SRC TR 87-167

Matrix-Geometric Solution for Finite Capacity Queues with Phase-Type Distributions

by

L. Gun and A. M. Makowski

To appear in the *Proceedings of Performance 87* Brussels, Belgium, December 1987

MATRIX-GEOMETRIC SOLUTION FOR FINITE CAPACITY QUEUES WITH PHASE-TYPE DISTRIBUTIONS

by

Levent Gün* and Armand M. Makowski**

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, Maryland 20742.

ABSTRACT

This paper presents a class of Quasi-Birth-and-Death processes with *finite* state space for which the invariant probability vector is found to admit a *matrix-geometric* representation. The corresponding rate matrix is given *explicitly* in terms of the model parameters, and the resulting closed-form expression is proposed as a basis for *efficient* calculation of the invariant probability vector. The framework presented in this paper provides a unified approach to the study of several well-known queueing systems.

* The work of this author was supported partially by the Office of Naval Research through Grant N00014-84-K-0614 and partially through a grant from AT&T Bell Laboratories.

** The work of this author was supported partially through NSF Grant ECS-83-51836 and partially through Grants from AT&T Bell Laboratories and GM Laboratories.

1. INTRODUCTION

As already apparent from the monograph by Neuts [9], Quasi-Birth-and-Death (QBD) processes naturally arise in the modelling of a wide variety of applications. In addition to the modelling flexibility they provide, QBD processes enjoy interesting structural properties which can be used to advantage in the computations. Indeed, under fairly general assumptions, the stationary probability vector $\pi = (\pi_0, \pi_1, \ldots)$ of a QBD process with *countably infinite* state space exhibits the *matrix-geometric* property [9], i.e., there exists a matrix R such that

$$\pi_{k+1} = \pi_k R, \quad k \geq 0 \tag{1.1}$$

where R is the minimal nonnegative solution of a matrix quadratic equation.

For QBD processes with *finite* state space, the situation is quite different owing to the presence of *boundary* states, and it is not possible *in general* to assert that the invariant distribution exhibits a matrix-geometric structure in the form (1.1). However, for arbitrary QBD processes with finite state space, Hajek [7] showed that the invariant probability vector can be written as a "sum of two matrix-geometric terms plus a linear term". The corresponding computations involve solving two matrix quadratic equations and then finding an invariant probability distribution on the boundary states.

It is the purpose of this paper to show that more precise results can be obtained for a class of finite state QBD processes which arise in modelling finite capacity queues with phase-type servers. In that case, the basic method – presented in Section 2 – yields a closed-form expression for the invariant probability vector π such that (1.1) holds on the *non*-boundary states for some matrix R. An *explicit* expression is available for this matrix R in terms of the model parameters, and no matrix quadratic equation needs to be solved as is generally the case in the algorithmic approaches suggested by Neuts and Hajek. The closed-form expressions obtained here are proposed as a basis for efficient computation of the invariant probability vector.

This matrix R is obtained through purely *algebraic* manipulations; it has no *probabilistic* interpretation and therefore does not coincide in general with the rate matrix introduced by Neuts. The rate matrix R introduced here is *not* always positive and this could lead to some numerical instabilities. The matrix-geometric form of π on the non-boundary states can be exploited to yield several *necessary* conditions satisfied by the component π_1, none of them being of the right rank in all generality. However, for the specific models discussed in Section 3, the vector π can be expressed in terms of the component π_0, and while no proof is available, extensive numerical evidence suggests that the necessary conditions thus obtained yield a *unique* vector π_0.

Although the discussion is given specifically for continuous-time QBD processes, the ideas apply *mutatis mutandis* in the discrete-time set-up to obtain closed-form matrix-geometric expressions for the steady-state probabilities of the system states. Moreover, the adopted framework is broad enough to allow for a unified approach to the study of several well-known queueing systems with phase-type (PH) distributions. In fact, the ideas of Section 2 are illustrated in Section 3 on the following continuous-time systems

1. The two node tandem system with finite buffers and feedback, under the assumption that the first node server is always busy. At each node PH-type servers, possibly subject to failures with PH-type repair distributions, are in attendance [6].
2. Two node closed queueing system with blocking and feedback [1].
3. The $PH/PH/1/K$ queue with feedback and arrival rejection [3, 5].
4. Queues with paired customers and arrival rejection [8], [9, pp. 300-320].

In each case, the necessary conditions are shown to hold so as to apply the solution technique of Section 2.

In both Sections 2 and 3, it is indicated how the structural results of this paper can be exploited to algorithmically solve for the invariant probability distribution of the QBD processes of interest.

A word on the notation used hereafter: The $r \times r$ identity matrix is denoted by I_r and the $r \times 1$ column vector of ones is denoted by e_r, while the $r \times r$ matrix and the $1 \times r$ dimensional row vector with zero entries are denoted by $0_{r\times r}$ and 0_r, respectively. The notation $\bar{x}$ is used to denote $1 - x$ for $0 \leq x \leq 1$.

2. THE MODEL AND ITS SOLUTION

2.1. The model

Consider a QBD process with finite state space given by

$$E = \begin{cases} (0,i)\ , & \text{if } k = 0 \text{ and } 1 \leq i \leq s, \\ (k,i)\ , & \text{if } 1 \leq k < K \text{ and } 1 \leq i \leq r\,, \\ (K,i)\ , & \text{if } k = K \text{ and } 1 \leq i \leq p, \end{cases}$$

and assume its generator matrix to be of the form

$$T = \begin{pmatrix} B_1 & B_0 & & & & & \\ B_2 & A_1 & A_0 & & & & \\ & A_2 & A_1 & A_0 & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & A_2 & A_1 & C_0 \\ & & & & & C_2 & C_1 \end{pmatrix} . \tag{2.1}$$

The block entries B_0, B_1, and B_2 have dimensions $s \times r$, $s \times s$ and $r \times s$, respectively, the matrices A_0, A_1 and A_2 are all of dimensions $r \times r$, while the matrices C_0, C_1 and C_2 have dimensions $r \times p$, $p \times p$ and $p \times r$, respectively. If the underlying Markov chain with generator matrix T is irreducible, then the matrices along the diagonal can be shown to be nonsingular [9, p. 13]. Here, only the *nonsingularity* of the matrices B_1 and C_1 is assumed and *no* irreducibility assumption is made on the matrix T.

Any invariant probability vector π for T is partitioned as $\pi = (\pi_0, \pi_1, \ldots, \pi_K)$, where the row vectors π_0, π_k, $1 \leq k < K$, and π_K are of dimension s, r and p, respectively. Pose $|E| = (K-1)r + s + p$, and observe that the equation

$$\pi T = 0_{|E|} \ , \ \pi e_{|E|} = 1 \tag{2.2}$$

satisfied by the invariant vector π can be rewritten in the form

$$\pi_0 B_1 + \pi_1 B_2 = 0_s \tag{2.3.a}$$

$$\pi_0 B_0 + \pi_1 A_1 + \pi_2 A_2 = 0_r \tag{2.3.b}$$

$$\pi_{k-1} A_0 + \pi_k A_1 + \pi_{k+1} A_2 = 0_r \ , \quad 1 < k < K-1 \ , \tag{2.3.c}$$

$$\pi_{K-2} A_0 + \pi_{K-1} A_1 + \pi_K C_2 = 0_r \tag{2.3.d}$$

$$\pi_{K-1} C_0 + \pi_K C_1 = 0_p \ . \tag{2.3.e}$$

Although the model and the correponding balance equations are given for a continuous-time Markov chain, the solution technique also applies to problems formulated in discrete-time, with the understanding that the underlying probability transition matrix is now $T + I_{|E|}$.

2.2. The solution technique

The QBD processes of interest are characterized by the properties (P0)-(P1), where

(P0): *The matrices B_1 and C_1 are nonsingular.*

(P1): *There exist $r \times r$ matrices X and Y such that the equalities*

$$A_0 X = A_2 Y = 0_{r\times r} \ , \quad B_0 X = 0_{s\times r} \quad \text{and} \quad C_2 Y = 0_{p\times r} \tag{2.4}$$

$$A_1(I_r - X)(I_r - Y) = -A_0(I_r - Y) - A_2(I_r - X) \tag{2.5}$$

$$B_2 B_1^{-1} B_0(I_r - Y) = -A_2(I_r - X) \tag{2.6}$$

and

$$XY = YX$$

hold, and one of the $r \times r$ matrices M and N defined by

$$N := A_1 X - A_0 \ , \quad M := A_1 Y - A_2 \tag{2.7}$$

is invertible.

In view of (2.4), postmultiplication of (2.3.b)-(2.3.c) by X and of (2.3.b)-(2.3.d) by Y yields

$$\pi_k A_1 X + \pi_{k+1} A_2 X = 0_r \ , \quad 1 \leq k < K-1 \ , \tag{2.8}$$

and

$$\pi_0 B_0 Y + \pi_1 A_1 Y = 0_r \ , \tag{2.9.a}$$

$$\pi_{k-1} A_0 Y + \pi_k A_1 Y = 0_r \ , \quad 1 < k < K \ , \tag{2.9.b}$$

respectively. With the definitions (2.7) of the matrices M and N, equations (2.8) and (2.9) can be rewritten as

$$\pi_{k-1} (N + A_0) + \pi_k A_2 X = 0_r \ , \ 1 < k < K \ , \tag{2.10}$$

and

$$\pi_0 B_0 Y + \pi_1 M = -\pi_1 A_2 \ , \tag{2.11.a}$$

$$\pi_{k-1} A_0 Y + \pi_k (M + A_2) = 0_r \ , \ 1 < k < K \ , \tag{2.11.b}$$

respectively.

On the other hand, postmultiplication of (2.3.c) by $(I_r - X)(I_r - Y)$ and use of (2.4) yield

$$\pi_{k-1} A_0 (I_r - Y) + \pi_k A_1 (I_r - X)(I_r - Y) + \pi_{k+1} A_2 (I_r - X) = 0_r \ , \ 1 < k < K - 1 \ . \tag{2.12}$$

Substitution of (2.5) into this last relation implies

$$\gamma_k = \gamma_{k-1} \ , \quad 2 < k < K \ , \tag{2.13}$$

where the notation

$$\gamma_k := \pi_k A_2 (I_r - X) - \pi_{k-1} A_0 (I_r - Y) \ , \qquad 1 \leq k \leq K \ ,$$

has been adopted.

The matrix B_1 being invertible, it follows from (2.3.a) that $\pi_0 = -\pi_1 B_2 B_1^{-1}$ and (2.3.b) now becomes

$$\pi_1 (A_1 - B_2 B_1^{-1} B_0) + \pi_2 A_2 = 0_r \ . \tag{2.14}$$

Postmultiplication of (2.14) by $(I_r - X)(I_r - Y)$ and use of the properties (2.4)-(2.6) readily yield

$$\gamma_2 = \pi_2 A_2 (I_r - X) - \pi_1 A_0 (I_r - Y) = 0_r \ .$$

Therefore, it is plain from (2.13) that

$$\gamma_k = 0_r \ , \ 1 < k < K \ ,$$

or equivalently,

$$\pi_k A_2 (I_r - X) = \pi_{k-1} A_0 (I_r - Y) \ , \ 1 < k < K \ . \tag{2.15}$$

Although the equations (2.10), (2.11) and (2.15) all provide a relation between the vectors π_{k-1} and π_k for $1 < k < K$, these relations cannot be exploited to yield in general a ***recursive*** solution for the vectors π_k, $1 \leq k < K$, since the coefficient matrices in these equations are typically singular. In fact, it is easy to check that all the coefficients in (2.10)-(2.11) are singular as a result of the enforced assumptions (2.4). However, the models that motivated the work reported here [6] suggest a different approach which is now briefly discussed. Upon equating the left handsides of (2.10) and (2.11.b), it follows from (2.15) that the recursion

$$\pi_k M = \pi_{k-1} N \ , \ 1 < k < K \ , \tag{2.16}$$

holds, and this leads to the following *structural* result.

Theorem 2.1. *For any finite state QBD process enjoying the properties (P0) and (P1) with M invertible, the invariant probabilities of the non-boundary states are given in matrix-geometric form by*

$$\pi_k = \pi_1 R^{k-1} \ , \ 1 \leq k < K \ , \tag{2.17.a}$$

with

$$R := N M^{-1} \ , \tag{2.17.b}$$

while the invariant probabilities of the boundary states can be expressed in terms of π_1 as

$$\pi_0 = -\pi_1 B_2 B_1^{-1} \ , \qquad \pi_K = -\pi_1 R^{K-2} C_0 C_1^{-1} \ . \tag{2.17.c}$$

The case where N is invertible can be treated in a similar way; details are omitted for sake of brevity. It is worth pointing out that Theorem 2.1 only gives the *structure* of the vector π and that the vector π_1 still needs to be determined. In principle, this could be done by substituting (2.17) back to the balance equation (2.2) where now

$$\pi = \pi_1[-B_2 B_1^{-1}, I_r, R, R^2, \cdots, R^{K-2}, -R^{K-2} C_0 C_1^{-1}] \ .$$

More precisely, π_1 could be computed from either one of the equations

$$\begin{aligned} \pi_1[-B_2 B_1^{-1} B_0 + A_1 + R A_2] &= 0_r \\ \pi_1 R^{k-2}[A_0 + R A_1 + R^2 A_2] &= 0_r \ , \quad 1 < k < K-1 \ , \\ \pi_1 R^{K-3}[A_0 + R A_1 - C_0 C_1^{-1} C_2] &= 0_r \ . \end{aligned} \tag{2.18}$$

Although this approach leads to various necessary conditions for π_1, there does not appear to be any *general* guidelines on how to proceed from here. Indeed, it would be desirable that at least one of the equations (2.18) (or linear combination thereof) have nullity one, so as to allow for unique determination of π_1. Unfortunately, specific examples suggest that the rank of each one of the equations (2.18) is essentially arbitrary, and as of this writing, computation of π_1 from these equations is still under investigation. However, an alternative approach can be taken for many QBD processes that appear in applications; this is discussed in Section 3.5 for the models presented in the next section.

In most applications, s is much smaller than r, and it is thus computationally much more convenient to express the vector π in terms of the vector π_0. To that end, assume that in addition to properties (P0) and (P1) the following property (P2) also holds, where

(P2): *There exists an* $s \times r$ *matrix* V *such that* $B_2 V = A_2$.

In view of (P2) and (2.11.a), postmultiplication of (2.3.a) by V yields

$$\pi_0 B_0 Y + \pi_1 M = -\pi_1 A_2 = \pi_0 B_1 V \ , \tag{2.19}$$

and equations (2.17) and (2.19) combine to give the following Lemma.

Lemma 2.1. *For any finite state QBD process enjoying the properties (P0)-(P2) with M invertible, the invariant probabilities of the system states satisfy the matrix-geometric property*

$$\pi_k = \begin{cases} \pi_0 \, S \, R^{k-1} \ , & 1 \le k < K \ , \\ -\pi_0 \, S \, R^{K-2} C_0 C_1^{-1} \ , & k = K \ , \end{cases} \tag{2.20}$$

where the matrix R *is given in Theorem 2.1 and the* $s \times r$ *matrix* S *is defined by*

$$S := (B_1 V - B_0 Y) M^{-1} \ .$$

3. APPLICATIONS

In this section, the solution methodology just outlined is discussed for four different queueing models. The first-come first-served service discipline is assumed in these models, which all lead to Markov processes with generator matrices of the form (2.1).

The first two models are concerned with two node tandem queueing systems with blocking. In each case, *immediate* blocking is assumed in that blocking of a server occurs as soon as the destination buffer becomes full. The server remains blocked until the congestion is reduced at the destination node, at which time the blocked server resumes service and *begins* to process its next job (if any). The methodology developed in Section 2 also applies to such two node systems under the *non-immediate* blocking policy. Under this policy, the server is blocked at a service completion time if the job that has just completed service cannot proceed to the next buffer due to congestion. When the congestion is reduced downstream, this job proceeds to the next buffer without receiving any further service, and the blocked server resumes service and begins processing its next job (if any). These models can be generalized to capture the situation where the servers are *unreliable* with PH-type up and down time distributions. Similar results hold for this case and the reader is referred to [6] for details.

In both models, the service times at each node are assumed to be *independent* and *identically distributed (i.i.d)* with common PH-distribution given by the *irreducible* representations (α, A) and (β, B) for the first and second node server, respectively. The service times at different servers are also assumed mutually independent. The row vectors α and β, and the matrices A and B have dimensions $1 \times l$, $1 \times m$, $l \times l$ and $m \times m$, respectively, and the corresponding $l \times 1$ and $m \times 1$ column vectors of absorption rates for the first and the second node server are denoted by a and b, respectively.

3.1. Two node tandem system with PH-type servers and feedback

The model consists of two nodes separated by a *finite* intermediate buffer of capacity K, i. e., there are exactly K positions in the buffer, inclusive of the one taken by the job in service at the second node server. There is an *infinite* supply of jobs available in front of the first node server and the second node server *never* gets blocked. The service distributions are of PH-type as described above. A job whose service is completed in the i^{th} node server receives another service from this server with probability p_i, where $0 \le p_i < 1$, $i = 1, 2$, i. e., a job serviced at station 1 joins the intermediate buffer with probability $\overline{p}_1$ and a job serviced at station 2 leaves the system with probability $\overline{p}_2$.

A natural state space E for this system is the one that contains $|E| := (K-1)lm + l + m$ states with

$$E = \begin{cases} (0,i)\ , & k=0,\ \ 1 \le i \le l, \\ (k,i,j)\ , & 0 < k < K,\ \ 1 \le i \le l, \text{ and } 1 \le j \le m, \\ (K,j)\ , & k=K,\ \ 1 \le j \le m, \end{cases}$$

where k indicates the buffer size, and i and j represent the service phase in the first and the second node server, respectively. The phase of the second node server is not defined when it has no jobs to process and the phase of the first node server is not defined when the buffer is full since blocked. With the notation of Section 2, $r = lm$, $s = l$ and $p = m$.

By lexicographically ordering the states, the generator matrix T of the underlying Markov process can be put in the form (2.1) with

$$A_0 = \overline{p}_1 a\alpha \otimes I_m, \qquad A_1 = (A + p_1 a\alpha) \oplus (B + p_2 b\beta), \qquad A_2 = I_l \otimes \overline{p}_2 b\beta, \tag{3.1.a}$$

$$B_0 = \overline{p}_1 a\alpha \otimes \beta, \qquad B_1 = A + p_1 a\alpha, \qquad B_2 = I_l \otimes \overline{p}_2 b, \tag{3.1.b}$$

$$C_0 = \overline{p}_1 a \otimes I_m, \qquad C_1 = B + p_2 b\beta, \qquad C_2 = \alpha \otimes \overline{p}_2 b\beta, \tag{3.1.c}$$

where $\otimes$ and $\oplus$ denote the Kronecker product and the Kronecker sum [4], respectively.

Properties (P1) and (P2) are satisfied by choosing

$$X = (I_l - e_l\alpha) \otimes I_m, \qquad Y = I_l \otimes (I_m - e_m\beta) \quad \text{and} \quad \mathrm{V} = \mathrm{I_l} \otimes \beta. \tag{3.2}$$

The matrices X and Y obviously commute, and the invertibility of the matrices B_1, C_1, M and N is shown in the Appendix.

3.2. Two node closed queueing system with blocking and feedback

This model consists of two nodes which handle a total of N jobs. The service distributions are of PH-type as described above. There is a buffer of capacity K_i, inclusive of the job in service, at node i, $i = 1, 2$, in front of the server. Buffers can have infinite capacity but N is assumed *finite*.

It is also assumed that there is sufficient capacity in the system to accomodate all the jobs, i. e., $N < K_1 + K_2$. A job whose service is completed at the i^{th} node server proceeds to the j^{th} buffer with probability p_{ij}, $i, j = 1, 2$, where $p_{ii} + p_{ij} = 1$, $i \neq j$, $i, j = 1, 2$.

A natural state space E for this system again contains $|E| := (K-1)lm + l + m$ states with

$$E = \begin{cases} (\kappa, i) \;, & k = \kappa, \;\; 1 \leq i \leq l, \\ (k, i, j) \;, & \kappa < k < K, \;\; 1 \leq i \leq l, \text{ and } 1 \leq j \leq m, \\ (K, j) \;, & k = K, \;\; 1 \leq j \leq m, \end{cases}$$

where κ and K are defined as

$$\kappa := \max\{N - K_1, 0\} \;, \qquad K := \min\{K_2, N\} \;.$$

Here, k indicates the number of jobs in the *second* buffer, while i and j represent the service phase in the first and the second node server, respectively. Since the number of jobs in the system is fixed, knowledge of the number of jobs in one buffer gives complete information about the number of jobs in the other buffer. The phase of the first (resp., second) node server is not defined when it has no jobs to process, i. e., $K = N$ (resp., $\kappa = 0$), or when it is blocked, i. e., $K = K_2$ (resp., $\kappa = N - K_1$).

By lexicographically ordering the states, the generator matrix T of the underlying Markov process can be seen to be exactly as in the previous model, but with $p_i = p_{ii}$, $i = 1, 2$. Therefore, the same choice of matrices X, Y and V given by (3.2) can be made for this model.

3.3. The $PH/PH/1/K$ queue with feedback and arrival rejection

A single server queue with a buffer of size K, inclusive of the service station, is considered. Jobs arriving when the buffer is full are considered *lost*. The arrival process is a PH-renewal process and its underlying PH-distribution has irreducible PH-representation (α, A) of order l. The service times are assumed *i.i.d* with a common PH-distribution whose *irreducible* PH-representation (β, B) is of order m. The vectors a and b are as defined above. Arrivals to the system when the buffer is *not* full are rejected and assumed *lost* with probability p_1, while a job whose service is completed is fed back to the buffer with probability p_2, where $0 \leq p_i < 1$, $i = 1, 2$.

The state space E for this system is given by

$$E = \begin{cases} (0, i) \;, & k = 0, \;\; 1 \leq i \leq l, \\ (k, i, j) \;, & 0 < k \leq K, \; 1 \leq i \leq l \text{ and } 1 \leq j \leq m, \end{cases}$$

where k indicates the buffer size, while i and j represent the service phase of the arrival and service processes, respectively. This time, $r = p = lm$ and $s = l$, and the block entries A_i and B_i, $i = 0, 1, 2$, of the matrix T are as given by (3.1.a) and (3.1.b), respectively, while

$$C_0 = A_0, \;\; C_1 = A_1 + A_0 \quad \text{and} \quad \mathrm{C}_2 = \mathrm{A}_2. \tag{3.3}$$

Properties (P1) and (P2) are again satisfied by choosing the matrices X, Y and V as in equation (3.2).

3.4. Queues with paired customers and arrival rejection

The following *infinite* capacity queue with two different job classes is now considered. Jobs for each class arrive into the system according to a PH-renewal processes whose underlying PH-distribution has *irreducible* PH-representation (α, A) and (β, B) for job class I and II, respectively. The PH-representations are of orders l and m for job class I and II, respectively. Job arrivals of class I and II are rejected and assumed *lost* with probability p_1 and p_2, respectively, where $0 \le p_i < 1$, $i = 1, 2$. Accepted arrivals form a queue in front of a single server. Jobs are serviced in pairs, one from each class, and when no such pair is present, the servers stays idle.

At any time, the queue length may be represented by a pair (Y, Z) of random variables where Y denotes the number of job *pairs* present in the system and Z denotes the *excess* number of jobs of class I over those of class II. The range of Y is the set of non-negative integers and Z may assume arbitrary integer values, in that $Z = j$ with $j \ge 0$ (resp. $j < 0$) indicates an excess of j (resp. $-j$) jobs of class I (resp. II).

With the service permitted only in pairs, the excess process Z will tend to $+\infty$ or $-\infty$ depending on the rejection probabilities and the average arrival rates of the job classes, or it will exhibit the behavior typical of symmetric random walks. Therefore, upper and lower bounds J_1 and $-J_2$, $J_1 > 0$, $J_2 > 0$, must be imposed on the excess for an algorithmic solution. With these bounds the excess process $\{Z(t),\ t \ge 0\}$ is $\{-J_2, -J_2 + 1, \cdots, J_1\}$-valued and when it reaches its upper (resp. lower) bound, arrivals of class I (resp. class II) are rejected until the excess no longer assumes the boundary value.

In studying such queueing systems, it is important to obtain the steady-state probabilities of the Markov process with *finite* state set

$$E = \{(k, i, j)\ ,\ -J_2 \le k \le J_1,\ \ 1 \le i \le l \quad \text{and} \quad 1 \le j \le m\}\,,$$

where k indicates the excess number of jobs, while i and j represent the phase of the arrival process for class I and class II jobs, respectively. This time, $r = p = s = lm$ and $K = J_1 + J_2 + 1$. The block entries A_i and C_i, $i = 0, 1, 2$, of the matrix T are given by (3.1.a) and (3.3), respectively, while

$$B_0 = A_0, \quad B_1 = A_1 + A_2 \quad \text{and} \quad \mathrm{B}_2 = \mathrm{A}_2. \tag{3.4}$$

Property (P1) is again satisfied by choosing the matrices X and Y as in equation (3.2), while (P2) is trivially satisfied by $V = I_{lm}$. It is an easy exercise to check that for this model $S = R$.

3.5. Computation of the vector π_0

For the models discussed in this section the special structure of the state space E allows for a computation of the vector π_0 (or π_1) through the calculation of the marginal probabilities of the PH-renewal processes (α, A) and (β, B). The argument is presented for the last model, and the results are only mentioned for the other models.

Let Z denote the number of excess jobs as before, and let PH_1 and PH_2 denote the phase of the PH-renewal processes (α, A) and (β, B), respectively. Let the $1 \times l$ and $1 \times m$ vectors x and y be the solutions of the equations

$$x(A + a\alpha) = 0_l \ , \quad xe_l = 1$$
$$y(B + b\beta) = 0_m \ , \quad ye_m = 1 \, .$$

Then, the marginal probability of the event $\{PH_1 = i, PH_2 = j\}$ can be computed as

$$\begin{aligned} P\{PH_1 = i, PH_2 = j\} &= \sum_{k=0}^{K} P\{Z = k, PH_1 = i, PH_2 = j\}, \quad 1 \le i \le l, \ 1 \le j \le m, \\ &= P\{PH_1 = i\} \, P\{PH_2 = j\} \, , \end{aligned} \tag{3.5}$$

where the second equality follows from the independence assumption of the random variables PH_1 and PH_2. Therefore, (3.5) can be rewritten as a vector equation in the form

$$x \otimes y = \sum_{k=0}^{K} \pi_k = \pi_0 U \tag{3.6}$$

where the $lm \times lm$ matrix U is given by

$$U := S \left(\sum_{k=0}^{K-1} R^{k-1} - R^{K-2} C_0 C_1^{-1} \right) \, .$$

For the other models, the phase of the process (α, A) is defined only when $k = 0$ and (3.6) takes the form

$$x = c\pi_0 \left[I_l + S \left(\sum_{k=1}^{K-1} R^{k-1} - dR^{K-2} C_0 C_1^{-1} \right) (I_l \otimes e_m) \right] \tag{3.7}$$

where the scalars c and d take different values depending on the model. The term $I_l \otimes e_m$ corresponds to summation of the right hand-side of (3.5) over j, $1 \le j \le m$. For the third model, $c = d = 1$. For the first two models, the phase of the process (α, A) is *not* defined when $k = K$. Therefore (3.7) is obtained by conditioning on the event that the buffer is not full, and $c = (1 - \pi_K e_m)^{-1}$ while $d = 0$. It is an easy exercise to show that for the first two examples π_0 can be obtained by solving the linear system

$$x = \pi_0 \left[I_l + S \sum_{k=1}^{K-1} R^{k-1} (I_l \otimes e_m) - S R^{K-2} C_0 C_1^{-1} e_m x \right] \, . \tag{3.8}$$

Although no proof is available, extensive numerical evidence indicates that equations (3.6)-(3.8) lead to a unique vector π_0. This is congruent with the fact that in all four models, the enforced irreducibility of the phase-type servers guarantees the corresponding Markov chain to be irreducible.

4. ACKNOWLEDGEMENTS

The authors are thankful to the reviewers for their careful reading of an earlier version of this paper, and for pointing out the reference [7]. The authors would also like to acknowledge conversations with Dr. Harry Heffes from AT&T Bell Laboratories, which provided the basis for some of the ideas discussed in this paper.

Proof. The Lemma is proved only for N as the nonsingularity of M follows along the same lines. The matrices $(A + p_1 a\alpha)$ and $(B + p_2 b\beta)$ being nonsingular by Lemma 1, the matrix N can be rewritten in the form

$$\begin{aligned} N &= (A + p_1 a\alpha) \otimes I_m + (I_l - e_l\alpha) \otimes (B + p_2 b\beta) \\ &= [(A + p_1 a\alpha) \otimes I_m]\, D\, [I_l \otimes (B + p_2 b\beta)] \end{aligned}$$

where

$$D := (A + p_1 a\alpha)^{-1}(I_l - e_l\alpha) \oplus (B + p_2 b\beta)^{-1} .$$

This clearly shows that N is nonsingular if and only if the matrix D is nonsingular. Each eigenvalue of the matrix D is the sum of the eigenvalues of the matrices $(A+p_1 a\alpha)^{-1}(I_l - e_l\alpha)$ and $(B+p_2 b\beta)^{-1}$ [4]. By Lemma A.1, it is an easy exercise to check that the eigenvalues of $(B+p_2 b\beta)^{-1}$ have strictly negative real parts. The matrix D is thus invertible if the real parts of the eigenvalues of the matrix $(A + p_1 a\alpha)^{-1}(I_l - e_l\alpha)$ are non-positive.

Let (γ, y) be any right eigenpair for $(A + p_1 a\alpha)^{-1}(I_l - e_l\alpha)$, and let $\Re e(\gamma)$ denote the real part of γ. The argument proceeds *ab absurdo* by assuming $\Re e(\gamma) > 0$. Since (γ, y) is a right eigenpair, the relation

$$(A + p_1 a\alpha)^{-1}(I_l - e_l\alpha)y = (A + p_1 a\alpha)^{-1}y - (\alpha y)(A + p_1 a\alpha)^{-1}e_l = \gamma y \qquad (A.2)$$

holds and therefore $\alpha y \neq 0$; indeed, otherwise (γ, y) would also be a right eigenpair for $(A+p_1 a\alpha)^{-1}$ and by Lemma A.1 this would contradict the assumption $\Re e(\gamma) > 0$.

Since $\alpha y \neq 0$, assume without loss of generality that $\alpha y = 1$ and note from (A.2) that $[I_l - \gamma(A + p_1 a\alpha)]y = e_l$. The assumption $\Re e(\gamma) > 0$ implies that the matrix $[I_l - \gamma(A + p_1 a\alpha)]$ is invertible and the relations

$$y = [I_l - \gamma(A + p_1 a\alpha)]^{-1}e_l = \frac{1}{\gamma}[\frac{1}{\gamma}I_l - (A + p_1 a\alpha)]^{-1}e_l \qquad (A.3)$$

thus follow. Premultiplying the last equation by α now yields

$$F^*(\frac{1}{\gamma}) = 1 - \frac{1}{\gamma}\alpha[\frac{1}{\gamma}I_l - (A + p_1 a\alpha)]^{-1}e_l = 0 \qquad (A.4)$$

where $F^*(\cdot)$ is the Laplace transform of the PH-distribution with representation $(\alpha, A + p_1 a\alpha)$ [9]. However, the Laplace transform of a non-negative random variable is strictly positive in the right half-plane and the assumption $\Re e(\gamma) > 0$ thus contradicts (A.4). Consequently, $\Re e(\gamma) \leq 0$ and the proof is now completed by the arguments given in the first paragraph. $\square$

REFERENCES

[1] I. F. Akyıldız, "Exact product form solution for queueing networks with blocking", *IEEE Trans. on Computers*, **36**, 1987, pp. 122-125.

[2] K.E. Atkinson, *An Introduction to Numerical Analysis*, Wiley, 1980.

[3] P.P. Bocharov and V.A. Naumov, "Matrix-geometric stationary distribution for the $PH/PH/1/N$ queue," *INRIA Rapport de Recherche* **304**, INRIA, France, 1984.

[4] J.W. Brewer, " Kronecker products and matrix calculus in systems theory," *IEEE Trans. on Circuits and Systems*, **25**, 1978, pp. 772-781.

[5] S. Chakravarthy and M.F. Neuts, "Algorithms for the design of finite capacity service units," *Working Paper* **87-002**, College of Engineering, The University of Arizona, Tucson, Arizona, 1987.

[6] L. Gün, " Tandem queueing systems subject to blocking with phase type servers: Analytic solutions and approximations," *SRC Technical Report* **87-002**, Systems Research Center, University of Maryland, College Park, MD, 1987.

[7] B. Hajek, " Birth-and-death processes on the integers with phases and general boundaries," *J. Appl. Prob.*, **19**, 1982, pp. 488-499.

[8] G. Latouche, "Queues with paired customers," *J. Appl. Prob*, **18**, 1981, pp. 684-696.

[9] M.F. Neuts, *Matrix-geometric Solutions in Stochastic Models - An Algorithmic Approach ,* The John Hopkins Univ. Press, Baltimore, Maryland, 1981.