ell,n,m}$. If type-1 and type-2 services have equal means, the mean waiting time can also be calculated from the mean queue length with Little's formula. If not, to apply Little's formula, the probability of no delay is needed to calculate the overall mean service time, which is not known in advance.

5. Examples

In this section some numerical examples are given. These will concern the $H_2/H_2/1$, $H_2^X/H_2/1$ and $MMPP_2/H_2/1$ models. The parameters are chosen such that the mean interarrival time and mean service time are equal to 1 (for $\rho=1$). The hyperexponential distributions have variance 2 and balanced means. The probability that in the H_2^X arrival process the batch size equals m is 0.25 for $m=1, \dots, 4$. In the $MMPP_2$ arrival process, the interarrival times also have variance 2, the steady-state distribution of the underlying Markov chain is $[0.5, 0.5]$ and two subsequent interarrival times have correlation coefficient 0.125 (if c^2 is the variance divided by the squared mean, the correlation coefficient is at most $\frac{1}{2}(1-c^{-2})$, which is 0.25 in this case). For these models the expectation and variance of the number of customers in the system are calculated for different values of K , the number of calculated coefficients of these moments. In all cases

the mapping (3.36) was used, where G was chosen such that the maximal coefficient in absolute value was not too large. The epsilon algorithm was used, in a similar way as in Blanc [3]. The results are given for $\rho=0.7$ in table 1 and for $\rho=0.9$ in table 2. A dot indicates that the value rounded to the first 4 digits is the same as the value above it. In table 3 the results are given for the same models, but with the variance of the interarrival times and the service times equal to 4 and the correlation coefficient equal to 0.15.

Comparing table 1 and table 2, it can be concluded that the algorithm converges faster for smaller values of ρ , which is no surprise since it calculates the power-series expansions around $\rho=0$. The difference between the model in table 2 and the model in table 3 is that the system parameters in the latter models have a wider range, which results in slower convergence. Perhaps this could be avoided by some kind of scaling, but this has not been thoroughly investigated yet.

K	$H_2/H_2/1$ ($G=2$)		$H_2^X/H_2/1$ ($G=2$)		$MMPP_2/H_2/1$ ($G=3$)	
	E	V	E	V	E	V
5	4.014	28.02	7.701	96.57	3.452	63.56
10	3.897	27.68	7.448	95.08	4.922	43.31
15	3.985	27.71	7.445	95.05	4.879	43.67
20	4.878	43.59
30

Table 1. Expectation and variance of the number of customers in the system, for $\rho=0.7$

K	$H_2/H_2/1$ ($G=2$)		$H_2^X/H_2/1$ ($G=2$)		$MMPP_2/H_2/1$ ($G=3$)	
	E	V	E	V	E	V
5	17.45	365.4	34.95	1199	.3501	1265
10	17.19	352.1	32.23	1220	22.23	536.4
15	17.14	355.2	32.13	1232	21.37	570.6
20	.	.	32.12	.	21.34	559.2
30	21.36	562.8
40	560.2
50	560.3
75

Table 2. Expectation and variance of the number of customers in the system, for $\rho=0.9$

K	$H_2/H_2/1 \quad (G=6)$		$H_2^X/H_2/1 \quad (G=2)$		$MMPP_2/H_2/1 \quad (G=7)$	
	E	V	E	V	E	V
5	34.76	1349	76.54	3905	56.76	1252
10	33.36	1441	63.88	3714	42.67	2138
15	33.49	1418	61.17	4384	41.75	2237
20	33.48	1411	59.46	4636	41.93	2220
30	.	1413	60.61	4494	42.02	2193
40	.	.	60.61	4498	41.88	2175
50	.	.	60.56	4504	41.90	2145
75	.	.	.	4505	.	2208
100	2209
125	2208
150

Table 3. Expectation and variance of the number of customers in the system, for higher variance and correlation and $\rho=0.9$

6. Conclusions

The power-series algorithm has been extended to the single server queue with Batch Markovian Arrival Process (with finite maximal batch size) and independent general phase-type service time distributions (possibly different for the first service after each idle period). It has been shown that the steady-state probabilities are analytic at $\rho=0$ and a lower bound on the radius of convergence of the power-series expansions has been given, which has improved the theoretical justification of the PSA. The extension to non-Poissonian arrival processes and non-Coxian service time distributions can be achieved by changing the order of computation and a limited number of matrix inversions.

For the single server queue, comparing the PSA with the matrix-geometric approach, the latter seems preferable. With the latter, the service time distribution need not be phase-type, the required memory space is much smaller and the method is more stable. On the other hand, the main advantage of the PSA is its flexibility, illustrated by the wide range of models it has been applied to. The analysis of more general arrival and service processes in this paper can be extended to multi-queue systems, like fork-join models, networks of queues and polling models. Because of the high dimensionality, the number of

queues and the sizes of the supplementary spaces of the arrival and service processes has to be limited, but for moderately sized systems the PSA is applicable, which will be the subject of future research.

References

- [1] Blanc, J.P.C. On a numerical method for calculating state probabilities for queueing systems with more than one waiting line, *J. Comput. Appl. Math.* **20** (1987), 119-125.
- [2] Blanc, J.P.C. A numerical approach to cyclic-service queueing models, *Queueing Systems* **6** (1990), 173-188.
- [3] Blanc, J.P.C. Performance evaluation of polling systems by means of the power-series algorithm, *Annals Oper. Res.* **35** (1992), 155-186.
- [4] Blanc, J.P.C. The power-series algorithm applied to the shortest-queue model, *Operat. Res.* **40** (1992), 157-167.
- [5] Blanc, J.P.C., R.D. van der Mei. Optimization of polling systems by means of gradient methods and the power-series algorithm, *Report FEW 575, Department of Economics, Tilburg University, Tilburg* (1992).
- [6] Cohen, J.W. *The Single Server Queue*, 2nd ed. North-Holland, Amsterdam, 1982.
- [7] Hooghiemstra, G., M. Keane, S. van de Ree. Power series for stationary distributions of coupled processor models, *SIAM J. Appl. Math.* **48** (1988), 1159-1166.
- [8] Lucantoni, D.M. New results on the single server queue with a batch Markovian arrival process, *Commun. Statist.-Stochastic Models* **7** (1991), 1-46.
- [9] Neuts, M.F. A versatile Markovian point process, *J. Appl. Prob.* **16** (1979), 764-769.
- [10] Neuts, M.F. *Matrix Geometric Solutions in Stochastic Models: an algorithmic approach*. John Hopkins Univ. Press, Baltimore, 1981.
- [11] Wynn, P. On the convergence and stability of the epsilon algorithm, *SIAM J. Numer. Anal.* **3** (1966), 91-122.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9221	S. Smulders and Th. van de Klundert	Monopolistic Competition, Product Variety and Growth: Chamberlin vs. Schumpeter
9222	H. Bester and E. Petrakis	Price Competition and Advertising in Oligopoly
9223	A. van den Nouweland, M. Maschler and S. Tijs	Monotonic Games are Spanning Network Games
9224	H. Suehiro	A "Mistaken Theories" Refinement
9225	H. Suehiro	Robust Selection of Equilibria
9226	D. Friedman	Economically Applicable Evolutionary Games
9227	E. Bomhoff	Four Econometric Fashions and the Kalman Filter Alternative - A Simulation Study
9228	P. Borm, G.-J. Otten and H. Peters	Core Implementation in Modified Strong and Coalition Proof Nash Equilibria
9229	H.G. Bloemen and A. Kapteyn	The Joint Estimation of a Non-Linear Labour Supply Function and a Wage Equation Using Simulated Response Probabilities
9230	R. Beetsma and F. van der Ploeg	Does Inequality Cause Inflation? - The Political Economy of Inflation, Taxation and Government Debt
9231	G. Almekinders and S. Eijffinger	Daily Bundesbank and Federal Reserve Interventions - Do they Affect the Level and Unexpected Volatility of the DM/\$-Rate?
9232	F. Vella and M. Verbeek	Estimating the Impact of Endogenous Union Choice on Wages Using Panel Data
9233	P. de Bijl and S. Goyal	Technological Change in Markets with Network Externalities
9234	J. Angrist and G. Imbens	Average Causal Response with Variable Treatment Intensity
9235	L. Meijdam, M. van de Ven and H. Verbon	Strategic Decision Making and the Dynamics of Government Debt
9236	H. Houba and A. de Zeeuw	Strategic Bargaining for the Control of a Dynamic System in State-Space Form
9237	A. Cameron and P. Trivedi	Tests of Independence in Parametric Models: With Applications and Illustrations

No.	Author(s)	Title
9238	J.-S. Pischke	Individual Income, Incomplete Information, and Aggregate Consumption
9239	H. Bloemen	A Model of Labour Supply with Job Offer Restrictions
9240	F. Drost and Th. Nijman	Temporal Aggregation of GARCH Processes
9241	R. Gilles, P. Ruys and J. Shou	Coalition Formation in Large Network Economies
9242	P. Kort	The Effects of Marketable Pollution Permits on the Firm's Optimal Investment Policies
9243	A.L. Bovenberg and F. van der Ploeg	Environmental Policy, Public Finance and the Labour Market in a Second-Best World
9244	W.G. Gale and J.K. Scholz	IRAs and Household Saving
9245	A. Bera and P. Ng	Robust Tests for Heteroskedasticity and Autocorrelation Using Score Function
9246	R.T. Baillie, C.F. Chung and M.A. Tieslau	The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis
9247	M.A. Tieslau, P. Schmidt and R.T. Baillie	A Generalized Method of Moments Estimator for Long-Memory Processes
9248	K. Wärneryd	Partisanship as Information
9249	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9250	H.G. Bloemen	Job Search Theory, Labour Supply and Unemployment Duration
9251	S. Eijffinger and E. Schaling	Central Bank Independence: Searching for the Philosophers' Stone
9252	A.L. Bovenberg and R.A. de Mooij	Environmental Taxation and Labor-Market Distortions
9253	A. Lusardi	Permanent Income, Current Income and Consumption: Evidence from Panel Data
9254	R. Beetsma	Imperfect Credibility of the Band and Risk Premia in the European Monetary System
9301	N. Kahana and S. Nitzan	Credibility and Duration of Political Contests and the Extent of Rent Dissipation
9302	W. Güth and S. Nitzan	Are Moral Objections to Free Riding Evolutionarily Stable?

No.	Author(s)	Title
9303	D. Karotkin and S. Nitzan	Some Peculiarities of Group Decision Making in Teams
9304	A. Lusardi	Euler Equations in Micro Data: Merging Data from Two Samples
9305	W. Güth	A Simple Justification of Quantity Competition and the Cournot-Oligopoly Solution
9306	B. Peleg and S. Tijs	The Consistency Principle For Games in Strategic Form
9307	G. Imbens and A. Lancaster	Case Control Studies with Contaminated Controls
9308	T. Ellingsen and K. Wärneryd	Foreign Direct Investment and the Political Economy of Protection
9309	H. Bester	Price Commitment in Search Markets
9310	T. Callan and A. van Soest	Female Labour Supply in Farm Households: Farm and Off-Farm Participation
9311	M. Pradhan and A. van Soest	Formal and Informal Sector Employment in Urban Areas of Bolivia
9312	Th. Nijman and E. Sentana	Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes
9313	K. Wärneryd	Communication, Complexity, and Evolutionary Stability
9314	O.P. Attanasio and M. Browning	Consumption over the Life Cycle and over the Business Cycle
9315	F. C. Drost and B. J. M. Werker	A Note on Robinson's Test of Independence
9316	H. Hamers, P. Borm and S. Tijs	On Games Corresponding to Sequencing Situations with Ready Times
9317	W. Güth	On Ultimatum Bargaining Experiments - A Personal Review
9318	M.J.G. van Eijs	On the Determination of the Control Parameters of the Optimal Can-order Policy
9319	S. Hurkens	Multi-sided Pre-play Communication by Burning Money
9320	J.J.G. Lemmen and S.C.W. Eijffinger	The Quantity Approach to Financial Integration: The Feldstein-Horioka Criterion Revisited
9321	A.L. Bovenberg and S. Smulders	Environmental Quality and Pollution-saving Technological Change in a Two-sector Endogenous Growth Model

No.	Author(s)	Title
9322	K.-E. Wärneryd	The Will to Save Money: an Essay on Economic Psychology
9323	D. Talman, Y. Yamamoto and Z. Yang	The $(2^{n+m+1} - 2)$ -Ray Algorithm: A New Variable Dimension Simplicial Algorithm For Computing Economic Equilibria on $S^n \times R^m$
9324	H. Huizinga	The Financing and Taxation of U.S. Direct Investment Abroad
9325	S.C.W. Eijffinger and E. Schaling	Central Bank Independence: Theory and Evidence
9326	T.C. To	Infant Industry Protection with Learning-by-Doing
9327	J.P.J.F. Scheepens	Bankruptcy Litigation and Optimal Debt Contracts
9328	T.C. To	Tariffs, Rent Extraction and Manipulation of Competition
9329	F. de Jong, T. Nijman and A. Röell	A Comparison of the Cost of Trading French Shares on the Paris Bourse and on SEAQ International
9330	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9331	H. Huizinga	Time Preference and International Tax Competition
9332	V. Feltkamp, A. Koster, A. van den Nouweland, P. Borm and S. Tijs	Linear Production with Transport of Products, Resources and Technology
9333	B. Lauterbach and U. Ben-Zion	Panic Behavior and the Performance of Circuit Breakers: Empirical Evidence
9334	B. Melenberg and A. van Soest	Semi-parametric Estimation of the Sample Selection Model
9335	A.L. Bovenberg and F. van der Ploeg	Green Policies and Public Finance in a Small Open Economy
9336	E. Schaling	On the Economic Independence of the Central Bank and the Persistence of Inflation
9337	G.-J. Otten	Characterizations of a Game Theoretical Cost Allocation Method
9338	M. Gradstein	Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning
9339	W. Güth and H. Kliemt	Competition or Co-operation
9340	T.C. To	Export Subsidies and Oligopoly with Switching Costs
9341	A. Demirgüç-Kunt and H. Huizinga	Barriers to Portfolio Investments in Emerging Stock Markets

No.	Author(s)	Title
9342	G.J. Almekinders	Theories on the Scope for Foreign Exchange Market Intervention
9343	E.R. van Dam and W.H. Haemers	Eigenvalues and the Diameter of Graphs
9344	H. Carlsson and S. Dasgupta	Noise-Proof Equilibria in Signaling Games
9345	F. van der Ploeg and A.L. Bovenberg	Environmental Policy, Public Goods and the Marginal Cost of Public Funds
9346	J.P.C. Blanc and R.D. van der Mei	The Power-series Algorithm Applied to Polling Systems with a Dormant Server
9347	J.P.C. Blanc	Performance Analysis and Optimization with the Power-series Algorithm
9348	R.M.W.J. Beetsma and F. van der Ploeg	Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions
9349	A. Simonovits	Intercohort Heterogeneity and Optimal Social Insurance Systems
9350	R.C. Douven and J.C. Engwerda	Is There Room for Convergence in the E.C.?
9351	F. Vella and M. Verbeek	Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages
9352	C. Meghir and G. Weber	Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel
9353	V. Feltkamp	Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values
9354	R.J. de Groof and M.A. van Tuijl	Aspects of Goods Market Integration. A Two-Country-Two-Sector Analysis
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the <i>BMAP/PH/1</i> Queue

Bibliotheek K. U. Brabant



17 000 01133571 9