Quasi-Birth-and-Death Processes with Rational Arrival Process Components

N.G. Bean[1] and B.F. Nielsen[2]

1 *Applied Mathematics, University of Adelaide, Australia 5005.*

2 *Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.*

IMM-Technical Report-2007-20

ABSTRACT

In this paper we introduce the concept of a Quasi-Birth-and-Death process (QBD) with Rational Arrival Process components. We use the physical interpretation of a Rational Arrival Process (RAP), developed by Asmussen and Bladt, to consider such a Markov process. We exploit this interpretation to develop an analytic method for such a process, that parallels the analysis of a traditional QBD. We demonstrate the analysis by considering a queue where the arrival process and the sequence of service times are derived from two different RAPs that are not just Markovian Arrival processes. We also introduce an element of correlation between the arrival process and the sequence of service times.

Key words: Matrix Exponential Distribution, Phase-Type Distribution, Rational Arrival Process, Markovian Arrival Process, Quasi-Birth-and-Death Process.

1 Introduction

Since the work of Lipsky [7] there has been an expectation that the matrix-analytic results of Neuts [12], and others, would carry over *verbatim* to the more general setting of Matrix-exponential distributions. This became more likely with the characterization of Rational Arrival Processes by Asmussen and Bladt [1], as a similar extension of Markovian Arrival Processes, and was in fact explicitly stated therein. See also Mitchell [11] for related work. However, queueing results in the matrix-analytic methods setting rely on path-wise probabilistic arguments, exploiting the Markovian nature via the discrete states.

Asmussen and Bladt [1] developed an intriguing physical interpretation of a Rational Arrival Process (RAP), which was explored further by Bladt and Neuts [4]. In this paper we use this physical interpretation to consider a class of processes that we call Quasi-Birth-and-Death processes with Rational Arrival Process components (QBDs with RAP components). We demonstrate how the analysis of such a process can be performed by interpreting the appropriate matrices as operators on measures, rather than as transition matrices of Markov chains. Information on the process is now carried by a vector of weighting factors of measures, which in the standard Markov chain setting have a direct interpretation as posterior probabilities. This vector is essentially the same (up to continuous re-normalization) as the vector of container levels referred to by Bladt and Neuts [4, Section 5].

We do not prove new formulae, instead our results extend well-known matrix-analytic results to the more general setting of queues with Matrix-exponential and Rational Arrival Process components. Thus most, if not all, formulae in this paper are well-known from the traditional matrix-analytic methods setting.

The approach followed in this paper is a direct approach in the sense that our proofs are generalizations of elegant and relatively elementary probabilistic proofs first suggested by Ramaswami [13].

In a forthcoming paper we demonstrate a more abstract technique which demonstrates how the framework of this paper fits in to a more general framework. We address numerical issues in this forthcoming paper too.

The structure of the paper is as follows. In Section 2 we briefly introduce Matrix-exponential (ME) distributions and Rational Arrival Processes (RAPs). In Section 2.1 we describe the physical interpretation of the RAP developed by Asmussen and Bladt [1],

which we call the *orbit representation*. In Section 2.2 we present a brief discussion of the related interpretation of Bladt and Neuts [4] and attempt to motivate our choice to use the interpretation of Asmussen and Bladt rather than that of Bladt and Neuts.

In Section 3 we describe a QBD with RAP components as a specific example of a Markov process on a mixed continuous-discrete state space. We prove existence of such processes by presenting a sequence of examples, ranging from well-known structures like the MAP/PH/1 queue through to RAP/RAP/1 queues with dependencies between the service and arrival processes. Section 4 contains all the theorems of the paper, presenting the required justification for the use of the traditional analysis of a traditional QBD in the extended world of the QBD with RAP components. All our formulae are well-known from the traditional matrix-analytic methods literature, but the physical interpretations are now re-formulated to support the extended world of the QBD with RAP components.

In Section 5 we present our example queue. The arrival process and the sequence of service times come from two different RAPs (outside the space of MAPs) that share the same phase space. Thus we also have cross-correlations between the arrival process and the sequence of service times. To the authors' knowledge there are no existing methods for analysing such a queue. We deliberately chose the RAP representing the sequence of service times to have a rank-one jump matrix. This means that we can identify the matrix G *a priori* and therefore do not require numerical algorithms for its determination. Finally, in Section 6 we present our conclusions, including a discussion of further research required in this area as well as some speculation about how these techniques could find application in practical situations.

2 Matrix–exponential distributions (ME) and Rational Arrival Processes (RAP)

A matrix–exponential distribution is a distribution on the non–negative real numbers with rational Laplace (Stieltjes) transform. Equivalently [9, 8] the distribution can be expressed via a matrix–exponential density function: that is, for positive x, $f(x) = \boldsymbol{\alpha} e^{Tx} \boldsymbol{t}$, where $\boldsymbol{\alpha}$ is a $1 \times p$ row vector, T is a $p \times p$ matrix and $\boldsymbol{t}$ is a $p \times 1$ column vector; and with point mass α_0 for $x = 0$. Throughout, we assume that $\alpha_0 = 0$ as we do not wish to consider distributions with a point mass at zero. One can choose an ME-representation such that

$\boldsymbol{\alpha e} = 1$, $T\boldsymbol{e} = -\boldsymbol{t}$ (where $\boldsymbol{e}$ is the $p \times 1$ column vector consisting entirely of ones) and further such that $(\boldsymbol{\alpha}_i, T)$, is the representation for a matrix–exponential distribution for all $i = 1, \ldots, p$ where $\boldsymbol{\alpha}_i$ has a one in the i'th position and zeros elsewhere [1, Page 135].

As in Asmussen and Bladt [1] we will consider the three ME–distributions $f_1(x) = e^{-x}$, $f_2(x) = \frac{2}{3}(1 + \sin(x))e^{-x}$, and $f_3(x) = \frac{2}{3}(1 + \cos(x))e^{-x}$. The representations are given by $(\boldsymbol{\alpha}_1, T)$, $(\boldsymbol{\alpha}_2, T)$, and $(\boldsymbol{\alpha}_3, T)$, with

$$T = \begin{bmatrix} -1 & 0 & 0 \\ -\frac{2}{3} & -1 & 1 \\ \frac{2}{3} & -1 & -1 \end{bmatrix} . \tag{1}$$

Asmussen and Bladt generalized the concept of a renewal process with matrix–exponentially distributed inter–event times by defining the class of point processes which are contained in a finite dimensional subspace under the time shift operator. As their main result they proved that a RAP can always be described, for some finite positive integer p, by two $p \times p$ matrices, C and D, and a row vector of length p, $\boldsymbol{\alpha}$, such that

1. the dominant eigenvalue of C is negative,
2. the dominant eigenvalue of $C + D$ is zero, and
3. $(C + D)\boldsymbol{e} = \boldsymbol{0}$.

The density of the RAP is then given by

$$f_n(x_1, x_2, \ldots, x_n) = \boldsymbol{\alpha} e^{Cx_1} D e^{Cx_2} D \ldots e^{Cx_n} D \boldsymbol{e},$$

where $\boldsymbol{\alpha e} = 1$.

The two matrices can be chosen such that additionally the process is a RAP whenever $\boldsymbol{\alpha} = \boldsymbol{\alpha}_i$, [1, page 130].

Thus Asmussen and Bladt's quite general definition proved to be **analytically** equivalent to the Markovian Arrival Process [10] (MAP) in terms of its representation in the form of two matrices C and D.

For a MAP we need to impose the following additional requirements: D is a nonnegative matrix, $\boldsymbol{\alpha}$ is a nonnegative vector and C has nonnegative off-diagonal entries and negative diagonal entries. These restrictions guarantee that a MAP has a very simple physical interpretation as a continuous-time Markov chain that moves around the state space of

dimension p according to the generator matrix $Q = C + D$, where transitions due to the matrix D also correspond to points in the point process. In contrast, the matrix representation of a RAP does not make these extra requirements, but merely requires that the conditions 1–3 are obeyed.

Although Matrix–exponential distributions contain and are analytically equivalent with phase–type distributions and Rational Arrival Processes contain and are analytically equivalent with Markovian Arrival Processes their probabilistic interpretation differ significantly: the probabilistic interpretation of PH distributions and MAPs in terms of finite dimensional Markov chains is not valid in the general ME and RAP cases.

However, Asmussen and Bladt [1, page 129] provide an intriguing physical interpretation, where they state that

> "... the RAP N can be seen as generated by a piecewise deterministic Markov process $\{\boldsymbol{A}(t)\}_{t\geq 0}$ on a compact convex subset of $\mathbb{R}^p$, such that the epochs of N are identical to the jump times of $\boldsymbol{A}(\cdot)$. Note that it is not surprising that a finite-state space does not suffice – it is well known that the MAP is the most general point process such that the counting process is an additive process on a finite Markov chain."

A consequence of the extra flexibility in RAPs is that, given a set of matrices according to the conditions 1–3 above, it is not immediately possible to determine whether they represent a RAP or not. Further conditions are required, as we shall see later.

We refer the reader to [1, 2, 5] for further results and discussion. Also note that there is an alternative construction of RAPs in Mitchell [11] as a correlated sequence of matrix exponentials. However, Mitchell does not delve into the details of the construction.

Asmussen and Bladt [1] provide an example of an ME renewal version of a RAP (C_s, D_s) with $C_s = T$, as given in equation (1), and $D_s = \left(1, \frac{2}{3}, \frac{4}{3}\right)^T (3, -1, -1)$ and of a non-ME-renewal process (C_a, D_a) with $C_a = T$ and

$$D_a = \begin{bmatrix} \frac{14}{5} & -\frac{9}{10} & -\frac{9}{10} \\ \frac{26}{15} & -\frac{8}{15} & -\frac{8}{15} \\ \frac{58}{15} & -\frac{19}{15} & -\frac{19}{15} \end{bmatrix} . \tag{2}$$

We will use these two RAPs later in our Section 5.

2.1 The piecewise deterministic Markov process $\{\boldsymbol{A}(t)\}_{t\geq 0}$

In this section we discuss the properties of the piecewise deterministic phase vector $\boldsymbol{A}(t)$ of the RAP, which is the fundamental aspect of Asmussen and Bladt's [1] physical interpretation. The phase vector $\boldsymbol{A}(t)$ is also the continuously re-normalised version of the vector of container levels considered by Bladt and Neuts [4].

In Proposition 2.1(a), Asmussen and Bladt [1] show that the piecewise deterministic Markov process, $\boldsymbol{A}(\cdot)$, evolves according to

$$\mathbb{E}\left[\boldsymbol{A}(t+s)|\mathcal{F}_t\right] = \boldsymbol{A}(t)e^{Qs}, \tag{3}$$

where $Q = C + D$ and $\mathcal{F}_t$ represents the knowledge of the entire history of the process up to and including time t.

They also show, in Proposition 2.4 and the ensuing discussion, that the piecewise deterministic Markov process, $\boldsymbol{A}(\cdot)$, evolves **between jumps** (that is, points in the point process) according to the deterministic differential equation

$$\begin{aligned} \dot{\boldsymbol{a}}(t) &= \boldsymbol{a}(t)C - \boldsymbol{a}(t)C\boldsymbol{e}\cdot\boldsymbol{a}(t), \\ &= \boldsymbol{a}(t)\left(C + D\boldsymbol{e}\cdot\boldsymbol{a}(t)\right), \\ &= \boldsymbol{a}(t)\left(C + \boldsymbol{a}(t)D\boldsymbol{e}I)\right), \end{aligned} \tag{4}$$

the latter expression being more transparent in showing how the uninterrupted piecewise process evolves – including the explicit continuous re-normalisation mentioned above. It is important to note that the state space on which the process $\boldsymbol{A}(\cdot)$ exists, which we denote by the set $\mathcal{A}$, is a subset of the hyperplane $\{\boldsymbol{a} \in \mathbb{R}^p : \boldsymbol{a}\boldsymbol{e} = 1\}$.

Jumps (or points) occur at intensity $\boldsymbol{a}D\boldsymbol{e}$ in state $\boldsymbol{a}$, and the new state after the jump/point is deterministically given by

$$\frac{\boldsymbol{a}D}{\boldsymbol{a}D\boldsymbol{e}}.$$

The expression is well–defined as jumps can occur only when $\boldsymbol{a}D\boldsymbol{e} > 0$.

It is worth noting that a set of matrices that satisfies the conditions in the previous section represent a RAP if and only if

a. the state space, $\mathcal{A}$, of the process $\boldsymbol{A}(\cdot)$ is a compact and convex set, and

b. $\boldsymbol{a}D\boldsymbol{e} \geq 0$ for all $\boldsymbol{a} \in \mathcal{A}$, that is the stochastic intensity of points in the point process must be nonnegative at all times.

In Corollary 2.2 of [1], Asmussen and Bladt give an explicit representation of $\boldsymbol{A}(\cdot)$ conditional on the times of jumps/points before time t. Since the evolution of $\boldsymbol{A}(\cdot)$ is deterministic in periods between the stochastic jumps, we feel that the simplified representation of $\boldsymbol{A}(\cdot)$ between jumps/points is as useful. Therefore,

$$\boldsymbol{A}(t) = \frac{\boldsymbol{A}(0)e^{Ct}}{\boldsymbol{A}(0)e^{Ct}\boldsymbol{e}},$$

given that the process started in state $\boldsymbol{A}(0)$ after a jump/point at time 0 and no other jump/point has occurred before time t. Again, the continuous re-normalisation is made explicit here.

Terminating RAPs

We need the concept of a Terminating RAP in later sections of the paper. To define such a process we write the matrix $D \equiv D_1 + \boldsymbol{d}_0\boldsymbol{e}^{\mathrm{T}}$, where ^{T} denotes matrix transposition. We require that for all $\boldsymbol{a} \in \mathcal{A}$ $\boldsymbol{a}D_1\boldsymbol{e} \geq 0$, $\boldsymbol{a}\boldsymbol{d}_0 \geq 0$, and $\frac{\boldsymbol{a}D_1}{\boldsymbol{a}D_1\boldsymbol{e}} \in \mathcal{A}$ whenever $\boldsymbol{a}D_1\boldsymbol{e} > 0$. The terminating RAP behaves almost as the ordinary RAP, whereby the behaviour of the phase process between events are given by the usual differential equation and events are generated with intensity $\boldsymbol{A}(t)D_1\boldsymbol{e}$. However, with intensity $\boldsymbol{A}(t)\boldsymbol{d}_0$ the process terminates, which can be interpreted as $\boldsymbol{A}(t)$ leaving $\mathcal{A}$ from that point in time. For convenience, we set $\boldsymbol{A}(t) = \boldsymbol{0}$.

2.2 Interpretation in terms of fluid flow

Bladt and Neuts [4] introduced the interpretation of the state vector of the ME distributions as fluid flowing between various containers with the important restriction that the content of the last container should be non–decreasing non–negative and tending towards 1 as time tends to infinity. Their interpretation is quite intuitive and they illustrate through a number of examples how one can generalize proofs in the PH–context that rely on differential equations involving the generator matrix. They conjecture that one can generalize any such proof and we believe that.

However, we are not aware of any proofs for the existence of the (infinite) matrix polynomial that solves the M/G/1 or GI/M/1 queue that rely solely on differential equations of the Chapman–Kolmogorov type. Rather such proofs rely on analyses based on the path–wise behavior of the process. These proofs cannot be generalized immediately by the fluid

interpretation as the concept of a state in a Markov chain no longer exists. Rather one must keep the whole vector in mind simultaneously in all calculations – in the language of Bladt and Neuts analyze all flows simultaneously. Our contribution illustrates one way of doing this.

Although the notion of flows is intuitively appealing and thus a very good tool for the understanding of many aspects of ME–distributions and RAPs we have chosen the slightly more abstract notion of the phase vector of the RAP/ME process as the time-dependent vector of weights for measures. We refer to this as the orbit representation and hope that the reader will bear with us in this respect and might appreciate the distinction – which is purely in the interpretation – in some places. Also, as in our arguments we use conditional probabilities (thus requiring continuous renormalization) it is less natural to use the fluid description.

3 A QBD with RAP Components

3.1 Definition

Suppose we have a Markov process $X(t) = (N(t), \boldsymbol{A}(t))$ on a state space $\mathbb{Z} \times \mathcal{A}$ with the property that $\mathcal{A}$ is a compact and convex subset of $\mathbb{R}^m$.

We will refer to the first component $N(t)$ of $X(t)$ as the level and to the second component $\boldsymbol{A}(t)$ as the phase vector. The process $X(t)$ is further defined by three $p \times p$ matrices A_0, A_1, and A_2 as follows. At any moment in time the process changes levels in such a way that $P(N(t + \mathrm{d}t) = i + 1 | X(t) = (i, \boldsymbol{a})) = \boldsymbol{a} A_0 \boldsymbol{e} + o(\mathrm{d}t)$ and $P(N(t + \mathrm{d}t) = i - 1 | X(t) = (i, \boldsymbol{a})) = \boldsymbol{a} A_2 \boldsymbol{e} + o(\mathrm{d}t)$. Whenever an upwards jump occurs from $\boldsymbol{a}$ the new value of $A(t)$ is deterministically determined to be $\boldsymbol{A}(t+) = \frac{\boldsymbol{a} A_0}{\boldsymbol{a} A_0 \boldsymbol{e}}$, and correspondingly for a downwards jump $\boldsymbol{A}(t+) = \frac{\boldsymbol{a} A_2}{\boldsymbol{a} A_2 \boldsymbol{e}}$, which are both well-defined as level changes can only occur when the relevant denominator is positive. Between level changes the phase vector deterministically follows the orbit given by the set of linear differential equations

$$\dot{\boldsymbol{A}}(t) \quad = \quad \boldsymbol{A}(t) \left(A_1 + \boldsymbol{A}(t)(A_0 + A_2)\boldsymbol{e} I \right)) \, .$$

Suppose we would like to restrict our attention to the case where the level is non–negative. To accomplish this we introduce a $p \times p$ matrix B_1 to describe the behaviour at level zero. At this level we no longer allow downward jumps; upward jumps still occur, but

now according to $\boldsymbol{a}B_0\boldsymbol{e}$, where B_0 is also a $p \times p$ matrix. The deterministic behaviour of $\boldsymbol{A}(t)$ when $N(t) = 0$ is thus given by

$$\dot{\boldsymbol{A}}(t) \quad = \quad \boldsymbol{A}(t)\,(B_1 + \boldsymbol{A}(t)B_0\boldsymbol{e}I)\,.$$

This Markov process now has the state space $(\{0\} \times \mathcal{A}_0) \cup (\mathbb{N}_+ \times \mathcal{A}_+)$, since the state space of the phase vector on level 0 is generally distinct from the state space $\mathcal{A}_+$ on levels above 0. Both $\mathcal{A}_0$ and $\mathcal{A}_+$ are now compact and convex subsets of $\mathbb{R}^p$.

3.2 Existence

Such a process clearly exists as the MAP/PH/1-queue, or for that sake the QBD, is an example. In these cases the probability distribution among the phases amounts to the phase vector $\boldsymbol{A}(t)$. Similarly the counting process of a RAP can be formulated as such a process, with $A_0 = B_0 = D$, $A_1 = B_1 = C$, and $A_2 = 0$. Here the interpretation of the vector $\boldsymbol{A}(t)$ is as the vector of weights of the measures at time t - which we have called the orbit. The important difference now being that some of these weights might be negative.

The RAP/ME/1 queue can be formulated as a QBD with RAP components. More generally we will consider the RAP(C, D)/RAP(E, F)/1 queue. Let the compact and convex set of the phase vector of the first RAP be denoted by $\mathcal{C}$, while the compact and convex set of the second RAP be denoted by $\mathcal{E}$. The state spaces of the phase vector are simply defined as $\mathcal{A}_0 = \mathcal{A}_+ = \mathcal{C} \times \mathcal{E}$. The governing matrices are then given by $A_1 = C \oplus E$, $A_2 = I_1 \otimes F$, and $A_0 = D \otimes I_2$, where I_1 and I_2 are the identity matrices of the same dimensions as the matrices C and E respectively. To model the behaviour in level 0, we simply assume that the service-time RAP is suspended in level 0 and so choose $B_0 = A_0$ and $B_1 = C \otimes I_2$.

As the two processes behave independently of each other it should be obvious that everything is well-defined. However, a formal proof goes as follows. We first show that the upward intensities are all non–negative. First we consider the expression $\exp(A_1 t) = \exp([C \oplus Et])$. The matrix–exponential solves two completely decoupled systems of differential equations thus the matrix–exponential decouples too. In terms of the Kronecker sum we see that $C \otimes I_2$ commutes with $I_1 \otimes E$ thus $\exp(A_1 t) = \exp(Ct) \otimes \exp(Et)$. Now

$$\begin{aligned} \boldsymbol{a}\exp(A_1 t)A_0\boldsymbol{e} &= (\boldsymbol{a}_1 \otimes \boldsymbol{a}_2)\exp(C \oplus Et)(D \otimes I)(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2), \\ &= \boldsymbol{a}_1 \exp(Ct)D\boldsymbol{e}_1\boldsymbol{a}_2 \exp(Et)\boldsymbol{e}_2, \end{aligned}$$

where $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are column vectors consisting entirely of ones with the number of rows given by that of C and E, respectively.

This is non–negative since

- $\boldsymbol{a}_1 \exp(Ct) D \boldsymbol{e}_1 \geq 0$ as RAP(C, D) is well-defined, and
- $\boldsymbol{a}_2 \exp(Et) \boldsymbol{e}_2$ is the probability of RAP(E, F) having no points in time t, which is clearly positive.

Further, we see that an upward jump from $(\boldsymbol{a}_1 \otimes \boldsymbol{a}_2)$ at time t will be to state

$$\begin{aligned} \frac{\boldsymbol{a} \exp(A_1 t) A_0}{\boldsymbol{a} \exp(A_1 t) A_0 \boldsymbol{e}} &= \left(\frac{\boldsymbol{a}_1 \exp(Ct) D \otimes \boldsymbol{a}_2 \exp(Et)}{\boldsymbol{a}_1 \exp(Ct) D \boldsymbol{e}_1 \boldsymbol{a}_2 \exp(Et) \boldsymbol{e}_2} \right), \\ &= \left(\frac{\boldsymbol{a}_1 \exp(Ct) D}{\boldsymbol{a}_1 \exp(Ct) D \boldsymbol{e}_1} \otimes \frac{\boldsymbol{a}_2 \exp(Et)}{\boldsymbol{a}_2 \exp(Et) \boldsymbol{e}_2} \right), \end{aligned}$$

which obviously is within $\mathcal{A}_+$. The argument is similar for downward jumps.

With potential interdependence between arrivals and departures the situation becomes more intricate. In the standard Markovian case we have no difficulties (as usual) while in the general case we have to add additional conditions, reflecting conditions a. and b. in Section 2.

- The state spaces of the phase vector in level zero, $\mathcal{A}_0$, and in all levels above zero, $\mathcal{A}_+$, are compact and convex spaces.
- $\boldsymbol{a} B_0 \boldsymbol{e} \geq 0$ for all $\boldsymbol{a} \in \mathcal{A}_0$, $\boldsymbol{a} A_0 \boldsymbol{e} \geq 0$ for all $\boldsymbol{a} \in \mathcal{A}_+$ and $\boldsymbol{a} A_2 \boldsymbol{e} \geq 0$ for all $\boldsymbol{a} \in \mathcal{A}_+$.
- If $\boldsymbol{a} \in \mathcal{A}_0$ is such that $\boldsymbol{a} B_0 \boldsymbol{e} > 0$ then $\frac{\boldsymbol{a} B_0}{\boldsymbol{a} B_0 \boldsymbol{e}} \in \mathcal{A}_+$, if $\boldsymbol{a} \in \mathcal{A}_+$ is such that $\boldsymbol{a} A_0 \boldsymbol{e} > 0$ then $\frac{\boldsymbol{a} A_0}{\boldsymbol{a} A_0 \boldsymbol{e}} \in \mathcal{A}_+$, and if $\boldsymbol{a} \in \mathcal{A}_+$ is such that $\boldsymbol{a} A_2 \boldsymbol{e} > 0$ then $\frac{\boldsymbol{a} A_2}{\boldsymbol{a} A_2 \boldsymbol{e}} \in \mathcal{A}_+$.

This is no different from the standard challenge when modeling with RAP's and ME–distributions. Given a matrix–exponential kernel it is in general a complicated task to determine whether the parameters actually reflect a distribution or even a process. See the original work by Asmussen and Bladt [1] or more recent work by Fackrell [5] for a discussion of these issues.

What we show in the next section is, that whenever the matrices involved satisfy the above conditions then the well-known non–linear matrix equations can be used to solve the system with adjustments needed only in their interpretation.

We now provide a non–trivial example with cross-correlation between the arrival process and the sequence of service times. A queue which to our knowledge is not immediately amenable to analysis by any other technique. The example is constructed by choosing $A_1 = C_a = C_s$, $A_0 = \gamma D_a$, and $A_2 = (1-\gamma)D_s$, where $\gamma \in [0,1]$. The orbit of both of the two RAPs (C_a, D_a), and (C_s, D_s) is always contained in the same set $\mathcal{C}$ as discussed in detail in [1]. Accordingly, we know that $\mathcal{A}_0 = \mathcal{A}_+ = \mathcal{C}$ if we choose $B_0 = A_0$ and $B_1 = \gamma A_1$.

Thus this QBD queue with RAP-components satisfies the above conditions and is at the same time a non–trivial example extending beyond the standard QBD case and with cross–correlation between the arrival process and the sequence of service times.

4 Analysis and Interpretations

For all $n = \{0, 1, 2, \ldots\}$, let τ_n be the first passage time to any state in the level n. We impose the condition that, if $N(0) = n$, then $\tau_n > \min(\tau_{n-1}, \tau_{n+1})$.

At the risk of being slightly repetitive of the previous section, we take this opportunity to succinctly restate the essential structure of the QBD with RAP components including statements of the physical interpretations of the key quantities in terms of first-hitting times:

The Markov process $X(t)$ evolves on level 1 and above as follows (with obvious minor notational modifications for level 0). During intervals where the level does not change, the phase evolves according to the differential equation (4) with matrix $C = A_1$. The level decreases by one with intensity $\boldsymbol{A}(t)A_2\boldsymbol{e}$ at time t and if such a level change occurs at time t then the phase process jumps immediately to the state $\dfrac{\boldsymbol{A}(t)A_2}{\boldsymbol{A}(t)A_2\boldsymbol{e}}$. Similarly, the level increases by one with intensity $\boldsymbol{A}(t)A_0\boldsymbol{e}$ at time t and if such a level change occurs at time t then the phase process jumps immediately to the state $\dfrac{\boldsymbol{A}(t)A_0}{\boldsymbol{A}(t)A_0\boldsymbol{e}}$. Thus,

$$
\begin{aligned}
\boldsymbol{a}A_0\boldsymbol{e} &= \frac{\mathrm{d}}{\mathrm{d}h}\mathbb{P}\left[\tau_{n+1} \le h | X(0) = (n, \boldsymbol{a})\right]\Big|_{h=0}, \\
\frac{\boldsymbol{a}A_0}{\boldsymbol{a}A_0\boldsymbol{e}} &= \mathbb{E}\left[\boldsymbol{A}(\tau_{n+1}) | X(0) = (n, \boldsymbol{a}) \ \&\ \tau_{n+1} = 0\right], \\
\boldsymbol{a}A_2\boldsymbol{e} &= \frac{\mathrm{d}}{\mathrm{d}h}\mathbb{P}\left[\tau_{n-1} \le h | X(0) = (n, \boldsymbol{a})\right]\Big|_{h=0}, \quad \text{and} \\
\frac{\boldsymbol{a}A_2}{\boldsymbol{a}A_2\boldsymbol{e}} &= \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1}) | X(0) = (n, \boldsymbol{a}) \ \&\ \tau_{n-1} = 0\right].
\end{aligned}
$$

One powerful tool we shall be using is that of a censored (or restricted) Markov process, see [6, Section 5.3]. We shall specifically consider the process censored to be on level n. That is, the Markov process that is observed, from the entire Markov process, when and only when it is in the level n. This construction naturally creates two different measures of time, which we shall call global time and local time. Global time is the time measured by the entire Markov process. Local time is the time measured by the censored process, consisting of level n only. Local time therefore increases when and only when the process is in level n. When the process is in level n both measures of time advance at exactly the same rate. For all $n = \{0, 1, 2, \ldots\}$ we let $\ell_n(t)$ denote the local time spent in level n after t units of global time and $g_n(t)$ denote the global time when the local time in level n first reaches t units.

Let $\{\boldsymbol{B}(t)\}_{t\geq 0}$ be the censored process consisting of level n only, measured in the local time of level n, and with level $n-1$ taboo.

The process $\{\boldsymbol{B}(t)\}_{t\geq 0}$ exists as it is merely a censored process derived from a taboo Markov process. However, the censored process $\boldsymbol{B}(t)$ is not necessarily a RAP, with the epochs of entering upper levels interpreted as the jumps, since jumps in the phase process corresponding to successive visits to level $n+1$ will not in general return to a state deterministically derived from $\boldsymbol{B}(t)$. In order to describe the distribution of the return state we introduce

$$\psi(\mathcal{B}; \boldsymbol{a}) = \mathbb{P}(\boldsymbol{A}(\tau_n) \in \mathcal{B} | X(0) = (n+1, \boldsymbol{a})), \qquad \mathcal{B} \subset \mathcal{A}_+.$$

Since the behaviour of the process on sample paths constrained to lie above level n is the same for all levels $n > 0$, we note that we could equivalently define $\psi(\mathcal{B}; \boldsymbol{a})$ as

$$\psi(\mathcal{B}; \boldsymbol{a}) = \mathbb{P}(\boldsymbol{A}(\tau_{n-1}) \in \mathcal{B} | X(0) = (n, \boldsymbol{a})), \qquad \mathcal{B} \subset \mathcal{A}_+.$$

To later enable us to follow more closely the conventions in the matrix-analytic methods literature where it is traditional that the matrix G refers to those sample paths starting at level n with level $n-1$ taboo, we use this second form of $\psi(\mathcal{B}; \boldsymbol{a})$. If we had chosen to continue with the first form of $\psi(\mathcal{B}; \boldsymbol{a})$ then the matrix G that we derive would refer to those sample paths starting at level $n+1$ with level n taboo. This is a purely notational change for convenience sake.

It is useful also to consider the process $\bar{\boldsymbol{B}}(t)$ which evolves continuously like $\boldsymbol{B}(t)$ according to the usual differential equation (4) with successive interruptions occurring with

intensity $\bar{\boldsymbol{B}}(t)A_0\boldsymbol{e}$. The value of this process immediately after an interruption is the expected value of $\boldsymbol{B}(t)$ after an interruption from state $\boldsymbol{B}(t) = \bar{\boldsymbol{B}}(t)$.

We now show that this process is a terminating RAP with events corresponding to successive returns to level n from excursions to higher levels. To do this we define, for $k \geq 1$, the vector valued functions $\Psi_k(\boldsymbol{a}) : \mathcal{A}_+ \to \mathcal{A}_+$ such that $\Psi_k(\boldsymbol{a})$ is the expected value of the phase process $\boldsymbol{A}(\cdot)$ on the first visit to level $n-1$, without visiting level $n+k$ or above, conditioned on the starting state $X(0) = (n, \boldsymbol{a})$. In short

$$\Psi_k(\boldsymbol{a}) = \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1})I(\tau_{n-1} < \tau_{n+k})|X(0) = (n, \boldsymbol{a})\right].$$

Correspondingly we define

$$\Psi(\boldsymbol{a}) = \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1})I(\tau_{n-1} < \infty)|X(0) = (n, \boldsymbol{a})\right] = \int_{\mathcal{A}_+} \boldsymbol{b}\psi(\mathrm{d}\boldsymbol{b}; \boldsymbol{a}).$$

The process $\bar{\boldsymbol{B}}(t)$ is a RAP if $\Psi(\boldsymbol{a}) = \boldsymbol{a}G$ for some matrix G, and $\boldsymbol{a}G \in \mathcal{A}_0 \cap \mathcal{A}_+$, for all $\boldsymbol{a} \in \mathcal{A}_+$. The following arguments show that this is indeed true.

Lemma 1 *For $k \geq 1$, the vector valued functions $\Psi_k(\boldsymbol{a})$ are linear, that is, $\Psi_k(\boldsymbol{a}) = \boldsymbol{a}G_k$ for some matrix G_k.*

Proof: The lemma is proved by induction.

First let's calculate $\Psi_1(\boldsymbol{a})$. The probability that the process spends time t in level n before first leaving, given it starts in $(n, \boldsymbol{a})$, is $\boldsymbol{a}e^{A_1 t}\boldsymbol{e}$, in which case the state will then be $\left(n, \frac{\boldsymbol{a}e^{A_1 t}}{\boldsymbol{a}e^{A_1 t}\boldsymbol{e}}\right)$. The conditional probability $P(\tau_{n-1} \in (t, t + \mathrm{d}t)|\tau_{n-1}, \tau_{n+1} > t)$ is therefore $\frac{\boldsymbol{a}e^{A_1 t}}{\boldsymbol{a}e^{A_1 t}\boldsymbol{e}}A_2\boldsymbol{e}\mathrm{d}t$. Then,

$$\mathbb{E}\left[\boldsymbol{A}(\tau_{n-1})|t = \tau_{n-1} < \tau_{n+1}, X(0) = (n, \boldsymbol{a})\right] = \frac{\dfrac{\boldsymbol{a}e^{A_1 t}}{\boldsymbol{a}e^{A_1 t}\boldsymbol{e}}A_2}{\dfrac{\boldsymbol{a}e^{A_1 t}}{\boldsymbol{a}e^{A_1 t}\boldsymbol{e}}A_2\boldsymbol{e}} = \frac{\boldsymbol{a}e^{A_1 t}A_2}{\boldsymbol{a}e^{A_1 t}A_2\boldsymbol{e}}.$$

Finally,

$$\begin{aligned}
\Psi_1(\boldsymbol{a}) &= \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1})I(\tau_{n-1} < \tau_{n+1})|X(0) = (n, \boldsymbol{a})\right], \\
&= \int_0^\infty \boldsymbol{a}e^{A_1 t}\boldsymbol{e}\frac{\boldsymbol{a}e^{A_1 t}A_2\boldsymbol{e}}{\boldsymbol{a}e^{A_1 t}\boldsymbol{e}}\frac{\boldsymbol{a}e^{A_1 t}A_2}{\boldsymbol{a}e^{A_1 t}A_2\boldsymbol{e}}\mathrm{d}t, \\
&= \int_0^\infty \boldsymbol{a}e^{A_1 t}A_2\mathrm{d}t, \\
&= \boldsymbol{a}(-A_1)^{-1}A_2 = \boldsymbol{a}G_1,
\end{aligned}$$

and thus is linear.

We now introduce, for $k \geq 1$, the measures associated with $\Psi_k(\boldsymbol{a})$,

$$\psi_k(\mathcal{B}; \boldsymbol{a}) = \mathbb{P}(\boldsymbol{A}(\tau_{n-1}) \in \mathcal{B} \ \& \ \tau_{n-1} < \tau_{n+k} | X(0) = (n, \boldsymbol{a})), \qquad \mathcal{B} \subset \mathcal{A}_+,$$

so that

$$\Psi_k(\boldsymbol{a}) = \int_{\mathcal{A}_+} \boldsymbol{b} \psi_k(\mathrm{d}\boldsymbol{b}; \boldsymbol{a}).$$

For all $k \geq 1$, we also introduce for all $\ell \geq 0$,

$$\Psi_{k,\ell}(\boldsymbol{a}) = \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1}) I(\tau_{n-1} < \tau_{n+k}, \text{exactly } \ell \text{ visits above level } n) | X(0) = (n, \boldsymbol{a})\right]$$

and note that

$$\Psi_k(\boldsymbol{a}) = \sum_{\ell=0}^{\infty} \Psi_{k,\ell}(\boldsymbol{a}).$$

Considering the first entry to level $n+1$ occurring at time t and the first entry to level $n-1$ occurring at local time $t+u$, in level n, we get

$$\begin{aligned}
\Psi_{k,1}(\boldsymbol{a}) &= \int_{t=0}^{\infty} \int_{\mathcal{A}_+} \int_{u=0}^{\infty} \boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e} \boldsymbol{b} e^{A_1 u} A_2 \boldsymbol{e} \frac{\boldsymbol{b} e^{A_1 u} A_2}{\boldsymbol{b} e^{A_1 u} A_2 \boldsymbol{e}} \mathrm{d}u \psi_{k-1}\left(\mathrm{d}\boldsymbol{b}, \frac{\boldsymbol{a} e^{A_1 t} A_0}{\boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e}}\right) \mathrm{d}t, \\
&= \int_{t=0}^{\infty} \boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e} \int_{\mathcal{A}_+} \boldsymbol{b} \psi_{k-1}\left(\mathrm{d}\boldsymbol{b}, \frac{\boldsymbol{a} e^{A_1 t} A_0}{\boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e}}\right) (-A_1)^{-1} A_2 \mathrm{d}t, \\
&= \int_{t=0}^{\infty} \boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e} \Psi_{k-1}\left(\frac{\boldsymbol{a} e^{A_1 t} A_0}{\boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e}}\right) (-A_1)^{-1} A_2 \mathrm{d}t.
\end{aligned}$$

Assuming the linearity of $\Psi_{k-1}(\boldsymbol{b}) = \boldsymbol{b} G_{k-1}$ for all $\boldsymbol{b} \in \mathcal{A}_+$, we have

$$\begin{aligned}
\Psi_{k,1}(\boldsymbol{a}) &= \int_{t=0}^{\infty} \boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e} \left(\frac{\boldsymbol{a} e^{A_1 t} A_0}{\boldsymbol{a} e^{A_1 t} A_0 \boldsymbol{e}}\right) G_{k-1} (-A_1)^{-1} A_2 \mathrm{d}t, \\
&= \boldsymbol{a} (-A_1)^{-1} A_0 G_{k-1} (-A_1)^{-1} A_2.
\end{aligned}$$

Since the process is Markovian, linearity of the first visit implies linearity for any fixed number ℓ of visits above n. Alternatively we could carry out the above calculations for any fixed $\ell > 1$, to achieve

$$\Psi_{k,\ell}(\boldsymbol{a}) = \boldsymbol{a} \left[(-A_1)^{-1} A_0 G_{k-1}\right]^{\ell} (-A_1)^{-1} A_2, \qquad \ell \geq 0.$$

Thus all the functions $\Psi_{k,\ell}(\boldsymbol{a})$ are linear and so, since $\Psi_k(\boldsymbol{a})$ is the expected value of a random variable with finite expectation,

$$\begin{aligned}
\Psi_k(\boldsymbol{a}) &= \sum_{\ell=0}^{\infty} \Psi_{k,\ell}(\boldsymbol{a}), \\
&= \sum_{\ell=0}^{\infty} \boldsymbol{a} \left[(-A_1)^{-1} A_0 G_{k-1}\right]^{\ell} (-A_1)^{-1} A_2, \\
&= \boldsymbol{a} G_k,
\end{aligned}$$

and is thus linear. Further, since $\mathcal{A}_0 \cap \mathcal{A}_+$ is a compact and convex set, $\boldsymbol{a}G \in \mathcal{A}_0 \cap \mathcal{A}_+$, for all $\boldsymbol{a} \in \mathcal{A}_+$. ■

Theorem 2 *For all $\boldsymbol{a} \in \mathcal{A}_+$, we have*

$$\Psi(\boldsymbol{a}) = \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1}) I(\tau_{n-1} < \infty) | X(0) = (n, \boldsymbol{a})\right] = \boldsymbol{a}G.$$

for some matrix G.

Proof: Clearly,

$$\Psi(\boldsymbol{a}) = \Psi_k(\boldsymbol{a}) + P(\tau_{n-1} \geq \tau_{n+k} | X(0) = (n, \boldsymbol{a})) \Psi_k^{\star}(\boldsymbol{a}),$$

where $\Psi_k^{\star}(\cdot)$ is some bounded function on the domain $\mathcal{A}_+$. Since $\mathrm{P}(\tau_{n-1} \geq \tau_{n+k} | X(0) = (n, \boldsymbol{a})) \to 0$ as $k \to \infty$, Lemma 1 implies that $\Psi(\boldsymbol{a}) = \boldsymbol{a}G$ for some matrix G. ■

It is immediately obvious that Theorem 2 implies that

$$\begin{aligned}
\boldsymbol{a}G\boldsymbol{e} &= \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1}) \boldsymbol{e} I(\tau_{n-1} < \infty) | X(0) = (n, \boldsymbol{a})\right], \\
&= \mathbb{E}\left[I(\tau_{n-1} < \infty) | X(0) = (n, \boldsymbol{a})\right], \\
&= \mathbb{P}\left[\tau_{n-1} < \infty | X(0) = (n, \boldsymbol{a})\right], \text{ and} \\
\frac{\boldsymbol{a}G}{\boldsymbol{a}G\boldsymbol{e}} &= \mathbb{E}\left[\boldsymbol{A}(\tau_{n-1}) | X(0) = (n, \boldsymbol{a}) \ \& \ \tau_{n-1} < \infty\right].
\end{aligned}$$

In words, we have that, given $X(0) = (n, \boldsymbol{a})$,

1. $\boldsymbol{a}G\boldsymbol{e}$ represents the probability that the process will ever reach level $n-1$, and
2. $\dfrac{\boldsymbol{a}G}{\boldsymbol{a}G\boldsymbol{e}}$ represents the expected state of the process $\boldsymbol{A}(\cdot)$ on the first visit to level $n-1$, given that the process ever visits level $n-1$.

Remark 1 *It is clear that the process is recurrent, if and only if* $\boldsymbol{a}G\boldsymbol{e} = 1$ *for all* $\boldsymbol{a} \in \mathcal{A}_+$.

The linearity result of Theorem 2 paves the way for the following important corollary.

Corollary 3 *The expected value of the processes* $\boldsymbol{B}(t)$ *and* $\bar{\boldsymbol{B}}(t)$ *are identical, that is,*

$$\mathbb{E}(\boldsymbol{B}(t)|X(0) = (n, \boldsymbol{a})) = \mathbb{E}(\bar{\boldsymbol{B}}(t)|X(0) = (n, \boldsymbol{a})), \qquad t \geq 0.$$

Proof: We prove the result by considering the number L of intermediate visits to higher levels. We show, for all $\ell \geq 0$, that

$$\begin{aligned}
&\mathbb{E}(\boldsymbol{B}(t)I(L = \ell)|X(0) = (n, \boldsymbol{b}_0)) \equiv \\
&\quad \int_{u_1=0}^{t} \int_{u_2=0}^{t-u_1} \cdots \int_{u_\ell=0}^{t-\sum_{i=1}^{\ell-1} u_i} \int_{\mathcal{A}_+} \cdots \int_{\mathcal{A}_+} \left[\prod_{i=1}^{\ell} \boldsymbol{b}_{i-1} e^{A_1 u_i} A_0 \boldsymbol{e} \right] \boldsymbol{b}_\ell e^{A_1 (t - \sum_{i=1}^{\ell} u_i)} \\
&\qquad\qquad \psi\left(\mathrm{d}\boldsymbol{b}_\ell, \frac{\boldsymbol{b}_{\ell-1} e^{A_1 u_\ell} A_0}{\boldsymbol{b}_{\ell-1} e^{A_1 u_\ell} A_0 \boldsymbol{e}} \right) \cdots \psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1 u_1} A_0}{\boldsymbol{b}_0 e^{A_1 u_1} A_0 \boldsymbol{e}} \right) \mathrm{d}u_\ell \cdots \mathrm{d}u_2 \mathrm{d}u_1, \\
&= \int_{u_1=0}^{t} \int_{u_2=0}^{t-u_1} \cdots \int_{u_\ell=0}^{t-\sum_{i=1}^{\ell-1} u_i} \boldsymbol{b}_0 \left[\prod_{i=1}^{\ell} e^{A_1 u_i} A_0 G \right] e^{A_1 (t - \sum_{i=1}^{\ell} u_i)} \mathrm{d}u_\ell \cdots \mathrm{d}u_2 \mathrm{d}u_1, \\
&\equiv \mathbb{E}(\bar{\boldsymbol{B}}(t)I(L = \ell)|X(0) = (n, \boldsymbol{b}_0)).
\end{aligned}$$

This is then sufficient to prove the desired result. The details of this argument are more tedious than insightful and so we have placed them in Appendix A. ■

Theorem 4 *The matrix*

$$U = A_1 + A_0 G \tag{5}$$

is a $p \times p$ *matrix that, for all* $t \in \mathbb{R}_+$ *and for all* $\boldsymbol{a} \in \mathcal{A}_+$, *satisfies the following expression*

$$\mathbb{E}\left[\boldsymbol{A}(g_n(t)) I\left(g_n(t) < \tau_{n-1}\right) | X(0) = (n, \boldsymbol{a})\right] = \boldsymbol{a} e^{Ut}. \tag{6}$$

Proof: It follows directly from Theorem 2 that the process $\bar{\boldsymbol{B}}(t)$ is a terminating RAP. Between epochs of visits to levels $n+1$ and above, the phase evolves at level n according to the differential equation (4) with matrix $C = A_1$. Excursions to level $n+1$ and above occur from $\boldsymbol{a}$ with intensity $\boldsymbol{a}A_0\boldsymbol{e}$. Depending on the recurrence or transience of the process, all or some of these excursions will return to level n. The intensity with which excursions to level $n+1$ and above that return to level n occur from $\boldsymbol{a}$ is $\boldsymbol{a}A_0G\boldsymbol{e}$. If such an excursion occurs, then the phase process returns to level n with expected state $\frac{\boldsymbol{a}A_0G}{\boldsymbol{a}A_0G\boldsymbol{e}}$. Thus the

right-hand side of (6) governs the phase process of $\bar{\boldsymbol{B}}(t)$ when initiated in $\boldsymbol{a}$. But, by Corollary 3, this is precisely the expected value of $\boldsymbol{B}(t)$, which is the left-hand side of (6). ■

Incidentally, it is immediately obvious that Theorem 4 implies that

$$\begin{aligned} \boldsymbol{a}e^{Ut}\boldsymbol{e} &= \mathbb{P}\left[g_n(t) < \tau_{n-1} | X(0) = (n, \boldsymbol{a})\right], \text{ and} \\ \frac{\boldsymbol{a}e^{Ut}}{\boldsymbol{a}e^{Ut}\boldsymbol{e}} &= \mathbb{E}\left[\boldsymbol{A}(g_n(t)) | X(0) = (n, \boldsymbol{a}) \ \& \ g_n(t) < \tau_{n-1}\right]. \end{aligned}$$

In words, we have, given $X(0) = (n, \boldsymbol{a})$, that

1. $\boldsymbol{a}e^{Ut}\boldsymbol{e}$ represents the probability that the process has spent at least t units of local time in level n before it is absorbed into the taboo level $n-1$, and
2. $\dfrac{\boldsymbol{a}e^{Ut}}{\boldsymbol{a}e^{Ut}\boldsymbol{e}}$ represents the expected state of the process $\boldsymbol{A}(\cdot)$ when the local time first reaches t units, given that the process spends at least t units of local time in level n.

This matrix U has a very similar interpretation to that of Q in Proposition 2.1 of Asmussen and Bladt [1]. The only differences between Q and U are

1. the concept of local time which is a direct consequence of the fact that we have embedded the RAP in a more complicated setting, and
2. the existence of the taboo level $n-1$, which means that there is a non-zero probability that the process may jump down to the taboo level $n-1$ and thus is terminating. This is analogous to U being a **non-conservative** generator in the traditional QBD case.

We are now ready to prove the following traditional QBD result in the more general setting of a QBD with RAP components.

Theorem 5 *The matrix G is given by*

$$G = (-U)^{-1} A_2. \tag{7}$$

Proof: The density of having a transition to level $n-1$ at local time t, when $X_0 = (n, \boldsymbol{a})$ is

$$\boldsymbol{a}e^{Ut}A_2\boldsymbol{e}.$$

The expected value of the phase process immediately after the first such transition occurring at local time t is

$$\mathbb{E}\left[\boldsymbol{A}(g_n(t))|X_0 = (n, \boldsymbol{a}) \ \& \ \tau_{n-1} = g_n(t)\right] = \boldsymbol{a}e^{Ut}A_2/\boldsymbol{a}e^{Ut}A_2\boldsymbol{e}.$$

Thus

$$\begin{aligned} \boldsymbol{a}G &= \mathbb{E}\left[\boldsymbol{A}(g_n(\tau_{n-1}))I(\tau_{n-1} < \infty)|X_0 = (n, \boldsymbol{a})\right], \\ &= \int_{t=0}^{\infty} \frac{\boldsymbol{a}e^{Ut}A_2}{\boldsymbol{a}e^{Ut}A_2\boldsymbol{e}}\boldsymbol{a}e^{Ut}A_2\boldsymbol{e}\ \mathrm{d}t = \int_{t=0}^{\infty} \boldsymbol{a}e^{Ut}A_2\ \mathrm{d}t = \boldsymbol{a}(-U)^{-1}A_2. \end{aligned}$$

Now, both G and $(-U)^{-1}A_2$ have the same physical interpretation for all $\boldsymbol{a} \in \mathcal{A}_+$ and hence we may take $G = (-U)^{-1}A_2$. ■

Corollary 6 *The matrix G is a solution to the equation*

$$A_2 + A_1G + A_0G^2 = 0. \tag{8}$$

Proof: Theorem 4 gives that $U = A_1 + A_0G$ and Theorem 5 shows $G = (-U)^{-1}A_2$. Exactly as in the traditional QBD environment, these two equations can be combined to give equation (8). ■

At this point we have developed all the necessary machinery. We are now ready to state and prove our main result.

Consider the censored process at level 0, with no levels taboo, which is denoted by $\{\boldsymbol{C}(t)\}_{t\geq 0}$ and define the corresponding RAP $\left\{\bar{\boldsymbol{C}}(t)\right\}_{t\geq 0}$ with a similar interpretation to that of $\left\{\bar{\boldsymbol{B}}(t)\right\}_{t\geq 0}$.

Theorem 7 *Assume that $X(\cdot)$ is a positive recurrent Markov process.*

1. *Let the vectors $\boldsymbol{\pi}_n$ denote $\lim_{t\to\infty}\mathbb{E}\left[\boldsymbol{A}(t)I(N(t) = n)|X(0) = (m, \boldsymbol{b})\right]$ $= \mathbb{E}\left[\boldsymbol{A}(t)I(N(t) = n)|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0) = m)\right) = \boldsymbol{\pi}_m,\ m = 0, 1, 2, \ldots\right]$ then*

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1}R \qquad \textit{for all } n \geq 1,$$

   with

$$R = A_0(-U)^{-1}. \tag{9}$$

2. *The matrix $V = B_1 + RA_2$ is the matrix that drives the evolution of $\bar{\boldsymbol{C}}(\cdot)$.*

3. *Let the vector $\hat{\boldsymbol{\pi}}_0$ be chosen so that it satisfies*

$$\boldsymbol{x}\left[B_1 + RA_2\right] = \mathbf{0}, \tag{10}$$

but is normalised according to $\hat{\boldsymbol{\pi}}_0\boldsymbol{e} = 1$. Then $\hat{\boldsymbol{\pi}}_0$ is such that, for all $t \geq 0$,

$$\hat{\boldsymbol{\pi}}_0 \;=\; \mathbb{E}\left[\boldsymbol{C}(t)|\mathbb{E}\left(\boldsymbol{C}(0)\right) = \hat{\boldsymbol{\pi}}_0\right].$$

That is, $\hat{\boldsymbol{\pi}}_0$ is the expected state of the stationary distribution of $\boldsymbol{C}(\cdot)$.

4. *Finally, $\boldsymbol{\pi}_0 = K \cdot \hat{\boldsymbol{\pi}}_{\mathbf{0}}$ where*

$$K = \mathbb{P}\left[N(t) = 0|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0) = m)\right) = \boldsymbol{\pi}_m,\ m = 0, 1, 2, \ldots\right],$$

and

$$\sum_{k=0}^{\infty} \boldsymbol{\pi}_0 R^k \boldsymbol{e} = \boldsymbol{\pi}_0 \left(I - R\right)^{-1} \boldsymbol{e} = 1. \tag{11}$$

Proof:

1. The existence, for all $(m, \boldsymbol{b}) \in \{(\{0\} \times \mathcal{A}_0) \cup (\mathbb{N}_+ \times \mathcal{A}_2)\}$, of

$$\begin{aligned}\boldsymbol{\pi}_n &= \lim_{t\to\infty} \mathbb{E}\left[\boldsymbol{A}(t)I(N(t) = n)|X(0) = (m, \boldsymbol{b})\right], \\ &= \mathbb{E}\left[\boldsymbol{A}(t)I(N(t) = n)|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0) = m)\right) = \boldsymbol{\pi}_m,\ m = 0, 1, 2, \ldots\right]\end{aligned}$$

follows from the positive recurrence of the Markov process $X(\cdot)$.

To show the form of $\boldsymbol{\pi}_n$ we closely follow the approach introduced by Ramaswami [13] in the traditional QBD environment with a few crucial changes. Actually, this approach shows that the extension of the classical framework goes beyond QBDs to the general case of processes of GI/M/1 type with RAP components. Consider the state of the RAP at time t. By conditioning on the last entry to level n from below, we obtain

$$\begin{aligned}&\mathbb{E}\left[\boldsymbol{A}(t)I\left(N(t) = n\right)|X(0) = (0, \boldsymbol{a})\right] \\ &= \int_0^t \int_{\mathcal{A}_+} P\left(X(t-x) \in (n-1, \mathrm{d}\boldsymbol{b})|X(0) = (0, \boldsymbol{a})\right) \boldsymbol{b}A_0\boldsymbol{e} \\ &\quad \times \mathbb{E}\left(A(x)I(N(x) = n, x < \tau_{n-1})|X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right)\right) \mathrm{d}x.\end{aligned}$$

From Theorem 4 we have

$$\mathbb{E}\left[A(t)I(N(t)=n,t<\tau_{n-1})|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(t)=u\right]$$
$$=\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}e^{Uu}P\left(N(t)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(t)=u\right).$$

Using this result we get

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{A}(t)I\left(N(t)=n\right)|X(0)=(0,\boldsymbol{a})\right]\\
&=\int_0^t\int_{\mathcal{A}_+}P\left(X(t-x)\in(n-1,\mathrm{d}\boldsymbol{b})|X(0)=(0,\boldsymbol{a})\right)\boldsymbol{b}A_0\boldsymbol{e}\\
&\quad\times\mathbb{E}_{\ell_n(x)}\left[\mathbb{E}\left(A(x)I(N(x)=n,x<\tau_{n-1})|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right)\right)|\ell_n(x)\right]\mathrm{d}x,\\
&=\int_0^t\int_{\mathcal{A}_+}P\left(X(t-x)\in(n-1,\mathrm{d}\boldsymbol{b}|X(0)=(0,\boldsymbol{a}))\right)\\
&\quad\times\boldsymbol{b}A_0\boldsymbol{e}\mathbb{E}_{\ell_n(x)}\left[\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}e^{U\ell_n(x)}P\left(N(x)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(x)\right)\right]\mathrm{d}x,\\
&=\int_0^t\int_{\mathcal{A}_+}P\left(X(t-x)\in(n-1,\mathrm{d}\boldsymbol{b})|X(0)=(0,\boldsymbol{a})\right)\boldsymbol{b}\\
&\quad\times A_0\mathbb{E}_{\ell_n(x)}\left[e^{U\ell_n(x)}P\left(N(x)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(x)\right)\right]\mathrm{d}x.
\end{aligned}$$

Taking limits as $t\to\infty$ on both sides, gives

$$\begin{aligned}
\boldsymbol{\pi}_n=&\int_{\mathcal{A}_+}\lim_{t\to\infty}\int_0^\infty P\left(X(t-x)\in(n-1,\mathrm{d}\boldsymbol{b})|X(0)=(0,\boldsymbol{a})\right)\boldsymbol{b}\\
&\times A_0\mathbb{E}_{\ell_n(x)}\left[e^{U\ell_n(x)}P\left(N(x)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(x)\right)\right]\mathrm{d}x\quad.
\end{aligned}$$

Now using similar arguments to those in [13, Page 263], we obtain

$$\begin{aligned}
\boldsymbol{\pi}_n\ =\ &\int_{\mathcal{A}_+}\left(\left[\lim_{t\to\infty}P\left(X(t)\in(n-1,\mathrm{d}\boldsymbol{b})|X(0)=(0,\boldsymbol{a})\right)\boldsymbol{b}\right]\right.\\
&\left.\times A_0\int_0^\infty\mathbb{E}_{\ell_n(x)}\left[e^{U\ell_n(x)}P\left(N(x)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(x)\right)\right]\mathrm{d}x\right)\quad.
\end{aligned}\tag{12}$$

The idea of the following argument is simple. Local time increases with rate 1, whenever the process is in level n and with rate 0 otherwise. Thus it seems natural and obvious that the following claim holds true,

$$\int_0^\infty\mathbb{E}_{\ell_n(x)}\left[e^{U\ell_n(x)}P\left(N(x)=n|X(0)=\left(n,\frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right),\ell_n(x)\right)\right]\mathrm{d}x=\int_0^\infty e^{Uu}\mathrm{d}u.$$

To formally prove this claim, first note that the events $\{\ell_n(x) \leq u\}$ and $\{g_n(u) \geq x\}$ are identical, since any sample path that belongs to the former belongs to the latter and vice versa. Introducing $L_x(u) = P\left(\ell_n(x) \leq u | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right)\right)$ and $G_u(x) = P\left(g_n(u) \geq x | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right)\right)$ one sees that $L_x(u) = G_u(x)$.

Now

$$\begin{aligned}
&\int_0^\infty \mathbb{E}_{\ell_n(x)}\left[e^{U\ell_n(x)} P\left(N(x) = n | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right), \ell_n(x)\right)\right] \mathrm{d}x \\
&= \int_{x=0}^\infty \int_{u=0}^x e^{Uu} P\left(N(x) = n | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right), \ell_n(x) = u\right) \mathrm{d}L_x(u)\mathrm{d}x.
\end{aligned}$$

We can now change the order of integration while using the relationship between $L_x(u)$ and $G_u(x)$ to get

$$\int_{u=0}^\infty e^{Uu} \int_{x=u}^\infty P\left(N(x) = n | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right), \ell_n(x) = u\right) \mathrm{d}G_u(x)\mathrm{d}u.$$

The integral

$$\int_{x=u}^\infty P\left(N(x) = n | X(0) = \left(n, \frac{\boldsymbol{b}A_0}{\boldsymbol{b}A_0\boldsymbol{e}}\right), \ell_n(x) = u\right) \mathrm{d}G_u(x)$$

is the probability that the process ever visits level n again, given that the local time spent at level n is u. This probability is one according to the positive recurrence of the Markov chain, and so the claim is true. Therefore, equation (12) becomes

$$\begin{aligned}
\boldsymbol{\pi}_n &= \int_{\mathcal{A}_+} \left[\lim_{t\to\infty} \boldsymbol{b}P\left(X(t-x) \in (n-1, \mathrm{d}\boldsymbol{b}) | X(0) = (0, \boldsymbol{a})\right)\right] \int_0^\infty A_0 e^{Uu}\mathrm{d}x, \\
&= \boldsymbol{\pi}_{n-1} \int_0^\infty A_0 e^{Uu}\mathrm{d}x, \\
&= \boldsymbol{\pi}_{n-1} A_0 (-U)^{-1}, \\
&= \boldsymbol{\pi}_{n-1} R,
\end{aligned}$$

by equation (9).

2. It is clear that the proof of Theorem 4 can be simply reworked to show that the matrix $V = B_1 + A_0 G$ is the matrix that drives the evolution of $\bar{\boldsymbol{C}}(\cdot)$ with evolution within level 0 driven by B_1 and jumps (that is, transitions that first go to level 1) driven by $A_0 G$. From equation (7) and (9) we have that V can also be written as $V = B_1 + RA_2$.

Since this process is merely a censored part of the positive recurrent process, with no taboo states, it must be positive recurrent and so equation (10) must have a solution.

3. Let $\hat{\boldsymbol{\pi}}_0$ be chosen so that it satisfies equation (10) and is normalised according to $\hat{\boldsymbol{\pi}}_0\boldsymbol{e} = 1$. Then Corollary 3 and equation (3) imply, for all $t \geq 0$, that

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{C}(t)|\mathbb{E}(\boldsymbol{C}(0)) = \hat{\boldsymbol{\pi}}_0\right] &= \mathbb{E}\left[\bar{\boldsymbol{C}}(t)|\mathbb{E}(\boldsymbol{C}(0)) = \hat{\boldsymbol{\pi}}_0\right], \\ &= \hat{\boldsymbol{\pi}}_0 e^{Vt}, \\ &= \hat{\boldsymbol{\pi}}_0.\end{aligned}$$

4. Since $C(\cdot)$ is the version of $X(\cdot)$, censored to be on level 0, we have that

$$\frac{\mathbb{E}\left[\boldsymbol{A}(t)I(N(t)=0)|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0)=m)\right) = \boldsymbol{\pi}_m, \quad m = 0,1,\ldots\right]}{\mathbb{E}\left[I(N(t)=0)|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0)=m)\right) = \boldsymbol{\pi}_m, \quad m = 0,1,\ldots\right]} = \hat{\boldsymbol{\pi}}_0, \quad \forall t \geq 0.$$

The denominator $\mathbb{E}\left[I(N(t)=0)|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0)=m)\right) = \boldsymbol{\pi}_m, \quad m = 0,1,\ldots\right]$ is equal to $\mathbb{P}\left[N(t)=0|\mathbb{E}\left(\boldsymbol{A}(0)I(N(0)=m)\right) = \boldsymbol{\pi}_m, \quad m = 0,1,\ldots\right]$, and the numerator is $\boldsymbol{\pi}_0$.

Since the process is recurrent we have

$$\sum_{i=0}^{\infty} \boldsymbol{\pi}_n \boldsymbol{e} = 1 = \boldsymbol{\pi}_0 (I - R)^{-1}\boldsymbol{e}.$$

■

Corollary 8 *The process is positive recurrent if and only if*

(i) the dominant eigenvalue of R, $Sp(R)$, is such that $Sp(R) < 1$, and

(ii) equation (10) has a solution.

Proof: If the process is positive recurrent, then as argued in the proof of Theorem 7(ii), equation (10) must have a solution. From Theorem 7(iv) we also know that $Sp(R) < 1$.

For the converse, we know that the invariant measure of the form described in Theorem 7 exists and is summable and so the process must be positive recurrent. ■

Remark 2 *It is also of interest to evaluate the marginal distribution that the queue has k customers in the queue, and this is given by*

$$m_k = \boldsymbol{\pi}_k \boldsymbol{e} = \boldsymbol{\pi}_0 R^k \boldsymbol{e}, \quad k \geq 0. \tag{13}$$

5 The example process

In the remainder of the paper we shall not pursue the full generality of a QBD with RAP components, but instead move to a simple concrete example, that of Section 3. That small example has all the ingredients needed. The arrival process and sequence of service times are driven from the same phase space thus providing dependence between the two. Further the sojourn time at each level are matrix–exponentially distributed, with a distribution that is not in PH. We repeat the matrices defining this queue:

$$A_0 = \gamma D_a = \gamma \begin{bmatrix} \frac{14}{5} & -\frac{9}{10} & -\frac{9}{10} \\ \frac{26}{15} & -\frac{8}{15} & -\frac{8}{15} \\ \frac{58}{15} & -\frac{19}{15} & -\frac{19}{15} \end{bmatrix} \quad ,$$

$$A_1 = \begin{bmatrix} -1 & 0 & 0 \\ -\frac{2}{3} & -1 & 1 \\ \frac{2}{3} & -1 & -1 \end{bmatrix} \quad ,$$

$$A_2 = (1-\gamma)\left(1, \frac{2}{3}, \frac{4}{3}\right)^T (3, -1, -1) \quad ,$$

$B_0 = A_0$, and $B_1 = \gamma A_1$.

The key to a solution of this problem is that A_2 is rank one. We can thus apply Theorem 8.5.1 of [6] directly as the proof of that theorem is purely analytical relying only on equation (7) and Remark 1 to show that $G = \boldsymbol{e}(3, -1, -1)$. We can then use equation (5) to show that

$$U = A_1 + A_0 G = \begin{bmatrix} -1+3\gamma & -\gamma & -\gamma \\ \frac{-2}{3} + 2\gamma & -1 - \frac{2}{3}\gamma & 1 - \frac{2}{3}\gamma \\ \frac{2}{3} + 4\gamma & -1 - \frac{4}{3}\gamma & -1 - \frac{4}{3}\gamma \end{bmatrix} \quad ,$$

and equation (9) to find

$$\begin{aligned} R &= \gamma \begin{bmatrix} \frac{14}{5} & -\frac{9}{10} & -\frac{9}{10} \\ \frac{26}{15} & -\frac{8}{15} & -\frac{8}{15} \\ \frac{58}{15} & -\frac{19}{15} & -\frac{19}{15} \end{bmatrix} \left(\begin{bmatrix} -1+3\gamma & -\gamma & -\gamma \\ \frac{-2}{3} + 2\gamma & -1 - \frac{2}{3}\gamma & 1 - \frac{2}{3}\gamma \\ \frac{2}{3} + 4\gamma & -1 - \frac{4}{3}\gamma & -1 - \frac{4}{3}\gamma \end{bmatrix} \right)^{-1} \quad , \\ &= \begin{bmatrix} -1/15\, \frac{\gamma\,(33+2\,\gamma)}{\gamma-1} & 0 & 1/10\, \frac{\gamma\,(\gamma+9)}{\gamma-1} \\ -\frac{2}{45}\, \frac{\gamma\,(31+4\,\gamma)}{\gamma-1} & 0 & 2/15\, \frac{\gamma\,(\gamma+4)}{\gamma-1} \\ -\frac{4}{45}\, \frac{\gamma\,(34+\gamma)}{\gamma-1} & 0 & 1/15\, \frac{\gamma\,(\gamma+19)}{\gamma-1} \end{bmatrix} \quad . \end{aligned}$$

Now, the eigenvalues of R are $(0, -\gamma/15, \gamma/(1-\gamma)$. This is important as it allows us to determine the stability of the queue by Corollary 8. We can easily see that the queue is stable whenever $\gamma < \frac{1}{2}$. Note, this is a mere coincedence of the parameter values rather than an intrinsic property of the construction.

To conclude this example, we present two results. First, we present a table giving the equilibrium distribution of the marginal queue length for $\gamma = 0.25$. In this table we present the analytic results from Theorem 7 and the sample mean and 95% confidence intervals for simulation results, where we ran 10 independent simulations each consisting of 10,000 level changes. Second, we present a graph presenting the mean queue length for the process as a function of γ.

Level	Analysis	Sample mean	95% Confidence Interval
0	0.6736	0.6580	(0.6419,0.6741)
1	0.2175	0.2275	(0.2192,0.2358)
2	0.0726	0.0758	(0.0709,0.0807)
3	0.0242	0.0259	(0.0220,0.0298)
4	0.0081	0.0085	(0.0060,0.0111)
5	0.0027	0.0028	(0.0014,0.0041)
6	0.0009	0.0010	(-0.0001,0.0021)
7	0.0003	0.0003	(-0.0002,0.0008)
8	0.0001	0.0001	(-0.0002,0.0004)

Table 1: Equilibrium distribution of the marginal queue length for $\gamma = 0.25$.

6 Conclusions

In this paper we have introduced the concept of a QBD with RAP components. We have extended the performance evaluation literature by providing an efficient analytic method for determining the equilibrium distribution for such a process. The efficiency comes from the fact that the complexity is determined by the dimension of the matrix exponential distribution and it is well known that some phase-type distributions have matrix exponential representations of much smaller dimension. The issue of justifying algorithms for the evaluation of the matrix G for such processes has not been undertaken,

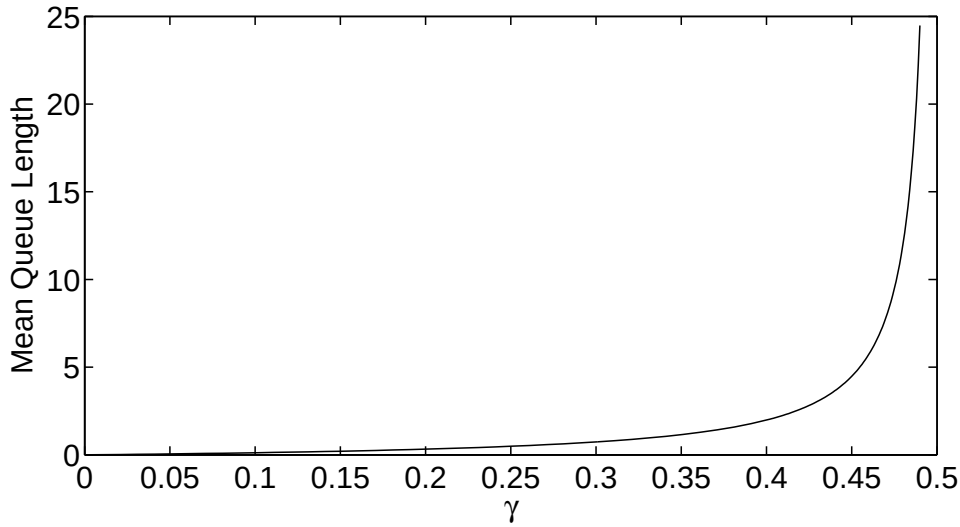

Figure 1: Figure of the Mean Queue Length against γ.

but has been delayed to a later paper [3]. The study of numerical stability is also important as there will be terms of opposing sign, unlike in the case of the traditional QBD.

We investigated a positive recurrent queue where the arrival process and sequence of service times were taken from different RAPs (outside the space of MAPs) that share the same phase space. This introduces cross-correlation between the arrival process and the sequence of service times. To our knowledge this is the first analysis of a queue with RAP components that is outside the domain of traditional matrix-analytic methods.

To avoid the issue of requiring algorithms for evaluating the matrix G, we deliberately chose the RAP representing the sequence of service times to have a rank-one jump matrix. This meant that the matrix G could be identified *a priori* for such a process.

The authors believe that the results in this paper are particularly important in the field of practical application where the arrival and service processes are fitted to observed data using some statistical technique. In this situation the attractiveness of the very simple physical interpretation of the MAP and the phase-type distribution is not missed as it was always artificial. This is in stark contrast to the situation where a MAP or phase-type distribution arises naturally from understanding the physical processes.

Further, the task of fitting phase-type distributions to data is awkward as the difficulty of representing a distribution efficiently by a phase-type distribution is always present. In contrast, now that it is possible to analyse QBDs with RAP components, this may

provide great impetus to the field of fitting data to matrix exponential distributions. In this direction, Fackrell [5] has made significant progress.

ACKNOWLEDGEMENTS

Nigel Bean would like to thank the University of Adelaide for assisting with the funding of the Special Studies Program that enabled him to visit the Technical University of Denmark and initiate the research contained in this paper. He would also like to acknowledge the financial support of the Australian Research Council through Discovery Project DP0209921. Bo Friis Nielsen would like to thank the Technical University of Denmark for providing the opportunity of sabbatical leave, the Technical Research Council of Denmark for the support under grant no. 26-02-0155, and the Department of Statistics, University of California at Berkeley, for the hospitality of hosting him during his sabbatical leave.

References

[1] S. Asmussen and M. Bladt. Point processes with finite-dimensional conditional probabilities. *Stochastic Processes and their Applications*, **82**:127-142, 1999.

[2] S. Asmussen and M. Bladt. Renewal Theory and Queueing Algorithms for Matrix-Exponential Distributions. In *Matrix-Analytic Methods in Stochastic Models, Lecture Notes in Pure and Applied Mathematics*, **183**:313–341, eds. S.R. Chakravarthy and A.S. Alfa, Marcel Dekker 1996.

[3] N.G. Bean and B.F. Nielsen. Analysis of Queues with Rational Arrival Process (RAP) Components – A General Approach. In preparation, 2007.

[4] M. Bladt and M.F. Neuts. Matrix-Exponential Distributions: Calculus and Interpretations via Flows. *Stochastic Models*, **19**:113–124, 2003.

[5] M. Fackrell. *Characterization of Matrix Exponential Distributions.* Ph. D. Thesis, Department of Applied Mathematics, The University of Adelaide, SA 5005, Australia. 2003.

[6] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic methods in Stochastic Modelling.* ASA-SIAM Series on Statistics and Applied Probability, Society for Industrial and Applied Mathematics, Philadelphia, USA, 1998.

[7] L. Lipsky *Queueing Theory: A Linear Algebra Approach.* Macmillan Publishing Company, New York, USA, 1992.

[8] L. Lipsky and Z. Fang. *Classification of functions with rational Laplace transforms.* In *Summer Computer Simulation Conference*, 20–25, Las Vegas, Nevada, 1986.

[9] L. Lipsky and V. Ramaswami. *A unique minimal representation of Coxian service centers.* Technical report, Department of Computer Science and Engineering, University of Nebraska, Lincoln, 1985.

[10] D.M. Lucantoni. *The BMAP/G/1 queue: A Tutorial.* In *Performance Evaluation of Computer and Communication Systems: Joint Tutorial Papers of Performance '93 and Sigmetrics '93*, 330–358, eds. L. Dantiello and R. Nelson, Springer Verlag 1993.

[11] K. Mitchell Constructing a correlated sequence of matrix exponentials with invariant first-order properties. *Operations Research Letters*, **28**:27–34, 2001.

[12] M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models.* The John Hopkins University Press, Baltimore, 1981.

[13] V. Ramaswami. A Tutorial Overview of Matrix Analytic Methods: with some Extensions & new Results In *Matrix-Analytic Methods in Stochastic Models, Lecture Notes in Pure and Applied Mathematics*, **183**:261-296, eds. S.R. Chakravarthy and A.S. Alfa, Marcel Dekker 1996.

A Details omitted from the proof of Corollary 3

The result is obvious for $L = 0$ since the processes $\boldsymbol{B}(t)$ and $\bar{\boldsymbol{B}}(t)$ are identical in this case. For $L = 1$ we get

$$
\begin{aligned}
&\mathbb{E}(\boldsymbol{B}(t)I(L=1)|X(0)=(n,\boldsymbol{b}_0)) \\
&\qquad \equiv \int_{u_1=0}^{t}\int_{\mathcal{A}_2} \boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}\boldsymbol{b}_1 e^{A_1(t-u_1)}\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)\mathrm{d}u_1, \\
&\qquad = \int_{u_1=0}^{t} \boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}\int_{\mathcal{A}_2}\boldsymbol{b}_1\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)e^{A_1(t-u_1)}\mathrm{d}u_1, \\
&\qquad = \int_{u_1=0}^{t} \boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}\frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}Ge^{A_1(t-u_1)}\mathrm{d}u_1, \\
&\qquad = \int_{u_1=0}^{t} \boldsymbol{b}_0 e^{A_1u_1}A_0Ge^{A_1(t-u_1)}\mathrm{d}u_1, \\
&\qquad \equiv \mathbb{E}(\bar{\boldsymbol{B}}(t)I(L=1)|X(0)=(n,\boldsymbol{b}_0)),
\end{aligned}
$$

where the third equation follows from Theorem 2.

We now use induction to conclude the argument.

$$
\begin{aligned}
&\mathbb{E}(\boldsymbol{B}(t)I(L=\ell+1)|X(0)=(n,\boldsymbol{b}_0)) \equiv \\
&\quad \int_{u_1=0}^{t}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-\sum_{i=1}^{\ell}u_i}\int_{\mathcal{A}_2}\cdots\int_{\mathcal{A}_2}\left[\prod_{i=1}^{\ell+1}\boldsymbol{b}_{i-1}e^{A_1u_i}A_0\boldsymbol{e}\right]\boldsymbol{b}_\ell e^{A_1(t-\sum_{i=1}^{\ell+1}u_i)} \\
&\qquad\qquad \psi\left(\mathrm{d}\boldsymbol{b}_{\ell+1}, \frac{\boldsymbol{b}_\ell e^{A_1u_{\ell+1}}A_0}{\boldsymbol{b}_\ell e^{A_1u_{\ell+1}}A_0\boldsymbol{e}}\right)\cdots\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)\mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\mathrm{d}u_1, \\
&= \int_{u_1=0}^{t}\int_{\mathcal{A}_2}\boldsymbol{b}_0e^{A_1u_1}A_0\boldsymbol{e}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-u_1-\sum_{i=2}^{\ell}u_i}\int_{\mathcal{A}_2}\cdots\int_{\mathcal{A}_2} \\
&\qquad\qquad \left[\prod_{i=2}^{\ell+1}\boldsymbol{b}_{i-1}e^{A_1u_i}A_0\boldsymbol{e}\right]\boldsymbol{b}_\ell e^{A_1(t-u_1-\sum_{i=2}^{\ell+1}u_i)} \\
&\quad \psi\left(\mathrm{d}\boldsymbol{b}_{\ell+1}, \frac{\boldsymbol{b}_\ell e^{A_1u_{\ell+1}}A_0}{\boldsymbol{b}_\ell e^{A_1u_{\ell+1}}A_0\boldsymbol{e}}\right)\cdots\psi\left(\mathrm{d}\boldsymbol{b}_2, \frac{\boldsymbol{b}_1 e^{A_1u_2}A_0}{\boldsymbol{b}_1 e^{A_1u_2}A_0\boldsymbol{e}}\right)\mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)\mathrm{d}u_1, \\
&= \int_{u_1=0}^{t}\int_{\mathcal{A}_2}\boldsymbol{b}_0e^{A_1u_1}A_0\boldsymbol{e}\mathbb{E}(\boldsymbol{B}(t-u_1)I(L=\ell)|X(0)=(n,\boldsymbol{b}_1))\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)\mathrm{d}u_1, \\
&= \int_{u_1=0}^{t}\int_{\mathcal{A}_2}\boldsymbol{b}_0e^{A_1u_1}A_0\boldsymbol{e}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-u_1-\sum_{i=2}^{\ell}u_i}\boldsymbol{b}_1\left[\prod_{i=2}^{\ell+1}e^{A_1u_i}A_0G\right]e^{A_1(t-u_1-\sum_{i=2}^{\ell+1}u_i)} \\
&\qquad\qquad \mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right)\mathrm{d}u_1,
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{u_1=0}^{t}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-\sum_{i=1}^{\ell}u_i} \boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}\int_{\mathcal{A}_2}\boldsymbol{b}_1\psi\left(\mathrm{d}\boldsymbol{b}_1, \frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}\right) \\
&\qquad\qquad \left[\prod_{i=2}^{\ell+1} e^{A_1u_i}A_0G\right] e^{A_1(t-\sum_{i=1}^{\ell+1}u_i)}\mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\mathrm{d}u_1, \\
&= \int_{u_1=0}^{t}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-\sum_{i=1}^{\ell}u_i} \boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}\frac{\boldsymbol{b}_0 e^{A_1u_1}A_0}{\boldsymbol{b}_0 e^{A_1u_1}A_0\boldsymbol{e}}G\left[\prod_{i=2}^{\ell+1} e^{A_1u_i}A_0G\right] e^{A_1(t-\sum_{i=1}^{\ell+1}u_i)} \\
&\qquad\qquad \mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\mathrm{d}u_1, \\
&= \int_{u_1=0}^{t}\int_{u_2=0}^{t-u_1}\cdots\int_{u_{\ell+1}=0}^{t-\sum_{i=1}^{\ell}u_i} \boldsymbol{b}_0\left[\prod_{i=1}^{\ell+1} e^{A_1u_i}A_0G\right] e^{A_1(t-\sum_{i=1}^{\ell+1}u_i)}\mathrm{d}u_{\ell+1}\cdots\mathrm{d}u_2\mathrm{d}u_1, \\
&\equiv \mathbb{E}(\bar{\boldsymbol{B}}(t)I(L=\ell+1)|X(0)=(n,\boldsymbol{b}_0)),
\end{aligned}
$$

where the fourth equation follows from the induction hypothesis and the sixth equation follows from Theorem 2. ■