Commuting matrices in the queue length and sojourn time analysis of MAP/MAP/1 queues

G. Horváth*[1], B. Van Houdt†[2] and M. Telek‡[3]

[1]Budapest University of Technology and Economics, 1521 Budapest, Hungary
[2]University of Antwerp, iMinds, B-2020 Antwerpen, Belgium
[3]MTA-BME Information Systems Research Group, 1521 Budapest, Hungary

May 20, 2014

Abstract

Queues with Markovian arrival and service processes, i.e., MAP/MAP/1 queues, have been useful in the analysis of computer and communication systems and different representations for their stationary sojourn time and queue length distribution have been derived. More specifically, the class of MAP/MAP/1 queues lies at the intersection of the class of QBD queues and the class of semi-Markovian queues.

While QBD queues have a matrix exponential representation for their queue length and sojourn time distribution of order N and N^2, respectively, where N is the size of the background continuous time Markov chain, the reverse is true for a semi-Markovian queue. As the class of MAP/MAP/1 queues lies at the intersection, both the queue length and sojourn time distribution of a MAP/MAP/1 queue has an order N matrix exponential representation.

The aim of this paper is to understand why the order N^2 distributions of the sojourn time of a QBD queue and the queue length of a semi-Markovian queue can be reduced to an order N distribution in the specific case of a MAP/MAP/1 queue. We show that the key observation exists in establishing the commutativity of some fundamental matrices involved in the analysis of the MAP/MAP/1 queue.

Keywords: QBD, MAP/MAP/1 queue, sojourn time distribution, queue length distribution, commuting matrices.

1 Introduction

The class of MAP/MAP/1 queues is a versatile and well-studied class of queueing systems used to model computer and communication systems [5, 6]. Its ef-

*ghorvath@hit.bme.hu
†benny.vanhoudt@uantwerpen.be
‡telek@hit.bme.hu

fectiveness lies in the generality of the Markovian arrival process (MAP) which can be used to fit very different arrival patterns with highly correlated inter-arrival times [12, 7, 19]. The MAP process can also be used to model the service process whenever significant correlation exists in the service times of consecutive customers, e.g. [1], and some authors therefore refer to it as the Markovian service process (MSP). The MAP process has also been extended and analyzed to allow for batch arrivals and multiple customer types [10, 2].

The queue length distribution of the MAP/MAP/1 queue is well-known to be matrix exponential of order N, where N is the product of the number of states of the arrival and service MAP, as its evolution can be captured by means of a Quasi-Birth-Death Markov chain [11]. The sojourn time distribution of the MAP/MAP/1 queue on the other hand can be obtained as a special case of a class of semi-Markovian queues studied by Sengupta [15, 16] and therefore has a matrix exponential form of order N as well. This result was later generalized in [4] for queues with multitype MAP arrivals. More recently, the queue length distribution of a semi-Markovian queue was shown to have a matrix exponential distribution of order N^2 [20], which also gives rise to an order N^2 representation for the queue length distribution of a MAP/MAP/1 queue.

On a different line of research Ozawa studied the sojourn time distribution of a class of so-called Quasi-Birth-Death (QBD) queues [14] and proved that it has a matrix exponential representation of order N^2, where N is the size of the background continuous time Markov chain. As the class of MAP/MAP/1 queues forms a subclass of the set of QBD queues (with N equal to the product of the number of phases of the arrival and service MAP), the result of Ozawa gives rise to an order N^2 representation for the sojourn time distribution of a MAP/MAP/1 queue.

While the order N^2 representations for the queue length of a semi-Markovian queue and the sojourn time in a QBD queue cannot be reduced in general [20], the aim of this paper exists in understanding why these representations collapse to an order N representation in case of the MAP/MAP/1 queue. It turns out that the key feature is the commutativity of some characteristic matrices that appear in the analysis of the queue length and sojourn time distribution of the MAP/MAP/1 queue. Apart from unifying these different representations for the queue length and sojourn time and proving the required commutativity property, we also identify several other sets of commuting matrices that have played a fundamental role in the analysis of the MAP/MAP/1 queue.

The paper is structured as follows. Sections 2 and 3 reintroduce the class of QBD and semi-Markovian queues, respectively, and also summarize the main results on their queue length and sojourn time distributions. In Section 4 we establish two key results that link some of the fundamental matrices and vec-

tors of the class of QBD and semi-Markovian queues in the specific case of a MAP/MAP/1 queue. Four sets of commuting matrices are identified next in Section 5. Finally, in Sections 6 and 7 we show how the well known order N representations for the sojourn time distribution and the queue length distribution of the MAP/MAP/1 queue, respectively, can be obtained by relying on the results established Sections 4 and 5.

2 The Quasi-Birth-Death queue

In a QBD queue the arrivals and the services are modulated by a common continuous time background Markov chain $\mathcal{Z}(t)$. Some of the transitions of the background process are accompanied by an arrival (the associated matrix is denoted by $\boldsymbol{F}$), other transitions of the background process are accompanied by a service completion, assuming that there is at least a customer in the system (given by matrix $\boldsymbol{B}$). There may be transitions by which neither an arrival, nor a service completion occurs (given by matrices $\boldsymbol{L}$ or $\boldsymbol{L}'$ depending on whether the system is busy or empty, respectively). When there is at least one customer in the system the generator of the background process is denoted by $\boldsymbol{Q} = \{q_{ij},\ i,j = 1,\ldots,N\}$. When there is no customer in the queue the generator of the background process might be different and is denoted by $\boldsymbol{Q}' = \{q'_{ij},\ i,j = 1,\ldots,N\}$. Note that $\boldsymbol{Q} = \boldsymbol{B} + \boldsymbol{L} + \boldsymbol{F}$ and $\boldsymbol{Q}' = \boldsymbol{L}' + \boldsymbol{F}$. The stochastic process that keeps track of the number of customers in the system is denoted by $\mathcal{X}(t)$.

With a lexicographical numbering of the states the two-dimensional process $\{\mathcal{X}(t), \mathcal{Z}(t), t > 0\}$ is a QBD Markov chain [8], with its generator given by

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{L}' & \boldsymbol{F} & & & \\ \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & & \\ & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{1}$$

The sojourn time in a QBD queue, $\mathcal{V}$, is defined as the time between an arrival event and the corresponding service instant in steady state assuming a first-come first-served (FCFS) service discipline.

Provided that the QBD Markov chain with transition matrix $\boldsymbol{\Pi}$ is irreducible and positive recurrent, denote its stationary distribution by $\pi = (\pi_0, \pi_1, \ldots)$. The j-th entry of the vector π_k corresponds to the steady state probability that there are k customers in the queue while the background process $\mathcal{Z}(t)$ is in state j. As the steady state distribution of a QBD Markov chain is known to have a matrix geometric form [8], π_k can be written as

$$\pi_k = \pi_0 \boldsymbol{R}^k, \quad k > 0, \tag{2}$$

where $\boldsymbol{R}$ is the minimal non-negative solution of the quadratic matrix equation

$$\mathbf{0} = \boldsymbol{F} + \boldsymbol{R}\boldsymbol{L} + \boldsymbol{R}^2\boldsymbol{B}, \tag{3}$$

and vector π_0 is the unique solution of the following set of linear equations:

$$\begin{aligned} 0 &= \pi_0 \left(\boldsymbol{L}' + \boldsymbol{R}\boldsymbol{B}\right), \\ 1 &= \pi_0 \left(\boldsymbol{I} - \boldsymbol{R}\right)^{-1} \mathbb{1}. \end{aligned} \tag{4}$$

For later use we also introduce the matrix $\boldsymbol{U}$ and $\boldsymbol{G}$ as the smallest non-negative solution of

$$\boldsymbol{U} = \boldsymbol{L} + \boldsymbol{F}(-\boldsymbol{U})^{-1}\boldsymbol{B}, \tag{5}$$

$$\mathbf{0} = \boldsymbol{B} + \boldsymbol{L}\boldsymbol{G} + \boldsymbol{F}\boldsymbol{G}^2, \tag{6}$$

respectively. The matrices $\boldsymbol{R}, \boldsymbol{U}$ and $\boldsymbol{G}$ are all defined by $\boldsymbol{B}, \boldsymbol{L}, \boldsymbol{F}$ and they are related such that $\boldsymbol{R} = \boldsymbol{F}(-\boldsymbol{U})^{-1}$ and $\boldsymbol{G} = (-\boldsymbol{U})^{-1}\boldsymbol{B}$ [8]. The mean arrival rate λ of a QBD queue is given by

$$\lambda = \sum_{i=0}^{\infty} \pi_i \boldsymbol{F} \mathbb{1}.$$

From Equation (2) it is clear that the queue length distribution of a QBD queue has a matrix geometric form of order N. To express the distribution of the sojourn time, let entry j of the vector $\hat{\pi}_k$ denote the probability that the QBD queue is at level k just after the arrival epoch, while the background process is in state j. Ozawa [14] established the following two theorems, where the second theorem shows that the sojourn time distribution has a matrix exponential form of order N^2:

Theorem 1. *(Theorem 1 in [14]) The vectors $\hat{\pi}_k$ are given by*

$$\begin{aligned} \hat{\pi}_1 &= \frac{1}{\lambda}\pi_0 \boldsymbol{F}, \\ \hat{\pi}_k &= \hat{\pi}_1 \hat{\boldsymbol{R}}^{k-1}, \quad k = 2, \ldots, \infty, \end{aligned} \tag{7}$$

with $\hat{\boldsymbol{R}}$ given by

$$\hat{\boldsymbol{R}} = (-\boldsymbol{U})^{-1}\boldsymbol{F}. \tag{8}$$

Theorem 2. *(Theorem 2 in [14]) The distribution of the sojourn time is given by*

$$P(\mathcal{V} < t) = 1 - (\mathbb{1}^T \otimes \hat{\eta}) e^{((\boldsymbol{L}+\boldsymbol{F})^T \otimes \boldsymbol{I}) + (\boldsymbol{B}^T \otimes \hat{\boldsymbol{R}}))t} vec\langle \boldsymbol{I} \rangle, \tag{9}$$

where $\hat{\eta}$ is the stationary phase distribution at arrivals

$$\hat{\eta} = \hat{\pi}_1 \left(\boldsymbol{I} - \hat{\boldsymbol{R}}\right)^{-1}, \tag{10}$$

and $vec\langle\rangle$ denotes the column-stacking operator.

Remark 1: Theorem 1 was proven using probabilistic arguments in [14], but can also be proven easily in an algebraic manner as

$$\begin{aligned}\hat{\pi}_k &= \frac{\pi_{k-1}\boldsymbol{F}}{\sum_{i=0}^{\infty}\pi_i\boldsymbol{F}\mathbb{1}} = \frac{1}{\lambda}\pi_{k-1}\boldsymbol{F} = \frac{1}{\lambda}\pi_0\boldsymbol{R}^{k-1}\boldsymbol{F} \\ &= \frac{1}{\lambda}\pi_0\left(\boldsymbol{F}(-\boldsymbol{U})^{-1}\right)^{k-1}\boldsymbol{F} = \frac{1}{\lambda}\pi_0\boldsymbol{F}\left((-\boldsymbol{U})^{-1}\boldsymbol{F}\right)^{k-1} \\ &= \frac{1}{\lambda}\pi_0\boldsymbol{F}\hat{\boldsymbol{R}}^{k-1}.\end{aligned} \tag{11}$$

3 The semi-Markovian queue

The class of semi-Markovian queues considered in this paper was introduced by Sengupta in [16]. To define this class, consider a bi-variate Markov process $(\mathcal{X}_t, \mathcal{M}_t)_{t\geq 0}$, with $\mathcal{X}_t \geq 0$ and $\mathcal{M}_t \in \{1, \ldots, N\}$. Assume the process evolves as follows: $\mathcal{X}_t$ increases linearly unless a jump occurs. Three types of jumps can occur from (x, i)

1. a jump to (x, j) with rate $(\boldsymbol{A_0})_{i,j}$ (for $i \neq j$),
2. a jump in the interval $([x-u, x), j)$, for $0 < u < x$, with a rate $\boldsymbol{A}_{i,j}(u)$, where we denote $\boldsymbol{dA}_{i,j}(u)$ as its density function, and
3. a jump to $(0, j)$ with rate $\int_{u=x}^{\infty}\boldsymbol{dA}_{i,j}(u)$.

Finally, define the (negative) diagonal entries of $\boldsymbol{A_0}$ such that $(\boldsymbol{A_0} + \int_{u=0}^{\infty}\boldsymbol{dA}(u))\mathbb{1} = \mathbb{1}$ and assume $\boldsymbol{A} = \boldsymbol{A_0} + \int_{u=0}^{\infty}\boldsymbol{dA}(u)$ is irreducible.

Such a Markov process has a matrix exponential distribution [15]. In other words, there exists a size N matrix $\boldsymbol{T}$ such that the length N vector $\alpha(x)$, for $x \geq 0$, which holds the steady-state density of the states $(x, 1)$ to (x, m), can be written as

$$\alpha(x) = \alpha(0)e^{\boldsymbol{T}x}. \tag{12}$$

The matrix $\boldsymbol{T}$ is the smallest non-negative solution to

$$\boldsymbol{T} = \boldsymbol{A_0} + \int_{x=0}^{\infty} e^{\boldsymbol{T}x}\boldsymbol{dA}(x),$$

and $\alpha(0) = \zeta(-\boldsymbol{T})$, where ζ is the unique invariant vector of $\boldsymbol{A}$, i.e., $\zeta\boldsymbol{A} = 0$ and $\zeta\mathbb{1} = 1$.

Next, consider a single server FCFS queue with an infinite waiting room. Observe this queue only when the server is busy and define $\mathcal{A}_t \geq 0$ as the age of the customer in service at time t (of the censored process). Such a queue belongs to the class of semi-Markovian queues defined in [16] if and only if there exists a bi-variate Markov process $(\mathcal{X}_t, \mathcal{M}_t)_{t\geq 0}$ as defined above such

that $\mathcal{X}_t = \mathcal{A}_t$. In other words, there exists an underlying Markov process with generator $\boldsymbol{A} = \boldsymbol{A_0} + \int_{u=0}^{\infty} \boldsymbol{dA}(u)$, such that $\boldsymbol{A_0}$ captures the evolution of the underlying chain while the same customer remains in service and $(\boldsymbol{dA}(u))_{i,j}$ represents the density function of the rate at which service completions occur, while the inter-arrival time to the next customer equals u and the state of the underlying chain changes from i to j.

Sengupta showed that the sojourn time distribution of a semi-Markovian queue has an order N matrix geometric distribution as indicated by the next theorem:

Theorem 3. *(Theorem 3 in [16]) The distribution of the sojourn time of a semi-Markovian queue is given by*

$$P(\mathcal{V} < t) = 1 - \frac{1}{\mu}\zeta e^{\boldsymbol{T}t}(\boldsymbol{A} - \boldsymbol{A_0})\mathbb{1}, \tag{13}$$

where μ the service rate is given by

$$\mu = \int_0^{\infty} \alpha(x)(\boldsymbol{A} - \boldsymbol{A_0})\mathbb{1}\ dx = \zeta(\boldsymbol{A} - \boldsymbol{A_0})\mathbb{1}.$$

The queue length distribution of a semi-Markovian queue on the other hand has a matrix exponential distribution of order N^2 as proven in [20]:

Theorem 4. *(Theorem 2 in [20]) The distribution of the queue length given that the server is busy $\mathcal{N}_b$ of a semi-Markovian queue can be expressed as*

$$P(\mathcal{N}_b = n) = (\mathbb{1}^T \otimes \zeta)(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{M}^{n-1} vec\langle \boldsymbol{I} \rangle, \tag{14}$$

where $\boldsymbol{M}$ is given by

$$\boldsymbol{M} = \int_0^{\infty} ((-\boldsymbol{A_0})^{-1}\boldsymbol{dA}(x) \otimes e^{\boldsymbol{T}x}).$$

4 The MAP/MAP/1 queue

The class of MAP/MAP/1 queues lies in the intersection of the class of semi-Markovian queues introduced by Sengupta [16] and the QBD queues studied by Ozawa [14]. More specifically, if the arrival and service processes of a QBD queue are controlled by independent Markov chains $\mathcal{Z}^{(in)}(t)$ and $\mathcal{Z}^{(out)}(t)$, the QBD queue simplifies to a MAP/MAP/1 queue. By denoting the matrices of the MAP that generates the arrivals by $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ ($\boldsymbol{D_0} + \boldsymbol{D_1} = \boldsymbol{D}$, $\boldsymbol{D} = \{d_{ij},\ i,j = 1,\ldots,N^{(in)}\}$) and the matrices of the MAP generating the service events by $\boldsymbol{S_0}$ and $\boldsymbol{S_1}$ ($\boldsymbol{S_0} + \boldsymbol{S_1} = \boldsymbol{S}$, $\boldsymbol{S} = \{s_{ij},\ i,j = 1,\ldots,N^{(out)}\}$) the blocks

of the QBD Markov chain can be expressed as

$$\begin{aligned} \boldsymbol{F} &= \boldsymbol{D_1} \otimes \boldsymbol{I}, \\ \boldsymbol{L} &= \boldsymbol{D_0} \oplus \boldsymbol{S_0}, \\ \boldsymbol{B} &= \boldsymbol{I} \otimes \boldsymbol{S_1}, \\ \boldsymbol{L'} &= \boldsymbol{D_0} \otimes \boldsymbol{I}. \end{aligned} \tag{15}$$

Similarly, when the matrices $\boldsymbol{A_0}$ and $\boldsymbol{dA}(u)$ characterizing the semi-Markovian queue are of the form $\boldsymbol{A_0} = \boldsymbol{I} \otimes \boldsymbol{S_0}$ and

$$\boldsymbol{dA}(u) = e^{\boldsymbol{D_0}u}\boldsymbol{D_1} \otimes \boldsymbol{S_1},$$

such that $(\boldsymbol{D_0}, \boldsymbol{D_1})$ and $(\boldsymbol{S_0}, \boldsymbol{S_1})$ characterize a MAP process, the semi-Markovian queue reduces to a MAP/MAP/1 queue. In this case, the matrix $\boldsymbol{T}$ can be expressed via the matrix $\hat{\boldsymbol{R}}$ ([16], Equation (15)) as

$$\boldsymbol{T} = (\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_1}). \tag{16}$$

Further $\boldsymbol{A} = (\boldsymbol{I} \otimes \boldsymbol{S_0}) + ((-\boldsymbol{D_0})^{-1}\boldsymbol{D_1} \otimes \boldsymbol{S_1})$ and due to (12) the vector $\alpha(0)$ is given by

$$\alpha(0) = (\theta \otimes \beta)(-\boldsymbol{T}), \tag{17}$$

where the vectors β and θ are the solutions of $\beta(\boldsymbol{S_0} + \boldsymbol{S_1}) = 0, \beta\mathbb{1} = 1$ and $\theta(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1} = \theta, \theta\mathbb{1} = 1$, respectively.

As $(\boldsymbol{A} - \boldsymbol{A_0})\mathbb{1} = (\boldsymbol{I} \otimes \boldsymbol{S_1})\mathbb{1}$ and $\zeta = (\theta \otimes \beta)$, the sojourn time distribution given in Theorem 3 can therefore be written as

$$P(\mathcal{V} < t) = 1 - \frac{1}{\mu}(\theta \otimes \beta)e^{\boldsymbol{T}t}(\boldsymbol{I} \otimes \boldsymbol{S_1})\mathbb{1}, \tag{18}$$

where $\mu = \beta\boldsymbol{S_1}\mathbb{1}$.

Remark 2: It is important to note that in the above definitions we assumed that the phase of the service process is frozen (i.e., remains identical) whenever the server is idle. In fact, without this assumption the MAP/MAP/1 queue would not belong to the class of semi-Markovian queues discussed in Section 3, as the rate of the jumps to $(0, j)$ is no longer given by $\int_{u=x}^{\infty} \boldsymbol{dA}_{i,j}(u)$. Assuming a frozen phase during idle periods is quite common when studying queues with (semi-)Markovian service (e.g., [3]) as it is a natural generalization of the MAP/PH/1 case (which uses a frozen service phase), though examples in which the service process evolves also exist (e.g., [13]). It might be possible to generalize some of the results presented in this paper to the case where the service phase also evolves during idle periods by introducing semi-Markovian queues with a more general boundary behavior.

We end this section by linking some of the fundamental matrices and vectors associated with the QBD Markov chain and the age process:

Theorem 5. *For the MAP/MAP/1 queue the boundary vectors π_0 and $\alpha(0)$ defined by* (4) *and* (17)*, respectively, obey the following equation*

$$\frac{\pi_0 \boldsymbol{F}}{\lambda} = \hat{\pi}_1 = \frac{\alpha(0)}{\mu} \tag{19}$$

Proof. We first express the probability vector corresponding to an arrival to the empty queue in two different ways:

- Based on the queue process this probability vector equals $\hat{\pi}_1$.
- We can express the probability vector that an arrival finds the queue empty also via the age process: it is the probability that the next arrival occurs later than the sojourn time of a customer. Hence we get

$$\frac{\int_0^\infty \alpha(x)(\boldsymbol{I} \otimes \boldsymbol{S_1}) e^{(\boldsymbol{D_0} \otimes \boldsymbol{I})x} ((-\boldsymbol{D_0})^{-1} \boldsymbol{D_1} \otimes \boldsymbol{I}) dx}{\int_0^\infty \alpha(x)(\boldsymbol{I} \otimes \boldsymbol{S_1}) \mathbb{1} \, dx}, \tag{20}$$

  where the denominator is equal to μ (see Theorem 3) and the numerator is $\alpha(0)$ due to Lemma 2.4 in [15].

Thus, we can conclude that $\hat{\pi}_1 = \alpha(0)/\mu$ holds and the result follows from Theorem 1. □

Theorem 6. *For the MAP/MAP/1 queue the matrices $\boldsymbol{T}$ and $\boldsymbol{U}$ defined by* (16) *and* (5)*, respectively, obey the following equation*

$$\boldsymbol{T}(-\boldsymbol{U})^{-1} + (-\boldsymbol{U})^{-1}(\boldsymbol{D_0} \otimes \boldsymbol{I}) = -\boldsymbol{I}, \tag{21}$$

Proof. We start by showing that

$$(-\boldsymbol{U})^{-1} = \int_{u=0}^{\infty} e^{\boldsymbol{T}u} (e^{\boldsymbol{D_0}u} \otimes \boldsymbol{I}) du, \tag{22}$$

using the stochastic interpretation of $(-\boldsymbol{U})^{-1}$ and $e^{\boldsymbol{T}u}$. This equality is closely related to Theorem 6 in [16], in fact it follows from this theorem in case $\boldsymbol{D_1}$ can be inverted. Entry (i, j), with $i = (i_1, i_2)$ and $j = (j_1, j_2)$, of $(-\boldsymbol{U})^{-1}$ holds the expected amount of time that the arrival and service processes spend in state j_1 and j_2, respectively, while there is a single customer in the queue during a busy period that was initiated while the arrival and service process were in state i_1 and i_2, respectively. Next, consider the probabilistic interpretation of entry (i, k) of $e^{\boldsymbol{T}u}$ with $k = (k_1, k_2)$ [15]: it is the expected number of times during a busy period that the age of the customer c in service equals u, the current service state equals k_2 and the state of the arrival process was k_1 when customer c arrived, given that the busy period was initiated in state $i = (i_1, i_2)$. Thus, each of these visits contributes to entry (i, j) of $(-\boldsymbol{U})^{-1}$ if $j_2 = k_2$ and there are no arrivals in an interval of length u after customer c arrived and the state

of the arrival process is k_1 at the start and j_1 at the end of the interval, which is given by entry (k_1, j_1) of the matrix $e^{\boldsymbol{D_0}u}$. This establishes (22).

Further, as $e^{\boldsymbol{D_0}u} \otimes \boldsymbol{I} = e^{(\boldsymbol{D_0}\otimes \boldsymbol{I})u}$ and $\boldsymbol{X} = -\int_{u=0}^{\infty} e^{\boldsymbol{A}u}\boldsymbol{C}e^{\boldsymbol{B}u}du$ is the unique solution of $\boldsymbol{AX} + \boldsymbol{XB} = \boldsymbol{C}$ if both $\boldsymbol{A}$ and $\boldsymbol{B}$ are stable matrices [9, Theorem 13.19] (that is, the real parts of the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$ are negative). It is well known that the matrix $\boldsymbol{D_0}$ is stable, while $\boldsymbol{T}$ is stable due to Lemma 2.4(b) in [16]. □

5 Commuting matrices in MAP/MAP/1 queues

In this section we identify four sets of commuting matrices related to the MAP/MAP/1 queue, where the key equation to prove these is given by (21).

Theorem 7. *The matrices* $\boldsymbol{R}$, $(\boldsymbol{I}\otimes\boldsymbol{S_0}) + \boldsymbol{R}(\boldsymbol{I}\otimes\boldsymbol{S_1})$ *and* $(\boldsymbol{D_1}\otimes\boldsymbol{I}) + \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I})$ *commute.*

Proof. Introduce $\mathcal{S}\boldsymbol{R} = (\boldsymbol{I}\otimes\boldsymbol{S_0}) + \boldsymbol{R}(\boldsymbol{I}\otimes\boldsymbol{S_1})$ and $\mathcal{D}\boldsymbol{R} = \boldsymbol{D_1}\otimes\boldsymbol{I} + \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I})$. By pre-multiplying (21) with $(\boldsymbol{D_1}\otimes\boldsymbol{I})$ one finds

$$(\boldsymbol{D_1}\otimes\boldsymbol{I})\boldsymbol{T}(-\boldsymbol{U})^{-1} + \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I}) = -(\boldsymbol{D_1}\otimes\boldsymbol{I}).$$

Using the expression for $\boldsymbol{T}$ and $\tilde{\boldsymbol{R}}$ shows that

$$(\boldsymbol{I}\otimes\boldsymbol{S_0})\boldsymbol{R} + \boldsymbol{R}(\boldsymbol{I}\otimes\boldsymbol{S_1})\boldsymbol{R} = -(\boldsymbol{D_1}\otimes\boldsymbol{I}) - \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I}),$$

that is,

$$\mathcal{S}\boldsymbol{R}\ \boldsymbol{R} = -(\boldsymbol{D_1}\otimes\boldsymbol{I}) - \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I}). \tag{23}$$

The fact that $\boldsymbol{R}$ and $\mathcal{S}\boldsymbol{R}$ commute now follows from the fact that quadratic equation (3) for $\boldsymbol{R}$ can be written as

$$\boldsymbol{R}\ \mathcal{S}\boldsymbol{R}\ = -(\boldsymbol{D_1}\otimes\boldsymbol{I}) - \boldsymbol{R}(\boldsymbol{D_0}\otimes\boldsymbol{I}). \tag{24}$$

Equation (23) implies

$$\boldsymbol{R}\ \mathcal{D}\boldsymbol{R}\ = -\boldsymbol{R}\ \mathcal{S}\boldsymbol{R}\ \boldsymbol{R},$$

while (24) yields

$$\mathcal{D}\boldsymbol{R}\ \boldsymbol{R} = -\boldsymbol{R}\ \mathcal{S}\boldsymbol{R}\ \boldsymbol{R},$$

meaning $\boldsymbol{R}$ and $\mathcal{D}\boldsymbol{R}$ commute.

Finally, as $\boldsymbol{R}$ commutes with $\mathcal{S}\boldsymbol{R}$ and $\mathcal{D}\boldsymbol{R}$, we have

$$\begin{aligned}\mathcal{D}\boldsymbol{R}\ \mathcal{S}\boldsymbol{R} &= \mathcal{D}\boldsymbol{R}\ (\boldsymbol{I}\otimes\boldsymbol{S_0}) + \mathcal{D}\boldsymbol{R}\ \boldsymbol{R}(\boldsymbol{I}\otimes\boldsymbol{S_1}) \\ &= \mathcal{S}\boldsymbol{R}\ (\boldsymbol{D_0}\otimes\boldsymbol{I}) + \boldsymbol{R}\ \mathcal{S}\boldsymbol{R}\ \boldsymbol{R}(\boldsymbol{D_1}\otimes\boldsymbol{I}) = \mathcal{S}\boldsymbol{R}\ \mathcal{D}\boldsymbol{R}.\end{aligned}$$

□

Theorem 8. *The matrices* $\hat{\boldsymbol{R}}$ *and* $\boldsymbol{T} = \boldsymbol{I} \otimes \boldsymbol{S_0} + \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_1})$ *commute.*

Proof. Post-multiplying (21) by $(\boldsymbol{D_1} \otimes \boldsymbol{I})$ implies that

$$\boldsymbol{T}(-\boldsymbol{U})^{-1}(\boldsymbol{D_1} \otimes \boldsymbol{I}) = (\boldsymbol{U}^{-1}(\boldsymbol{D_0} \otimes \boldsymbol{I}) - \boldsymbol{I})(\boldsymbol{D_1} \otimes \boldsymbol{I}).$$

As noted before $\mathcal{S}\hat{\boldsymbol{R}} = \boldsymbol{T}$ and $\hat{\boldsymbol{R}} = (-\boldsymbol{U})^{-1}(\boldsymbol{D_1} \otimes \boldsymbol{I})$, meaning

$$\mathcal{S}\hat{\boldsymbol{R}}\ \hat{\boldsymbol{R}} = (\boldsymbol{U}^{-1}(\boldsymbol{D_0} \otimes \boldsymbol{I}) - \boldsymbol{I})(\boldsymbol{D_1} \otimes \boldsymbol{I}).$$

Since $\boldsymbol{U} = (\boldsymbol{D_0} \otimes \boldsymbol{I}) + \mathcal{S}\boldsymbol{R}$, we therefore get

$$\mathcal{S}\hat{\boldsymbol{R}}\ \hat{\boldsymbol{R}} = -[\boldsymbol{U}^{-1}(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \boldsymbol{U}^{-1}\boldsymbol{R}(\boldsymbol{I} \otimes \boldsymbol{S_1})](\boldsymbol{D_1} \otimes \boldsymbol{I}).$$

As $\boldsymbol{R} = (\boldsymbol{D_1} \otimes \boldsymbol{I})(-\boldsymbol{U})^{-1}$, $\hat{\boldsymbol{R}} = (-\boldsymbol{U})^{-1}(\boldsymbol{D_1} \otimes \boldsymbol{I})$ and $(\boldsymbol{D_1} \otimes \boldsymbol{I})$ commutes with $(\boldsymbol{I} \otimes \boldsymbol{S_0})$ and $(\boldsymbol{I} \otimes \boldsymbol{S_1})$, this implies

$$\mathcal{S}\hat{\boldsymbol{R}}\ \hat{\boldsymbol{R}} = \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}^2(\boldsymbol{I} \otimes \boldsymbol{S_1}) = \hat{\boldsymbol{R}}\ \mathcal{S}\hat{\boldsymbol{R}}\ .$$

□

Remark 3: Given Theorems 7 and 8 one may expect that $(\boldsymbol{D_1} \otimes \boldsymbol{I}) + \hat{\boldsymbol{R}}(\boldsymbol{D_0} \otimes \boldsymbol{I})$ and $\hat{\boldsymbol{R}}$ also commute, but numerical experiments indicate that this is not true in general.

Theorem 9. *The matrices* $\boldsymbol{G}$, $(\boldsymbol{D_0} \otimes \boldsymbol{I}) + (\boldsymbol{D_1} \otimes \boldsymbol{I})\boldsymbol{G}$ *and* $(\boldsymbol{I} \otimes \boldsymbol{S_1}) + (\boldsymbol{I} \otimes \boldsymbol{S_0})\boldsymbol{G}$ *commute.*

Proof. To simplify the notation we introduce $\mathcal{D}\boldsymbol{G} = (\boldsymbol{D_0} \otimes \boldsymbol{I}) + (\boldsymbol{D_1} \otimes \boldsymbol{I})\boldsymbol{G}$ and $\mathcal{S}\boldsymbol{G} = (\boldsymbol{I} \otimes \boldsymbol{S_1}) + (\boldsymbol{I} \otimes \boldsymbol{S_0})\boldsymbol{G}$. First, post-multiply (21) by $(\boldsymbol{I} \otimes \boldsymbol{S_1})$ and use the fact that $\boldsymbol{G} = (-\boldsymbol{U})^{-1}(\boldsymbol{I} \otimes \boldsymbol{S_1})$ to obtain

$$\boldsymbol{T}\boldsymbol{G} + \boldsymbol{G}(\boldsymbol{D_0} \otimes \boldsymbol{I}) = -(\boldsymbol{I} \otimes \boldsymbol{S_1}),$$

where we also used the fact that $(\boldsymbol{I} \otimes \boldsymbol{S_1})$ and $(\boldsymbol{D_0} \otimes \boldsymbol{I})$ commute. Using (16) and $\hat{\boldsymbol{R}} = (-\boldsymbol{U})^{-1}(\boldsymbol{D_1} \otimes \boldsymbol{I})$ yields

$$(\boldsymbol{I} \otimes \boldsymbol{S_0})\boldsymbol{G} + \boldsymbol{G}(\boldsymbol{D_1} \otimes \boldsymbol{I})\boldsymbol{G} + \boldsymbol{G}(\boldsymbol{D_0} \otimes \boldsymbol{I}) = -(\boldsymbol{I} \otimes \boldsymbol{S_1}).$$

In other words,

$$\boldsymbol{G}\ \mathcal{D}\boldsymbol{G}\ = -(\boldsymbol{I} \otimes \boldsymbol{S_0})\boldsymbol{G} - (\boldsymbol{I} \otimes \boldsymbol{S_1}). \tag{25}$$

From the quadratic equation (6) for $\boldsymbol{G}$ we find

$$\mathcal{D}\boldsymbol{G}\ \boldsymbol{G} = -(\boldsymbol{I} \otimes \boldsymbol{S_0})\boldsymbol{G} - (\boldsymbol{I} \otimes \boldsymbol{S_1}), \tag{26}$$

meaning $\mathcal{D}\boldsymbol{G}\ \boldsymbol{G} = \boldsymbol{G}\ \mathcal{D}\boldsymbol{G}$. By (25)

$$\mathcal{S}\boldsymbol{G}\ \boldsymbol{G} = -\boldsymbol{G}\ \mathcal{D}\boldsymbol{G}\ \boldsymbol{G},$$

while by (26), we have

$$\boldsymbol{G}\ \mathcal{SG}\ = -\boldsymbol{G}\ \mathcal{DG}\ \boldsymbol{G},$$

which yields $\boldsymbol{G}\ \mathcal{SG}\ =\ \mathcal{SG}\ \boldsymbol{G}$. Finally, if $\boldsymbol{G}$ commutes with $\mathcal{DG}$ and $\mathcal{SG}$, then

$$\begin{aligned}\mathcal{SG}\ \mathcal{DG} &= (\boldsymbol{I}\otimes\boldsymbol{S_1})\ \mathcal{DG}\ + (\boldsymbol{I}\otimes\boldsymbol{S_0})\ \mathcal{DG}\ \boldsymbol{G}\\ &= (\boldsymbol{D_0}\otimes\boldsymbol{I})\ \mathcal{SG}\ + (\boldsymbol{D_1}\otimes\boldsymbol{I})\ \mathcal{SG}\ \boldsymbol{G} = \mathcal{DG}\ \mathcal{SG}.\end{aligned}$$

□

Remark 4: Let $\boldsymbol{Q}$ be the $\boldsymbol{Q}$-matrix of the workload process of the MAP/MAP/1 queue as defined in [17]. Then entry (i,j), with $i=(i_1,i_2)$ and $j=(j_1,j_2)$, of $\exp(\boldsymbol{Q}u)$ holds the state transition probability during the first passage from (u,i_1,i_2) to $(0,j_1,j_2)$ [18]. This implies that the matrix $\boldsymbol{G}$ can be expressed as

$$\boldsymbol{G}=\int_{u=0}^{\infty}(\boldsymbol{I}\otimes\exp(\boldsymbol{S_0}u))(I\otimes\boldsymbol{S_1})\exp(\boldsymbol{Q}u)du, \tag{27}$$

as $(I\otimes\exp(\boldsymbol{S_0}u))(\boldsymbol{I}\otimes\boldsymbol{S_1})$ is the density of the amount of work remaining for the customer in service. Further, by Equation (2.13) in [18], $\boldsymbol{Q}$ can be written as

$$\boldsymbol{Q}=(\boldsymbol{D_0}\otimes\boldsymbol{I})+\int_{u=0}^{\infty}(\boldsymbol{D_1}\otimes\boldsymbol{I})(\boldsymbol{I}\otimes\exp(\boldsymbol{S_0}u)\boldsymbol{S_1})\exp(\boldsymbol{Q}u)du,$$

in other words $\boldsymbol{Q}=(\boldsymbol{D_0}\otimes\boldsymbol{I})+(\boldsymbol{D_1}\otimes\boldsymbol{I})\boldsymbol{G}$.

Theorem 10. *The matrices $\hat{\boldsymbol{G}}$ and $\boldsymbol{D_0}\otimes\boldsymbol{I}+(\boldsymbol{D_1}\otimes\boldsymbol{I})\hat{\boldsymbol{G}}$ commute.*

Proof. Let $\mathcal{D}\hat{\boldsymbol{G}}\ =\boldsymbol{D_0}\otimes\boldsymbol{I}+(\boldsymbol{D_1}\otimes\boldsymbol{I})\hat{\boldsymbol{G}}$ and $\mathcal{S}\hat{\boldsymbol{R}}\ =\boldsymbol{I}\otimes\boldsymbol{S_0}+\hat{\boldsymbol{R}}(\boldsymbol{I}\otimes\boldsymbol{S_1})$. Pre-multiplying (21) with $(\boldsymbol{I}\otimes\boldsymbol{S_1})$ gives

$$\hat{\boldsymbol{G}}(\boldsymbol{D_0}\otimes\boldsymbol{I})=(\boldsymbol{I}\otimes\boldsymbol{S_1})[\boldsymbol{T}\boldsymbol{U}^{-1}-\boldsymbol{I}],$$

which indicates that

$$\begin{aligned}\hat{\boldsymbol{G}}\ \mathcal{D}\hat{\boldsymbol{G}}\ &=(\boldsymbol{I}\otimes\boldsymbol{S_1})[\boldsymbol{T}\boldsymbol{U}^{-1}-\boldsymbol{I}]+\hat{\boldsymbol{G}}(\boldsymbol{D_1}\otimes\boldsymbol{I})(\boldsymbol{I}\otimes\boldsymbol{S_1})(-\boldsymbol{U})^{-1}\\ &=(\boldsymbol{I}\otimes\boldsymbol{S_1})[\boldsymbol{T}\boldsymbol{U}^{-1}-\boldsymbol{I}+\hat{\boldsymbol{R}}(\boldsymbol{I}\otimes\boldsymbol{S_1})(-\boldsymbol{U})^{-1}].\end{aligned}$$

Using the expression $\boldsymbol{T}=\ \mathcal{S}\hat{\boldsymbol{R}}$ yields

$$\hat{\boldsymbol{G}}\ \mathcal{D}\hat{\boldsymbol{G}}\ =(\boldsymbol{I}\otimes\boldsymbol{S_1})[(\boldsymbol{I}\otimes\boldsymbol{S_0})\boldsymbol{U}^{-1}-\boldsymbol{I}]. \tag{28}$$

Further, by definition of $\mathcal{D}\hat{\boldsymbol{G}}$ and the fact that $\hat{\boldsymbol{G}}=(\boldsymbol{I}\otimes\boldsymbol{S_1})(-\boldsymbol{U})^{-1}$ and $\hat{\boldsymbol{G}}^2=(\boldsymbol{I}\otimes\boldsymbol{S_1})\boldsymbol{G}(-\boldsymbol{U})^{-1}$, we have

$$\mathcal{D}\hat{\boldsymbol{G}}\ \hat{\boldsymbol{G}}=(\boldsymbol{I}\otimes\boldsymbol{S_1})[(\boldsymbol{D_0}\otimes\boldsymbol{I})+(\boldsymbol{D_1}\otimes\boldsymbol{I})\boldsymbol{G}](-\boldsymbol{U})^{-1}.$$

As $\boldsymbol{U}=(\boldsymbol{I}\otimes\boldsymbol{S_0})+\mathcal{DG}$, we get

$$\mathcal{D}\hat{\boldsymbol{G}}\ \hat{\boldsymbol{G}}=(\boldsymbol{I}\otimes\boldsymbol{S_1})[(\boldsymbol{I}\otimes\boldsymbol{S_0})\boldsymbol{U}^{-1}-\boldsymbol{I}]. \tag{29}$$

Hence, $\mathcal{D}\hat{\boldsymbol{G}}\ \hat{\boldsymbol{G}}=\hat{\boldsymbol{G}}\ \mathcal{D}\hat{\boldsymbol{G}}$ due to (28) and (29). □

6 Sojourn time distribution of the MAP/MAP/1 queue via QBD Markov chain

In this section we show how an order $N = N^{(in)}N^{(out)}$ representation for the sojourn time distribution of a MAP/MAP/1 queue can be obtained directly from the QBD Markov chain. To determine the distribution of the sojourn time it suffices to know the distribution of the queue length at arrival instants and the distribution of the time taken by the QBD queue to generate k service events, for $k \geq 1$.

Recall that entry j of the vector $\hat{\pi}_k$ denotes the probability that the QBD queue is at level k just after the arrival epoch, while the background process is in state j. Further, let entry (i, j) of the matrix $\boldsymbol{N}(k, t)$ denote the probability that exactly k service events occur in a non-idle interval of length t, while the phase of the underlying process is i and j at the start and end of the interval, respectively, that is

$$[\boldsymbol{N}(k,t)]_{i,j} = P(\mathcal{X}_s(t) = 1, \mathcal{Z}(t) = j | \mathcal{X}_s(0) = k+1, \mathcal{Z}(0) = i),$$

where $\mathcal{X}_s(t)$ corresponds to the level of the two-dimensional Markov chain $\{\mathcal{X}_s(t), \mathcal{Z}(t), t > 0\}$ with its generator given by

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{L'} + \boldsymbol{F} & & & \\ \boldsymbol{B} & \boldsymbol{L} + \boldsymbol{F} & & \\ & \boldsymbol{B} & \boldsymbol{L} + \boldsymbol{F} & \\ & & \ddots & \ddots \end{bmatrix}. \tag{30}$$

The matrices $\boldsymbol{N}(k, t)$ are determined by the following set of differential equations [8]:

$$\frac{\partial}{\partial t}\boldsymbol{N}(0,t) = \boldsymbol{N}(0,t)(\boldsymbol{L} + \boldsymbol{F}), \tag{31}$$

$$\frac{\partial}{\partial t}\boldsymbol{N}(k,t) = \boldsymbol{N}(k,t)(\boldsymbol{L} + \boldsymbol{F}) + \boldsymbol{N}(k-1,t)\boldsymbol{B}, \tag{32}$$

for $k = 1, \ldots, \infty$ with boundary conditions $\boldsymbol{N}(0, 0) = \boldsymbol{I}$ and $\boldsymbol{N}(k, 0) = \boldsymbol{0}$ for $k > 0$. The generating function of the departure events is defined by $\boldsymbol{N}^*(z, t) = \sum_{k=0}^{\infty} z^k \boldsymbol{N}(k, t)$. Multiplying (31) and (32) by z^k and summing up for $k = 0, 1, \ldots$ gives

$$\frac{\partial}{\partial t}\boldsymbol{N}^*(z,t) = \boldsymbol{N}(z,t)(\boldsymbol{L} + \boldsymbol{F} + z\boldsymbol{B}), \tag{33}$$

with initial condition $\boldsymbol{N}^*(z, 0) = \boldsymbol{I}$. Its solution is given by

$$\boldsymbol{N}^*(z,t) = e^{(\boldsymbol{L}+\boldsymbol{F}+z\boldsymbol{B})t}. \tag{34}$$

Remark 5: It was also noted in [14, Remark 1] that the sojourn time distribution can also be expressed as $P(\mathcal{V} > t) = \hat{\eta}\boldsymbol{W}(t)\mathbb{1}$, where

$$\boldsymbol{W}(t) = \sum_{k=0}^{\infty} \hat{\boldsymbol{R}}^k \boldsymbol{N}(k,t), \tag{35}$$

and that $\boldsymbol{W}(t)$ is the solution of the differential equation

$$\frac{d}{dt}\boldsymbol{W}(t) = \boldsymbol{W}(t)(\boldsymbol{L} + \boldsymbol{F}) + \hat{\boldsymbol{R}}\boldsymbol{W}(t)\boldsymbol{B}. \tag{36}$$

with $\boldsymbol{W}(0) = \boldsymbol{I}$. Note if $\hat{\boldsymbol{R}}$ and $\boldsymbol{W}(t)$ were to commute, this differential equation immediately leads to a matrix exponential distribution for the sojourn time of order N. Ozawa [14] notes that $\hat{\boldsymbol{R}}$ and $\boldsymbol{W}(t)$ commute for the M/PH/1 queue, but not in general for the QBD queue. In fact, even for the MAP/M/1 queue $\hat{\boldsymbol{R}}$ and $\boldsymbol{W}(t)$ do not commute in general, meaning (36) does not give immediate rise to an order N representation. More specifically, for the MAP/M/1 queue we can easily see that $\boldsymbol{W}(t)$ can be expressed as

$$\boldsymbol{W}(t) = \sum_{k=0}^{\infty} \hat{\boldsymbol{R}}^k e^{\boldsymbol{D}t} \frac{(\mu t)^k}{k!} e^{-\mu t} = e^{\hat{\boldsymbol{R}}\mu t} e^{(\boldsymbol{D}-\mu\boldsymbol{I})t}. \tag{37}$$

Thus $\hat{\boldsymbol{R}}$ and $\boldsymbol{W}(t)$ only commute if $\hat{\boldsymbol{R}}$ and $e^{(\boldsymbol{D}-\mu\boldsymbol{I})t}$ commute, which only holds in some special cases.

Next, we will make a slight modification to $\boldsymbol{W}(t)$ for the MAP/MAP/1 queue such that we obtain a differential equation where the modified $\boldsymbol{W}(t)$, denoted as $\tilde{\boldsymbol{W}}(t)$ directly leads to an order N sojourn time distribution. More specifically, we introduce the matrix $\tilde{\boldsymbol{W}}(t)$ similar to (35) as

$$\tilde{\boldsymbol{W}}(t) = \sum_{k=0}^{\infty} \hat{\boldsymbol{R}}^k \tilde{\boldsymbol{N}}(k,t), \tag{38}$$

where $\tilde{\boldsymbol{N}}(k,t)$ is defined as the solution to the differential equation

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{N}}(0,t) = \tilde{\boldsymbol{N}}(0,t)(\boldsymbol{I} \otimes \boldsymbol{S_0}), \tag{39}$$

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{N}}(k,t) = \tilde{\boldsymbol{N}}(k,t)(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \tilde{\boldsymbol{N}}(k-1,t)(\boldsymbol{I} \otimes \boldsymbol{S_1}), \tag{40}$$

for $k = 1, \ldots, \infty$ with $\tilde{\boldsymbol{N}}(0,0) = \boldsymbol{I}$ and $\tilde{\boldsymbol{N}}(k,0) = \boldsymbol{0}$ for $k > 0$. Observe that the definition of $\tilde{\boldsymbol{N}}(k,t)$ differs from $\boldsymbol{N}(k,t)$ in that $\tilde{\boldsymbol{N}}(k,t)$ does not follow the evolution of the arrival process, more precisely the phase of the arrival process remains fixed. This slight difference will turn out to be essential in the subsequent discussion.

We can now establish the following theorem, the proof of which is similar in nature to the one of Theorem 2 in [14] and is included for completeness:

Theorem 11. *The sojourn time distribution in a MAP/MAP/1 queue can be expressed as* $P(\mathcal{V} > t) = \hat{\eta}\tilde{\boldsymbol{W}}(t)\mathbb{1}$, *where* $\tilde{\boldsymbol{W}}(t)$ *is the unique solution to the differential equation*

$$\frac{d}{dt}\tilde{\boldsymbol{W}}(t) = \tilde{\boldsymbol{W}}(t)(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}\tilde{\boldsymbol{W}}(t)(\boldsymbol{I} \otimes \boldsymbol{S_1}). \tag{41}$$

with $\tilde{\boldsymbol{W}}(0) = \boldsymbol{I}$.

Proof. The probability that the sojourn time of an arriving customer is greater than t equals the probability that the number of service events generated up to time t is less than the number of customers the arriving customer found in the system (including itself). Hence, we have

$$\begin{aligned} P(\mathcal{V} > t) &= \sum_{n=1}^{\infty} \hat{\pi}_n \sum_{k=0}^{n-1} \tilde{\boldsymbol{N}}(k,t)\mathbb{1} \\ &= \sum_{n=1}^{\infty} \hat{\pi}_1 \hat{\boldsymbol{R}}^{n-1} \sum_{k=0}^{n-1} \tilde{\boldsymbol{N}}(k,t)\mathbb{1} \\ &= \sum_{k=0}^{\infty} \hat{\eta}\hat{\boldsymbol{R}}^{k} \tilde{\boldsymbol{N}}(k,t)\mathbb{1} = \hat{\eta}\tilde{\boldsymbol{W}}(t)\mathbb{1}, \end{aligned} \tag{42}$$

where $\hat{\eta} = \sum_{k=1}^{\infty} \hat{\pi}_1 \hat{\boldsymbol{R}}^{k-1}$ has a closed form given by (10). To obtain the differential equation in (41) for $\tilde{\boldsymbol{W}}(t)$, it suffices to sum (39) and (40) after left-multiplying them by $\hat{\boldsymbol{R}}^{k}$. □

Remark 6: Making use of the vec$\langle\rangle$ operator and utilizing its properties, Theorem 11 yields

$$\begin{aligned} \frac{d}{dt}\text{vec}\langle\tilde{\boldsymbol{W}}(t)\rangle &= ((\boldsymbol{I} \otimes \boldsymbol{S_0})^T \otimes \boldsymbol{I})\text{vec}\langle\tilde{\boldsymbol{W}}(t)\rangle \\ &\quad + ((\boldsymbol{I} \otimes \boldsymbol{S_1})^T \otimes \hat{\boldsymbol{R}})\text{vec}\langle\tilde{\boldsymbol{W}}(t)\rangle, \end{aligned}$$

for which the closed form solution is

$$\text{vec}\langle\tilde{\boldsymbol{W}}(t)\rangle = e^{((\boldsymbol{I}\otimes\boldsymbol{S_0})^T\otimes\boldsymbol{I}) + ((\boldsymbol{I}\otimes\boldsymbol{S_1})^T\otimes\hat{\boldsymbol{R}}))t}\text{vec}\langle\boldsymbol{I}\rangle, \tag{43}$$

by noting that $\tilde{\boldsymbol{W}}(0) = \boldsymbol{I}$. Thus the distribution of the sojourn time in a MAP/MAP/1 queue can also be expressed as

$$\begin{aligned} &P(\mathcal{V} < t) = 1 - \hat{\eta}\tilde{\boldsymbol{W}}(t)\mathbb{1} \\ &= 1 - (\mathbb{1}^T \otimes \hat{\eta})e^{((\boldsymbol{I}\otimes\boldsymbol{S_0})^T\otimes\boldsymbol{I}) + ((\boldsymbol{I}\otimes\boldsymbol{S_1})^T\otimes\hat{\boldsymbol{R}}))t}\text{vec}\langle\boldsymbol{I}\rangle. \end{aligned}$$

This distribution is a matrix exponential distribution of order N^2 and is therefore of little interest. Theorem 11 is however interesting as it directly leads to an order N representation for the sojourn time distribution:

Theorem 12. *The sojourn time distribution of a MAP/MAP/1 queue has an order N matrix exponential representation given by*

$$P(\mathcal{V} < t) = 1 - \hat{\eta} e^{((\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_1}))t} \mathbb{1}. \tag{44}$$

Proof. We prove that $\tilde{\boldsymbol{W}}(\boldsymbol{t}) = e^{\boldsymbol{T}t} = e^{((\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_1}))t}$ by showing that it is a solution of (41). If we plug $\tilde{\boldsymbol{W}}(\boldsymbol{t}) = e^{\boldsymbol{T}t}$ into (41), it suffices to verify that

$$\frac{d}{dt} e^{\boldsymbol{T}t} = e^{\boldsymbol{T}t}(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}} e^{\boldsymbol{T}t}(\boldsymbol{I} \otimes \boldsymbol{S_1}).$$

Now, by Theorem 8 the matrices $\hat{\boldsymbol{R}}$ and $\boldsymbol{T}$ commute, meaning $\hat{\boldsymbol{R}}$ and $e^{\boldsymbol{T}t}$ commute and $\tilde{\boldsymbol{W}}(\boldsymbol{t}) = e^{\boldsymbol{T}t}$ if

$$\frac{d}{dt} e^{\boldsymbol{T}t} = e^{\boldsymbol{T}t} \left[(\boldsymbol{I} \otimes \boldsymbol{S_0}) + \hat{\boldsymbol{R}}(\boldsymbol{I} \otimes \boldsymbol{S_1}) \right] = e^{\boldsymbol{T}t} \boldsymbol{T},$$

which clearly holds. □

Remark 7: For the MAP/M/1 queue we can easily see that $\tilde{\boldsymbol{W}}(t)$ is found as

$$\tilde{\boldsymbol{W}}(t) = \sum_{k=0}^{\infty} \hat{\boldsymbol{R}}^k \frac{(\mu t)^k}{k!} e^{-\mu t} = e^{-\mu t}\, e^{\hat{\boldsymbol{R}} \mu t}, \tag{45}$$

meaning $\hat{\boldsymbol{R}}$ and $\tilde{\boldsymbol{W}}(t)$ commute and Theorem 12 immediately follows from (41).

Remark 8: The two expressions for the distribution of the sojourn time in a MAP/MAP/1 queue given by (18) and (44) can be proven to be equal in a direct manner. Due to (10), we have

$$P(\mathcal{V} > t) = \hat{\eta} e^{\boldsymbol{T}t} \mathbb{1} = \hat{\pi}_1 (\boldsymbol{I} - \hat{\boldsymbol{R}})^{-1} e^{\boldsymbol{T}t} \mathbb{1}.$$

Theorem 5 and (17) therefore imply

$$P(\mathcal{V} > t) = \frac{1}{\mu} (\theta \otimes \beta)(-\boldsymbol{T})(\boldsymbol{I} - \hat{\boldsymbol{R}})^{-1} e^{\boldsymbol{T}t} \mathbb{1}.$$

Exploiting the fact that the matrices $\hat{\boldsymbol{R}}, \boldsymbol{T}$ and $e^{\boldsymbol{T}t}$ commute (due to Theorem 8) yields

$$\begin{aligned} P(\mathcal{V} > t) &= \frac{1}{\mu} (\theta \otimes \beta)(-\boldsymbol{T})(\boldsymbol{I} - \hat{\boldsymbol{R}})^{-1} e^{\boldsymbol{T}t} \mathbb{1} \\ &= \frac{1}{\mu} (\theta \otimes \beta) e^{\boldsymbol{T}t} (\boldsymbol{I} - \hat{\boldsymbol{R}})^{-1} (-\boldsymbol{T}) \mathbb{1} \\ &= \frac{1}{c} (\theta \otimes \beta) e^{\boldsymbol{T}t} (\boldsymbol{I} \otimes \boldsymbol{S_1}) \mathbb{1}, \end{aligned}$$

where in the last step we utilized that $(\boldsymbol{I}-\hat{\boldsymbol{R}})^{-1}(-\boldsymbol{T})\mathbb{1} = (\boldsymbol{I}\otimes\boldsymbol{S_1})\mathbb{1}$ which can be proven as follows: (16) clearly implies that

$$-\boldsymbol{T} + (\boldsymbol{I}\otimes\boldsymbol{S_1}) = -(\boldsymbol{I}\otimes\boldsymbol{S_0}) + (\boldsymbol{I}-\hat{\boldsymbol{R}})(\boldsymbol{I}\otimes\boldsymbol{S_1}),$$

which yields

$$(\boldsymbol{I}-\hat{\boldsymbol{R}})^{-1}(-\boldsymbol{T}) = (\boldsymbol{I}\otimes\boldsymbol{S_1}) - (\boldsymbol{I}-\hat{\boldsymbol{R}})^{-1}(\boldsymbol{I}\otimes(\boldsymbol{S_0}+\boldsymbol{S_1})),$$

and the equality follows by post-multiplying it with $\mathbb{1}$ as $(\boldsymbol{S_0}+\boldsymbol{S_1})\mathbb{1} = \boldsymbol{0}$.

7 Queue length distribution of the MAP/MAP/1 queue via age process

In this section we derive the well-known matrix geometric form of the queue length distribution of the MAP/MAP/1 queue via the age process by relying on some of the results presented in Section 4 and 5.

First, let us introduce the matrices $\tilde{\boldsymbol{L}}(k,u)$ whose entry (i,j) denotes the probability that k arrivals occur in an interval of length u while the phase of the underlying process is i at the start and j at the end of the interval, respectively. These matrices are determined by the following set of differential equations:

$$\frac{\partial}{\partial u}\tilde{\boldsymbol{L}}(0,u) = \tilde{\boldsymbol{L}}(0,u)(\boldsymbol{D_0}\otimes\boldsymbol{I}), \tag{46}$$

$$\frac{\partial}{\partial u}\tilde{\boldsymbol{L}}(k,u) = \tilde{\boldsymbol{L}}(k,u)(\boldsymbol{D_0}\otimes\boldsymbol{I}) + \tilde{\boldsymbol{L}}(k-1,u)(\boldsymbol{D_1}\otimes\boldsymbol{I}), \tag{47}$$

for $k = 1,\ldots,\infty$ with $\tilde{\boldsymbol{L}}(0,0) = \boldsymbol{I}$ and $\tilde{\boldsymbol{L}}(k,0) = \boldsymbol{0}$ for $k > 0$. Notice that the definition of $\tilde{\boldsymbol{L}}(k,u)$ and the corresponding set of differential equations are the dual of the ones defined by (39) and (40) in the sense that $\tilde{\boldsymbol{L}}(k,u)$ is related to the arrival process while $\tilde{\boldsymbol{N}}(k,u)$ defines the same quantity for the service process.

Before proceeding to the queue length distribution, let us introduce the matrices $\tilde{\boldsymbol{Q}}_k$, for $k \geq 0$, that will play an important role in the sequel as the counterpart of $\tilde{\boldsymbol{W}}(t)$ introduced in Section 6. The matrices $\tilde{\boldsymbol{Q}}_k$ are defined as

$$\tilde{\boldsymbol{Q}}_k = \int_{u=0}^{\infty} e^{\boldsymbol{T}u}\tilde{\boldsymbol{L}}(k,u)du. \tag{48}$$

The next theorem derives the steady state distribution based on the age process, similar to Example 5.2 in [4].

Theorem 13. *The stationary queue length distribution of the MAP/MAP/1 queue is given by*

$$p_0 = 1-\rho, \tag{49}$$

$$p_k = \rho\alpha(0)\tilde{\boldsymbol{Q}}_{k-1}\mathbb{1}, \quad k > 0, \tag{50}$$

where the matrices $\tilde{\boldsymbol{Q}}_k$ are the unique solution of the following matrix Sylvester equations:

$$\boldsymbol{T}\tilde{\boldsymbol{Q}}_0 + \tilde{\boldsymbol{Q}}_0(\boldsymbol{D_0} \otimes \boldsymbol{I}) = -\boldsymbol{I}, \tag{51}$$

$$\boldsymbol{T}\tilde{\boldsymbol{Q}}_k + \tilde{\boldsymbol{Q}}_k(\boldsymbol{D_0} \otimes \boldsymbol{I}) = -\tilde{\boldsymbol{Q}}_{k-1}(\boldsymbol{D_1} \otimes \boldsymbol{I}), \tag{52}$$

for $k > 0$.

Proof. Given that the queue is not empty (with probability ρ) the number of customers in the system is equal to the number of arrivals during the sojourn time (the age) of the customer residing in the server, plus one (which is the customer in the server itself). The age process keeps track of the age of the customer in service, together with the current service phase and the state of the arrival process when the customer in service arrived, its density function is given by $\alpha(u) = \alpha(0)e^{\boldsymbol{T}u}$, hence

$$p_k = \rho \int_{u=0}^{\infty} \alpha(0)e^{\boldsymbol{T}u}\tilde{\boldsymbol{L}}(k-1,u)\mathbb{1}du = \rho\alpha(0)\tilde{\boldsymbol{Q}}_{k-1}\mathbb{1}. \tag{53}$$

To prove (52) we pre-multiply (47) by $e^{\boldsymbol{T}u}$ and take the integral from 0 to ∞, yielding

$$\begin{aligned}\int_{u=0}^{\infty} e^{\boldsymbol{T}u}\frac{\partial}{\partial u}\tilde{\boldsymbol{L}}(k,u)du = &\underbrace{\int_{u=0}^{\infty} e^{\boldsymbol{T}u}\tilde{\boldsymbol{L}}(k,u)du}_{\tilde{\boldsymbol{Q}}_k}(\boldsymbol{D_0} \otimes \boldsymbol{I}) \\ &+ \underbrace{\int_{u=0}^{\infty} e^{\boldsymbol{T}u}\tilde{\boldsymbol{L}}(k-1,u)du}_{\tilde{\boldsymbol{Q}}_{k-1}}(\boldsymbol{D_1} \otimes \boldsymbol{I}),\end{aligned} \tag{54}$$

where the integration of the left-hand side by parts results in $-\boldsymbol{T}\tilde{\boldsymbol{Q}}_k$ if $k > 0$, establishing (52). Equation (51) can be proven similarly, by starting from (46) and applying the same steps. □

Remark 9: Based on the results of Theorem 13 and using the vec$\langle\rangle$ operator it is possible to obtain an explicit matrix-geometric distribution for the queue length. From (51) and (52) we have

$$\text{vec}\langle\tilde{\boldsymbol{Q}}_0\rangle = -\Big(\boldsymbol{I} \otimes \boldsymbol{T} + (\boldsymbol{D_0} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)^{-1}\text{vec}\langle\boldsymbol{I}\rangle, \tag{55}$$

$$\text{vec}\langle\tilde{\boldsymbol{Q}}_k\rangle = \Big(\boldsymbol{I} \otimes \boldsymbol{T} + (\boldsymbol{D_0} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)^{-1}\Big((\boldsymbol{D_1} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)\text{vec}\langle\tilde{\boldsymbol{Q}}_{k-1}\rangle, \tag{56}$$

which yields

$$\begin{aligned}p_k = -\rho(\mathbb{1}^T \otimes \alpha(0))&\left[\Big(\boldsymbol{I} \otimes \boldsymbol{T} + (\boldsymbol{D_0} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)^{-1}\Big((\boldsymbol{D_1} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)\right]^{k-1} \\ &\cdot \Big(\boldsymbol{I} \otimes \boldsymbol{T} + (\boldsymbol{D_0} \otimes \boldsymbol{I})^T \otimes \boldsymbol{I}\Big)^{-1}\text{vec}\langle\boldsymbol{I}\rangle.\end{aligned} \tag{57}$$

This distribution is, however, of order N^2, while it is known that standard matrix-analytic techniques lead to order N queue length distribution (see (2)). The following theorem states that the order N^2 matrix geometric solution collapses to order N due to the commuting property of some matrices proven in Section 5.

Theorem 14. *The stationary queue length distribution of the MAP/MAP/1 queue has an order N matrix-geometric representation given by*

$$p_0 = 1 - \rho, \tag{58}$$

$$p_k = \rho\alpha(0)\hat{\boldsymbol{R}}^{k-1}(-\boldsymbol{U})^{-1}\mathbb{1}, \quad k > 0. \tag{59}$$

Proof. Equation (59) follows from Theorem 13 once we show that $\tilde{\boldsymbol{Q}}_k = \hat{\boldsymbol{R}}^k(-\boldsymbol{U})^{-1}$. Plugging $\tilde{\boldsymbol{Q}}_k = \hat{\boldsymbol{R}}^k(-\boldsymbol{U})^{-1}$ into the matrix Sylvester equation (52) gives

$$\boldsymbol{T}\hat{\boldsymbol{R}}^k(-\boldsymbol{U})^{-1} + \hat{\boldsymbol{R}}^k(-\boldsymbol{U})^{-1}(\boldsymbol{D_0}\otimes\boldsymbol{I}) = -\hat{\boldsymbol{R}}^{k-1}\underbrace{(-\boldsymbol{U})^{-1}(\boldsymbol{D_1}\otimes\boldsymbol{I})}_{\hat{\boldsymbol{R}}}. \tag{60}$$

By observing that the right-hand side is equal to $-\hat{\boldsymbol{R}}^k$ (see (8)) and that $\boldsymbol{T}$ and $\hat{\boldsymbol{R}}$ commute (see Theorem 8), it suffices to show that

$$\boldsymbol{T}(-\boldsymbol{U})^{-1} + (-\boldsymbol{U})^{-1}(\boldsymbol{D_0}\otimes\boldsymbol{I}) = -\boldsymbol{I} \tag{61}$$

is satisfied, which is ensured by Theorem 6. Equation (58) can be proven similarly. □

Remark 10: For the M/MAP/1 queue we can easily see that $\tilde{\boldsymbol{Q}}_k$ can be expressed as

$$\tilde{\boldsymbol{Q}}_k = \int_{u=0}^{\infty} e^{\boldsymbol{T}u}\frac{(\lambda u)^k}{k!}e^{-\lambda u}du = \lambda^k(\lambda\boldsymbol{I} - \boldsymbol{T})^{-(k+1)}, \tag{62}$$

meaning $\boldsymbol{T}$ and $\tilde{\boldsymbol{Q}}_k$ commute. Further for the M/MAP/1 queue $\boldsymbol{R} = \hat{\boldsymbol{R}}$, which implies that $\boldsymbol{U} = \boldsymbol{T} - \lambda\boldsymbol{I}$ and Theorem 14 now immediately follows from Theorem 13.

Remark 11: Now we show that the queue length distribution defined by (59) and the one based on the matrix-analytic approach (2) are equivalent. We start by applying Theorem 5 on (59) and obtain

$$p_k = \rho\alpha(0)\hat{\boldsymbol{R}}^{k-1}(-\boldsymbol{U})^{-1}\mathbb{1} = \rho\frac{\mu}{\lambda}\pi_0(\boldsymbol{D_1}\otimes\boldsymbol{I})\hat{\boldsymbol{R}}^{k-1}(-\boldsymbol{U})^{-1}\mathbb{1}. \tag{63}$$

Making use of $\hat{\boldsymbol{R}} = (-\boldsymbol{U})^{-1}(\boldsymbol{D_1}\otimes\boldsymbol{I})$ we get

$$\begin{aligned} p_k &= \pi_0(\boldsymbol{D_1}\otimes\boldsymbol{I})\left[(-\boldsymbol{U})^{-1}(\boldsymbol{D_1}\otimes\boldsymbol{I})\right]^{k-1}(-\boldsymbol{U})^{-1}\mathbb{1} \\ &= \pi_0\left[(\boldsymbol{D_1}\otimes\boldsymbol{I})(-\boldsymbol{U})^{-1}\right]^k\mathbb{1}, \end{aligned} \tag{64}$$

from which, observing that $(\boldsymbol{D_1} \otimes \boldsymbol{I})(-\boldsymbol{U})^{-1} = \boldsymbol{R}$, the well known result $p_k = \pi_0 \boldsymbol{R}^k \mathbb{1}$ follows.

Acknowledgment

This work was supported by the Hungarian Government through the OTKA K101150 project, by the European Union (co-financed by the European Social Fund) through the TAMOP-4.2.2C-11/1/KONV-2012-0001 project, and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The second author was supported by the FWO-Flanders (project number G024514N).

References

[1] V. Gupta, M. Burroughs, and M. Harchol-Balter. Analysis of scheduling policies under correlated job sizes. *Perform. Eval.*, 67(11):996–1013, November 2010.

[2] Q. He. The versatility of the MMAP[K] and the MMAP[K]/G[K]/1 queue. *Queueing Systems*, 38:397–418, 2001.

[3] Q. He. Age process, workload process, sojourn times, and waiting times in a discrete-time SM[K]/PH[K]/1/FCFS queue. *Queueing Systems*, 49:363–403, 2005.

[4] Q. He. Analysis of a continuous time SM[K]/PH[K]/1/FCFS queue: Age process, sojourn times, and queue lengths. *Journal of Systems Science and Complexity*, 25:133–155, 2012.

[5] A. Heindl, Q. Zhang, and E. Smirni. ETAQA truncation models for the MAP/MAP/1 departure process. In *QEST '04: Proceedings of the The Quantitative Evaluation of Systems, First International Conference*, pages 100–109, Washington, DC, USA, 2004.

[6] A. Horváth, G. Horváth, and M. Telek. A joint moments based analysis of networks of MAP/MAP/1 queues. *Performance Evaluation*, 67:759–778, 2010.

[7] G. Horváth, P. Buchholz, and M. Telek. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. In *Proc of QEST 2005*. IEEE Computer Society, 2005.

[8] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics, 1999.

[9] A. J. Laub. *Matrix Analysis For Scientists And Engineers*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

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

[11] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models*. John Hopkins University Press, Baltimore, MD, USA, 1981.

[12] M. F. Neuts. *Structured Stochastic Matrices of M/G/1-type and their Applications*. Marcel Dekker, New York, NY, 1989.

[13] R. Núñez-Queija. A queueing model with varying service rate for ABR. In *Proc. TOOLS 1998*, volume 1469 of *Lecture Notes in Computer Science*, pages 93–104. Springer Berlin Heidelberg, 1998.

[14] T. Ozawa. Sojourn time distributions in the queue defined by a general QBD process. *Queueing Syst. Theory Appl.*, 53(4):203–211, August 2006.

[15] B. Sengupta. Markov processes whose steady state distribution is matrix exponential with an application to the GI/PH/1 queue. *Adv. in Appl. Probab.*, 21:159–180, 1989.

[16] B. Sengupta. The semi-Markovian queue: theory and applications. *Stochastic Models*, 6(3):383–413, 1990.

[17] T. Takine. A continuous version of matrix-analytic methods with skip-free to the left property. *Stochastic Models*, 12(4):673–682, 1996.

[18] T. Takine and T. Hasegawa. The workload in a MAP/G/1 queue with state-dependent services: its applications to a queue with preemptive resume priority. *Stochastic Models*, 10(1):183–204, 1994.

[19] M. Telek and G. Horváth. A minimal representation of markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, October 2007.

[20] B. Van Houdt. A phase-type representation for the queue length distribution of a semi-markovian queue. In *QEST*, pages 49–58. IEEE Computer Society, 2010.