M/G/1-type Markov Processes: A tutorial*

Alma Riska and Evgenia Smirni

Department of Computer Science
College of William and Mary
Williamsburg, VA 23187-8795
e-mail: {riska,esmirni}@cs.wm.edu

Abstract. M/G/1-type processes are commonly encountered when modeling modern complex computer and communication systems. In this tutorial, we present a detailed survey of existing solution methods for M/G/1-type processes, focusing on the matrix-analytic methodology. From first principles and using simple examples, we derive the fundamental matrix-analytic results and lay out recent advances. Finally, we give an overview of an existing, state-of-the-art software tool for the analysis of M/G/1-type processes.
Keywords: M/G/1-type processes; matrix analytic method; Markov chains.

1 Introduction

Matrix analytic techniques, pioneered by Marcel Neuts [25, 26], provide a framework that is widely used for the exact analysis of a general and frequently encountered class of queueing models. In these models, the embedded Markov chains are two-dimensional generalizations of elementary GI/M/1 and M/G/1 queues [13], and their intersection, i.e., quasi-birth-death (QBD) processes. GI/M/1 and M/G/1 Markov chains model systems with interarrival and service times characterized, respectively, by *general* distributions rather than simple exponentials and are often used as the modeling tool of choice in modern computer and communication systems [24, 30, 35, 6, 18]. As a consequence, considerable effort has been placed into the development of efficient matrix-analytic techniques for their analysis [26, 21, 8, 9, 11, 15]. Alternatively, GI/M/1 and M/G/1 Markov chains can be analyzed by means of eigenvalues and eigenvectors [7].

The class of models that can be analyzed using M/G/1-type Markov chains includes the important class of BMAP/G/1 queues, where the arrival process is a batch Markovian arrival process (BMAP) [17, 26, 3]. Special cases of BMAPs include phase-type renewal processes (e.g., Erlang or Hyperexponential processes) and non-renewal processes (e.g., the Markov modulated Poisson process). The importance of BMAPs lies in their ability to be more effective and powerful traffic models than the simple Poisson process or the batch Poisson process, as

* This work has been supported by National Science Foundation under grands EIA-9974992, CCR-0098278, and ACI-0090221.

they can effectively capture dependence and correlation, salient characteristics of Internet traffic [27, 12, 33].

In this paper, we focus on the solution techniques for M/G/1-type Markov chains. Neuts [25] defines various classes of infinite-state Markov chains with a repetitive structure, whose state space[1] is partitioned into the boundary states $\mathcal{S}^{(0)} = \{s_1^{(0)}, \ldots, s_m^{(0)}\}$ and the sets of states $\mathcal{S}^{(i)} = \{s_1^{(i)}, \ldots, s_n^{(i)}\}$, for $i \geq 1$, that correspond to the repetitive portion of the chain. For the class of M/G/1-type Markov chains, the infinitesimal generator $\mathbf{Q}_{M/G/1}$ has upper block Hessenberg form:

$$\mathbf{Q}_{M/G/1} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \widehat{\mathbf{F}}^{(2)} & \widehat{\mathbf{F}}^{(3)} & \widehat{\mathbf{F}}^{(4)} & \cdots \\ \widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{1}$$

We use the letters “L”, “F”, and “B” to describe “local”, ‘forward”, and “backward” transition rates, respectively, in relation to a set of states $\mathcal{S}^{(i)}$ for $i \geq 1$, and a “$\widehat{\ }$” for matrices related to $\mathcal{S}^{(0)}$.

For systems of the M/G/1-type, matrix analytic methods have been proposed for the solution of the basic equation $\boldsymbol{\pi} \cdot \mathbf{Q}_{M/G/1} = \mathbf{0}$ [26], where $\boldsymbol{\pi}$ is the (infinite) stationary probability vector of all states in the chain. Key to the matrix-analytic methods is the computation of an auxiliary matrix called $\mathbf{G}$. Traditional solution methodologies for M/G/1-type processes compute the stationary probability vector with a recursive function based on $\mathbf{G}$. Iterative algorithms are used to determine $\mathbf{G}$ [20, 16].

Another class of Markov-chains with repetitive structure that commonly occurs in modeling of computer systems is the class of GI/M/1-type processes,

[1] We use calligraphic letters to indicate sets (e.g., $\mathcal{A}$), lower case boldface Roman or Greek letters to indicate row vectors (e.g., $\mathbf{a}$, $\boldsymbol{\alpha}$), and upper case boldface Roman letters to indicate matrices (e.g., $\mathbf{A}$). We use superscripts in parentheses or subscripts to indicate family of related entities (e.g., $\mathcal{A}^{(1)}$, $\mathbf{A}_1$), and we extend the notation to subvectors or submatrices by allowing sets of indices to be used instead of single indices (e.g., $\mathbf{a}[\mathcal{A}]$, $\mathbf{A}[\mathcal{A}, \mathcal{B}]$). Vector and matrix elements are indicated using square brackets (e.g., $\mathbf{a}[1]$, $\mathbf{A}[1, 2]$). $RowSum(\cdot)$ indicates the diagonal matrix whose entry in position (r, r) is the sum of the entries on the r^{th} row of the argument (which can be a rectangular matrix). $Norm(\cdot)$ indicates a matrix whose rows are normalized. $\mathbf{0}$ and $\mathbf{1}$ indicate a row vector or a matrix of 0's, or a row vector of 1's, of the appropriate dimensions, respectively.

whose infinitesimal generator $\mathbf{Q}_{GI/M/1}$ has a lower block Hessenberg form:

$$\mathbf{Q}_{GI/M/1} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \widehat{\mathbf{B}}^{(1)} & \mathbf{L} & \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots \\ \widehat{\mathbf{B}}^{(2)} & \mathbf{B}^{(1)} & \mathbf{L} & \mathbf{F} & \mathbf{0} & \cdots \\ \widehat{\mathbf{B}}^{(3)} & \mathbf{B}^{(2)} & \mathbf{B}^{(1)} & \mathbf{L} & \mathbf{F} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (2)$$

The solution of GI/M/1-type processes is significantly simpler than the solution of M/G/1-type processes because of the *matrix geometric* relation [25] that exists among the stationary probabilities of sets $\mathcal{S}^{(i)}$ for $i \geq 1$. This property leads to significant algebraic simplifications resulting in the very elegant matrix-geometric solution technique that was pioneered by Neuts and that was later popularized by Nelson in the early '90s [23, 24]. Key to the matrix-geometric solution is a matrix called $\mathbf{R}$ which is used in the computation of the steady-state probability vector and measures of interest.

Quasi-Birth-Death (QBD) processes are the intersection of M/G/1-type and GI/M/1-type processes and their infinitesimal generator has the structure depicted in Eq.(3).

$$\mathbf{Q}_{QDB} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (3)$$

Since QBDs are special cases of both M/G/1-type processes and GI/M/1-type processes, either the matrix-analytic method or the matrix-geometric solution can be used for their analysis. The matrix-geometric solution is the preferable one because of its simplicity. Both matrices $\mathbf{G}$ and $\mathbf{R}$ are defined for QBD processes. We direct the interested reader to [16] for recent advances on the analysis of QBD processes.

Key to the solution of Markov chains of the M/G/1, GI/M/1, and QBD types, is the existence of a repetitive structure, as illustrated in Eqs. (1), (2), and (3), that allows for a certain recursive procedure to be applied for the computation of the stationary probability vector $\boldsymbol{\pi}^{(i)}$ corresponding to $\mathcal{S}^{(i)}$ for $i \geq 1$. It is this recursive relation that gives elegance to the solution for the case of $GI/M/1$ (and consequently QBD) Markov chains, but results in unfortunately more complicated mathematics for the case of the M/G/1-type.

The purpose of this tutorial is to shed light into the existing techniques for the analysis of Markov chains of the M/G/1 type that are traditionally considered not easy to solve. Our intention is to derive from first principles (i.e., global balance equations) the repetitive patterns that allow for their solution and illustrate that the mathematics involved are less arduous than initially feared.

Our stated goals and outline of this tutorial are the following:

- Give an overview of the matrix-geometric solution of GI/M/1 and QBD processes and establish from first principles why a geometric solution exists (Section 2).
- Use first principles to establish the most stable recursive relation for the case of M/G/1-type processes and essentially illustrate the absence of any geometric relation among the steady state probabilities of sets $\mathcal{S}^{(i)}$, $i \geq 0$, for such chains (Section 3).
- Present an overview of the current state of the art of efficient solutions for M/G/1-type processes (Section 4).
- State the stability conditions for M/G/1-type processes (Section 5).
- Summarize the features of an existing software tool that can provide M/G/1-type solutions (Section 6).

Our aim is to make these results more accessible to performance modelers. We do this by presenting simplified derivations that are often example driven and by describing an existing tool for the solution of such processes.

2 Matrix geometric solutions for GI/M/1-type and QBD processes

In this section we give a brief overview[2] of the matrix geometric solution technique for GI/M/1-type and QBD processes. While QBDs fall under both the M/G/1 and the GI/M/1-type cases, they are most commonly associated with GI/M/1 processes because they can be both solved using the very well-known matrix geometric approach [25].

Key to the general solution for the generator of Eqs.(2) and (3) is the assumption that a geometric relation[3] holds among the stationary probability vectors $\boldsymbol{\pi}^{(i)}$ of states in $\mathcal{S}^{(i)}$ for $i \geq 1$:

$$\forall i \geq 1, \;\; \boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{R}^{i-1}, \tag{4}$$

where, in the GI/M/1-type case, $\mathbf{R}$ is the solution of the matrix equation

$$\mathbf{F} + \mathbf{R} \cdot \mathbf{L} + \sum_{k=1}^{\infty} \mathbf{R}^{k+1} \cdot \mathbf{B}^{(k)} = \mathbf{0}, \tag{5}$$

and can be computed using iterative numerical algorithms. The above equation is obtained from the balance equations of the repeating portion of the process, i.e., starting from the third column of $\mathbf{Q}_{GI/M/1}$. Using Eq.(4) and substituting

[2] In this section and in the remainder of this tutorial we assume continuous time Markov chains, or CTMCs, but our discussion applies just as well to discrete time Markov chains, or DTMCs.

[3] This is similar to the simplest degenerate case of a QBD process, the straight forward birth-death M/M/1 case.

in the balance equation that corresponds to the second column of $\mathbf{Q}_{GI/M/1}$, and together with the normalization condition

$$\boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot \sum_{i=1}^{\infty} \mathbf{R}^{i-1} \cdot \mathbf{1}^T = 1 \quad \text{i.e.,} \quad \boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot (\mathbf{I} - \mathbf{R})^{-1} \cdot \mathbf{1}^T = 1,$$

we obtain the following system of linear equations

$$[\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}] \cdot \begin{bmatrix} \mathbf{e} & (\mathbf{L}^{(0)})^{\diamond} & \widehat{\mathbf{F}}^{(1)} \\ (\mathbf{I} - \mathbf{R})^{-1} \cdot \mathbf{e} & \left(\sum_{k=1}^{\infty} \mathbf{R}^{k-1} \cdot \widehat{\mathbf{B}}^{(k)}\right)^{\diamond} & \mathbf{L} + \sum_{k=1}^{\infty} \mathbf{R}^{k} \cdot \mathbf{B}^{(k)} \end{bmatrix} = [1, \mathbf{0}], \tag{6}$$

that yields a unique solution for $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$. The symbol "$\diamond$" indicates that we discard one (any) column of the corresponding matrix, since we added a column representing the normalization condition. For $i \geq 2$, $\boldsymbol{\pi}^{(i)}$ can be obtained numerically from Eq. (4), but many useful performance metrics such as expected system utilization, throughput, or queue length can be expressed explicitly in closed-form using $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\mathbf{R}$ only (e.g., the average queue length is simply given by $\boldsymbol{\pi}^{(1)} \cdot (\mathbf{I} - \mathbf{R})^{-2} \cdot \mathbf{1}^T$) [23].

In the case of QBD processes, Eq. (5) simply reduces to the matrix quadratic equation

$$\mathbf{F} + \mathbf{R} \cdot \mathbf{L} + \mathbf{R}^2 \cdot \mathbf{B} = \mathbf{0},$$

while $\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$ are obtained as the solution of the following system of linear equations [24]:

$$[\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}] \cdot \begin{bmatrix} \mathbf{e} & (\mathbf{L}^{(0)})^{\diamond} & \widehat{\mathbf{F}}^{(1)} \\ (\mathbf{I} - \mathbf{R})^{-1} \cdot \mathbf{e} & (\widehat{\mathbf{B}})^{\diamond} & \mathbf{L} + \mathbf{R} \cdot \mathbf{B} \end{bmatrix} = [1, \mathbf{0}].$$

Again, the average queue length is given by the same equation as in the GI/M/1 case.

2.1 Why does a geometric relation hold for QBD processes?

There is a clear intuitive appeal to the fact that a geometric relation holds for QBD processes. In this section, we first focus on the reasons for the existence of this relationship via a simple example. Our first example is a QBD process that models an $M/Cox_2/1$ queue. The state transition diagram of the CTMC that models this queue is depicted in Figure 1. The state space $\mathcal{S}$ of this CTMC is divided into subsets $\mathcal{S}^{(0)} = \{(0,0)\}$ and $\mathcal{S}^{(i)} = \{(i,1),(i,2)\}$ for $i \geq 1$, implying that the stationary probability vector is also divided in the respective subvectors $\boldsymbol{\pi}^{(0)} = [\boldsymbol{\pi}(0,0)]$ and $\boldsymbol{\pi}^{(i)} = [\boldsymbol{\pi}(i,1), \boldsymbol{\pi}(i,2)]$, for $i \geq 1$. The block-partitioned infinitesimal generator $\mathbf{Q}_{QBD}$ is a infinite block tridiagonal matrix as defined in Eq.(3) and its component matrices are:

$$\begin{array}{lll} \widehat{\mathbf{L}} = [-\lambda], & \widehat{\mathbf{F}} = [\lambda \quad 0], & \widehat{\mathbf{B}} = \begin{bmatrix} 0.2\mu \\ \gamma \end{bmatrix}, \\ \mathbf{B} = \begin{bmatrix} 0.2\mu & 0 \\ \gamma & 0 \end{bmatrix}, & \mathbf{L} = \begin{bmatrix} -(\lambda+\mu) & 0.8\mu \\ 0 & -(\lambda+\gamma) \end{bmatrix}, & \mathbf{F} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}. \end{array} \tag{7}$$

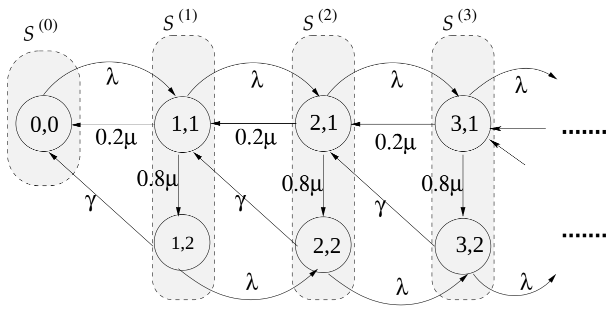

Fig. 1. The CTMC modeling an $M/Cox_2/1$ queue.

To illustrate the existence of the geometric relationship among the various stationary probability vectors $\boldsymbol{\pi}^{(i)}$, we use the concept of stochastic complementation [22]. A detailed summary of some important results on stochastic complementation is presented in Appendix A. We partition the state space into

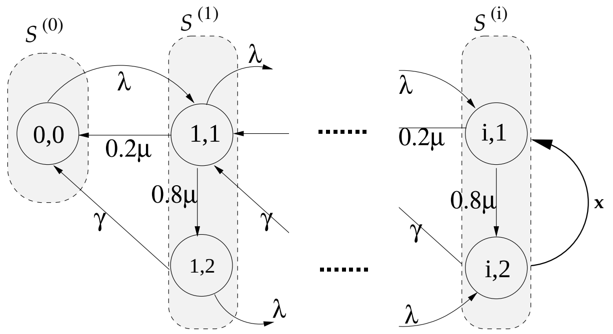

Fig. 2. The CTMC of the stochastic complement of $\mathcal{A} = \cup_{j=0}^{j=i} \mathcal{S}^{(j)}$ of the CTMC modeling an $M/Cox_2/1$ queue.

two subsets; $\mathcal{A} = \cup_{j=0}^{j=i} \mathcal{S}^{(j)}$ and $\overline{\mathcal{A}} = \cup_{j=i+1}^{\infty} \mathcal{S}^{(j)}$, i.e., a set with finite number of states and a set with an infinite number of states, respectively. The stochastic complement of the states in $\mathcal{A}$ is a new Markov chain that "skips over" all states in $\overline{\mathcal{A}}$. This Markov chain includes states in $\mathcal{A}$ only but all transitions out of $\mathcal{S}^{(i)}$ (i.e., the boundary set) to $\mathcal{S}^{(i+1)}$ (i.e., the first set in $\overline{\mathcal{A}}$) need to be "folded back" to $\mathcal{A}$ (see Figure 2). This folding introduces a new direct transition with rate x that ensures that the stochastic complement of the states in $\mathcal{A}$ is a stand-alone process. Because of the structure of this particular process, i.e., $\mathcal{A}$ is entered from $\overline{\mathcal{A}}$ *only* through state $(i, 1)$, x is simply equal to λ (see Lemma 1 in Appendix A). Furthermore, because of the repetitive structure of the original chain, this rate does not depend on i (which essentially defines the size of set $\mathcal{A}$).

The steady state probability vector $\overline{\boldsymbol{\pi}} = [\overline{\boldsymbol{\pi}}^{(0)}, \cdots, \overline{\boldsymbol{\pi}}^{(i)}]$ of the stochastic complement of the states in $\mathcal{A}$ relates to the steady state probability $\boldsymbol{\pi}_{\mathcal{A}}$ of the original process with: $\overline{\boldsymbol{\pi}} = \boldsymbol{\pi}_{\mathcal{A}}/\boldsymbol{\pi}_{\mathcal{A}} \cdot \mathbf{1}^T$. This implies that if a relation exists between $\overline{\boldsymbol{\pi}}^{(i-1)}$ and $\overline{\boldsymbol{\pi}}^{(i)}$, then the same relation holds for $\boldsymbol{\pi}^{(i-1)}$ and $\boldsymbol{\pi}^{(i)}$.

The flow balance equations for states $(i, 1)$ and $(i, 2)$ in the stochastic complement of $\mathcal{A}$, are:

$$(0.2\mu + 0.8\mu)\overline{\boldsymbol{\pi}}^{(i)}[1] = \lambda\overline{\boldsymbol{\pi}}^{(i-1)}[1] + \lambda\overline{\boldsymbol{\pi}}^{(i)}[2],$$

$$(\gamma + \lambda)\overline{\boldsymbol{\pi}}^{(i)}[2] = \lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2] + 0.8\mu\overline{\boldsymbol{\pi}}^{(i)}[1],$$

which can further be expressed as:

$$(-\mu\overline{\boldsymbol{\pi}}^{(i)}[1] + \lambda\overline{\boldsymbol{\pi}}^{(i)}[2]) = -\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[1],$$

$$(0.8\mu\overline{\boldsymbol{\pi}}^{(i)}[1] - (\gamma + \lambda)\overline{\boldsymbol{\pi}}^{(i)}[2]) = -\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2].$$

This last set of equations leads us to the following matrix equation

$$\left[\overline{\boldsymbol{\pi}}^{(i)}[1], \overline{\boldsymbol{\pi}}^{(i)}[2]\right] \begin{bmatrix} -\mu & 0.8\mu \\ \lambda & -(\gamma + \lambda) \end{bmatrix} = -\left[\overline{\boldsymbol{\pi}}^{(i-1)}[1], \overline{\boldsymbol{\pi}}^{(i-1)}[2]\right] \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix},$$

which implies that the relation between $\boldsymbol{\pi}^{(i-1)}$ and $\boldsymbol{\pi}^{(i)}$ can be expressed as

$$\boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}^{(i-1)}\mathbf{R}, \tag{8}$$

where matrix $\mathbf{R}$ is defined as

$$\mathbf{R} = -\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \cdot \begin{bmatrix} -\mu & 0.8\mu \\ \lambda & -(\gamma + \lambda) \end{bmatrix}^{-1} = -\mathbf{F}\left(\mathbf{L} + \mathbf{F}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\right)^{-1}. \tag{9}$$

Applying Eq.(8) recursively, one can obtain the result of Eq.(4). Observe that in this particular case an explicit computation of $\mathbf{R}$ is possible (i.e., there is no need to compute $\mathbf{R}$ [28] via an iterative numerical procedure as in the general case). This is a direct effect of the fact that in this example backward transitions from $\mathcal{S}^{(i)}$ to $\mathcal{S}^{(i-1)}$ are directed toward a single state only. In Appendix B, we give details on the cases when matrix $\mathbf{R}$ can be explicitly computed.

2.2 Generalization: geometric solution for the GI/M/1 processes

We generalize the finding in the previous example by considering a GI/M/1-queue with infinitesimal generator $\mathbf{Q}_{GI/M/1}$ similarly to the proof given in [14]. To evaluate the relation between $\boldsymbol{\pi}^{(i-1)}$ and $\boldsymbol{\pi}^{(i)}$ for $i > 1$, we construct the stochastic complement of the states in $\mathcal{A} = \cup_{j=0}^{i}\mathcal{S}^{(j)}$ ($\overline{\mathcal{A}} = \mathcal{S} - \mathcal{A}$). The stochastic complement of states in $\mathcal{A}$ has an infinitesimal generator defined by the following relation

$$\overline{\mathbf{Q}} = \mathbf{Q}[\mathcal{A}, \mathcal{A}] + \mathbf{Q}[\mathcal{A}, \overline{\mathcal{A}}] \cdot (-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}])^{-1} \cdot \mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}],$$

where

$$\mathbf{Q}[\mathcal{A},\mathcal{A}] = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}} & \cdots \mathbf{0} & \mathbf{0} \\ \widehat{\mathbf{B}}^{(1)} & \mathbf{L} & \cdots \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \;\; \vdots & \vdots \\ \widehat{\mathbf{B}}^{(i-1)} & \mathbf{B}^{(i-2)} & \cdots \mathbf{L} & \mathbf{F} \\ \widehat{\mathbf{B}}^{(i)} & \mathbf{B}^{(i-1)} & \cdots \mathbf{B}^{(1)} & \mathbf{L} \end{bmatrix}, \qquad \mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}] = \begin{bmatrix} \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \cdots \\ \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \cdots \\ \vdots \;\vdots\; \vdots\; \vdots\; \vdots \\ \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \cdots \\ \mathbf{F}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \cdots \end{bmatrix},$$

$$\mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}] = \begin{bmatrix} \widehat{\mathbf{B}}^{(i+1)} & \mathbf{B}^{(i)} & \cdots & \mathbf{B}^{(1)} \\ \widehat{\mathbf{B}}^{(i+2)} & \mathbf{B}^{(i+1)} & \cdots & \mathbf{B}^{(2)} \\ \widehat{\mathbf{B}}^{(i+3)} & \mathbf{B}^{(i+2)} & \cdots & \mathbf{B}^{(3)} \\ \widehat{\mathbf{B}}^{(i+4)} & \mathbf{B}^{(i+3)} & \cdots & \mathbf{B}^{(4)} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad \mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}] = \begin{bmatrix} \mathbf{L} & \mathbf{F} & \mathbf{0} & \mathbf{0} \cdots \\ \mathbf{B}^{(1)} & \mathbf{L} & \mathbf{F} & \mathbf{0} \cdots \\ \mathbf{B}^{(2)} & \mathbf{B}^{(1)} & \mathbf{L} & \mathbf{F} \cdots \\ \mathbf{B}^{(3)} & \mathbf{B}^{(2)} & \mathbf{B}^{(1)} & \mathbf{L} \cdots \\ \vdots & \vdots & \vdots & \vdots \; \ddots \end{bmatrix}. \tag{10}$$

Observe that $\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}]$ is the same matrix for any $i > 1$. We define its inverse to be as follows

$$(-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} & \mathbf{A}_{0,2} & \mathbf{A}_{0,3} & \cdots \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \cdots \\ \mathbf{A}_{2,0} & \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \cdots \\ \mathbf{A}_{3,0} & \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{11}$$

From the special structure of $\mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}]$ we conclude that the second term in the summation that defines $\overline{\mathbf{Q}}$ is a matrix with all block entries equal to zero except the very last block row, whose block entries $\mathbf{X}_j$ are of the form:

$$\mathbf{X}_j = \mathbf{F} \cdot \sum_{k=0}^{\infty} \mathbf{A}_{0,k} \widehat{\mathbf{B}}^{(j+1+k)} \qquad j = i$$

and

$$\mathbf{X}_j = \mathbf{F} \cdot \sum_{k=0}^{\infty} \mathbf{A}_{0,k} \mathbf{B}^{(j+1+k)}, \qquad 0 \le j < i.$$

Note that $\mathbf{X}_0 = \mathbf{F} \cdot \sum_{k=0}^{\infty} \mathbf{A}_{0,k} \mathbf{B}^{(1+k)}$ which means that $\mathbf{X}_0$ does not depend on the value of $i > 1$. The infinitesimal generator $\overline{\mathbf{Q}}$ of the stochastic complement of states in $\mathcal{A}$ is determined as

$$\overline{\mathbf{Q}} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}} & \cdots & \mathbf{0} & \mathbf{0} \\ \widehat{\mathbf{B}}^{(1)} & \mathbf{L} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \widehat{\mathbf{B}}^{(i-1)} & \mathbf{B}^{(i-2)} & \cdots & \mathbf{L} & \mathbf{F} \\ \widehat{\mathbf{B}}^{(i)} + \mathbf{X}_i & \mathbf{B}^{(i-1)} + \mathbf{X}_{i-1} & \cdots & \mathbf{B}^{(1)} + \mathbf{X}_1 & \mathbf{L} + \mathbf{X}_0 \end{bmatrix}. \tag{12}$$

Let $\overline{\boldsymbol{\pi}}$ be the stationary probability vector of the CTMC with infinitesimal generator $\overline{\mathbf{Q}}$ and $\boldsymbol{\pi}_{\mathcal{A}}$ the steady-state probability vector of the CTMC of states in $\mathcal{A}$ in the original process, i.e., the process with infinitesimal generator $\mathbf{Q}_{GI/M/1}$. There is a linear relation between $\overline{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}_{\mathcal{A}}$ given in the following equation:

$$\overline{\boldsymbol{\pi}} = \frac{\boldsymbol{\pi}_{\mathcal{A}}}{\boldsymbol{\pi}_{\mathcal{A}} \cdot \mathbf{1}^T}. \tag{13}$$

Since $\overline{\boldsymbol{\pi}}\overline{\mathbf{Q}} = \mathbf{0}$, we obtain the following relation

$$\overline{\boldsymbol{\pi}}^{(i)} \cdot (\mathbf{L} + \mathbf{X}_0) = -\overline{\boldsymbol{\pi}}^{(i-1)} \cdot \mathbf{F}$$

implying:

$$\boldsymbol{\pi}^{(i)} \cdot (\mathbf{L} + \mathbf{X}_0) = -\boldsymbol{\pi}^{(i-1)} \cdot \mathbf{F}.$$

The above equation holds for any $i > 1$, because their matrix coefficients do not depend on i. By applying it recursively over all vectors $\boldsymbol{\pi}^{(i)}$ for $i > 1$, we obtain the following geometric relation

$$\boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{R}^{i-1} \quad \forall\ i \geq 1.$$

Matrix $\mathbf{R}$, the geometric coefficient, has an important probabilistic interpretation: the entry (k, l) of $\mathbf{R}$ is the expected time spent in the state l of $\mathcal{S}^{(i)}$, before the first visit into $\mathcal{S}^{(i-1)}$, expressed in time unit Δ^i, given the starting state is k in $\mathcal{S}^{(i-1)}$. Δ^i is the mean sojourn time in the state k of $\mathcal{S}^{(i-1)}$ for $i \geq 2$ [25, pages 30-35].

3 Why M/G/1 processes are more difficult

For M/G/1-type processes there is no geometric relation among the various probability vectors $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$ as in the case of QBD and GI/M/1-type processes. In this section, we first demonstrate via a simple example why such a geometric relation does not exist and then we generalize and derive Ramaswami's recursive formula, i.e., the classic methodology for the solution of M/G/1 chains.

3.1 Example: a $BMAP_1/Cox_2/1$ queue

Figure 3 illustrates a Markov chain that models a $BMAP_1/Cox_2/1$ queue. This chain is very similar with the one depicted in Figure 1, the only difference is that the new chain models bulk arrivals of unlimited size. The infinitesimal generator $\mathbf{Q}_{M/G/1}$ of the process is block partitioned according to the partitioning of the state space $\mathcal{S}$ of this CTMC into subsets $\mathcal{S}^{(0)} = \{(0,0)\}$ and $\mathcal{S}^{(i)} = \{(i,1),(i,2)\}$ for $i \geq 1$. The definition of the component matrices of $\mathbf{Q}_{M/G/1}$ is as follows:

$$\widehat{\mathbf{L}} = [-2\lambda], \qquad \widehat{\mathbf{B}} = \begin{bmatrix} 0.2\mu \\ \gamma \end{bmatrix}, \qquad \widehat{\mathbf{F}}^{(i)} = \begin{bmatrix} 0.5^{i-1}\lambda & 0 \end{bmatrix} \; i \geq 1,$$

$$\mathbf{B} = \begin{bmatrix} 0.2\mu & 0 \\ \gamma & 0 \end{bmatrix}, \mathbf{L} = \begin{bmatrix} -(2\lambda+\mu) & 0.8\mu \\ 0 & -(2\lambda+\gamma) \end{bmatrix}, \mathbf{F}^{(i)} = \begin{bmatrix} 0.5^{i-1}\lambda & 0 \\ 0 & 0.5^{i-1}\lambda \end{bmatrix} \; i \geq 1.$$

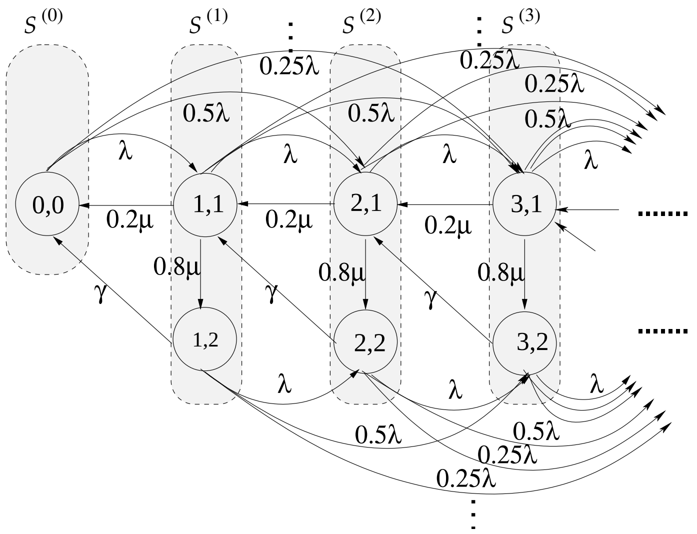

Fig. 3. The CTMC that models a $BMAP_1/Cox_2/1$ queue.

In the following, we derive the relation between $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$ and the rest of vectors in $\boldsymbol{\pi}$ using stochastic complementation, i.e., similarly to the approach described in Section 2. First we partition the state space $\mathcal{S}$ into two partitions $\mathcal{A} = \cup_{j=0}^{j=i} \mathcal{S}^{(j)}$ and $\overline{\mathcal{A}} = \cup_{j=i+1}^{\infty} \mathcal{S}^{(j)}$ and then we construct the stochastic complement of states in $\mathcal{A}$. The Markov chain of the stochastic complement of states in $\mathcal{A}$, (see Figure 4), illustrates how transitions from states $(j, 1)$ and $(j, 2)$ for $j \leq i$ and state $(0, 0)$ to states $(l, 1)$ and $(l, 2)$ for $l > i$ are folded back to state $(i, 1)$, which is the single state to enter $\mathcal{A}$ from states in $\overline{\mathcal{A}}$. These "back-folded" transitions are marked by $x_{k,h}$ for $k \leq i$ and $h = 1, 2$ and represent the "correction" needed to make the stochastic complement of states in $\mathcal{A}$, a stand-alone process. Because of the single entry state in $\mathcal{A}$ the stochastic complement of states in $\mathcal{A}$ for this example can be explicitly derived (see Lemma 1 in Appendix A) and the definition of rates $x_{k,h}$ is as follows:

$$x_{k,h} = 2 \cdot 0.5^{i-k}\lambda = 0.5^{i-k-1}\lambda, \quad i \geq 1, \quad k \leq i, \quad h = 1, 2.$$

The flow balance equations for states $(i, 1)$ and $(i, 2)$ for the stochastic complement of states in $\mathcal{A}$ are:

$$\begin{aligned}(0.2\mu + 0.8\mu)\overline{\boldsymbol{\pi}}^{(i)}[1] = {} & 2\lambda\overline{\boldsymbol{\pi}}^{(i)}[2] \\ & + 2 \cdot 0.5^{i-1}\lambda\overline{\boldsymbol{\pi}}^{(0)}[1] \\ & + 2 \cdot 0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}\ [1] + 0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}\ [2] + ... \\ & + 2 \cdot 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[1] + 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2]\end{aligned}$$

and

$$(2\lambda + \gamma)\overline{\boldsymbol{\pi}}^{(i)}[2] = 0.8\mu\overline{\boldsymbol{\pi}}^{(i)}[1] + 0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}[2] + ... + 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2].$$

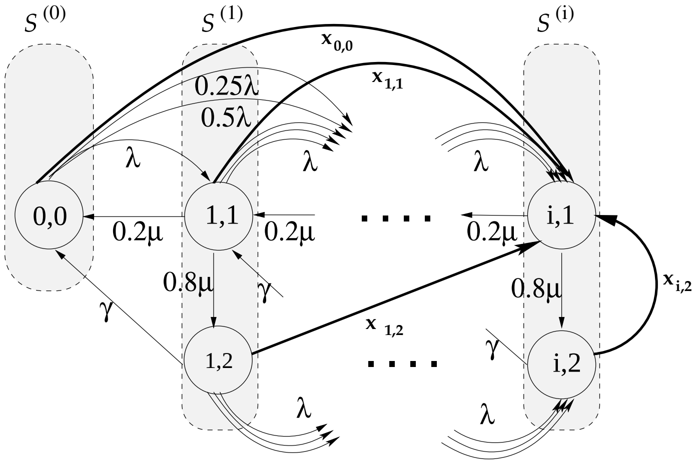

Fig. 4. Stochastic complement of $BMAP_1/Cox_2/1$-type queue at level i.

In the above equalities we group the elements of $\overline{\boldsymbol{\pi}}^{(i)}$ on the left and the rest of the terms on the right in order to express their relation in terms of the block matrices that describe the infinitesimal generator $\overline{\mathbf{Q}}$ of the stochastic complement of states in $\mathcal{A}$. By rearranging the terms, the above equations can be re-written as:

$$\begin{aligned}-\mu\overline{\boldsymbol{\pi}}^{(i)}[1] + 2\lambda\overline{\boldsymbol{\pi}}^{(i)}[2] = -(&2\cdot 0.5^{i-1}\lambda\overline{\boldsymbol{\pi}}^{(0)}[1] \\ &+ 2\cdot 0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}\ \ [1] + 0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}\ \ [2] + ... \\ &+ 2\cdot 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[1] + 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2])\end{aligned}$$

and

$$0.8\mu\overline{\boldsymbol{\pi}}^{(i)}[1] - (2\lambda+\gamma)\overline{\boldsymbol{\pi}}^{(i)}[2] = -(0.5^{i-2}\lambda\overline{\boldsymbol{\pi}}^{(1)}[2] + ... + 0.5^{i-i}\lambda\overline{\boldsymbol{\pi}}^{(i-1)}[2]).$$

We can now re-write the above equations in the following matrix equation form:

$$\begin{aligned}[\overline{\boldsymbol{\pi}}^{(i)}[1],\ \overline{\boldsymbol{\pi}}^{(i)}[2]]\cdot\begin{bmatrix}-\mu & 0.8\mu\\ 2\lambda & -(2\lambda+\gamma)\end{bmatrix} = -\Big(&\overline{\boldsymbol{\pi}}^{(0)}[1]\begin{bmatrix}2\cdot 0.5^{i-1}\lambda & 0\end{bmatrix} \\ &+ [\overline{\boldsymbol{\pi}}^{(1)}\ \ [1],\ \overline{\boldsymbol{\pi}}^{(1)}\ \ [2]]\begin{bmatrix}2\cdot 0.5^{i-2}\lambda & 0\\ 0.5^{i-2}\lambda & 0.5^{i-2}\lambda\end{bmatrix} + ... \\ &+ [\overline{\boldsymbol{\pi}}^{(i-1)}[1],\ \overline{\boldsymbol{\pi}}^{(i-1)}[2]]\begin{bmatrix}2\cdot 0.5^{i-i}\lambda & 0\\ 0.5^{i-i}\lambda & 0.5^{i-i}\lambda\end{bmatrix}\Big).\end{aligned}$$

By substituting $[\overline{\boldsymbol{\pi}}^{(i)}[1],\ \overline{\boldsymbol{\pi}}^{(i)}[2]]$ with $\overline{\boldsymbol{\pi}}^{(i)}$ and expressing the coefficient matrices in the above equation in terms of the component matrices of the infinitesimal generator $\overline{\mathbf{Q}}$ of the stochastic complement of states in $\mathcal{A}$, we obtain[4]:

$$\overline{\boldsymbol{\pi}}^{(i)} \cdot (\mathbf{L} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}\mathbf{G}) = -(\overline{\boldsymbol{\pi}}^{(0)} \sum_{j=i}^{\infty} \widehat{\mathbf{F}}^{(j)}\mathbf{G} + \overline{\boldsymbol{\pi}}^{(1)} \sum_{j=i-1}^{\infty} \mathbf{F}^{(j)}\mathbf{G} + \ldots + \overline{\boldsymbol{\pi}}^{(i-1)} \sum_{j=1}^{\infty} \mathbf{F}^{(j)}\mathbf{G}),$$

where $\mathbf{G}$ is a matrix with the following structure:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Note that at this point, we have introduced a new matrix, $\mathbf{G}$, that has an important probabilistic interpretation. In this specific example, the matrix $\mathbf{G}$ can be explicitly derived [28]. This is a direct outcome of the fact that all states in set $\mathcal{S}^{(i)}$ for $i \geq 1$ return to the same *single* state in set $\mathcal{S}^{(i-1)}$. Equivalently, the matrix $\mathbf{B}$ of the infinitesimal generator $\mathbf{Q}_{M/G/1}$ has only a single column different from zero.

3.2 Generalization: Derivation of Ramaswami's Recursive formula

In this section, we investigate the relation between $\boldsymbol{\pi}^{(i)}$ for $i > 1$ and $\boldsymbol{\pi}^{(j)}$ for $0 \leq j < i$ for the general case in the same spirit as [34]. We construct the stochastic complementation of the states in $\mathcal{A} = \cup_{j=0}^{i} \mathcal{S}^{(j)}$ ($\overline{\mathcal{A}} = \mathcal{S} - \mathcal{A}$). We obtain

$$\mathbf{Q}[\mathcal{A},\mathcal{A}] = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \cdots & \widehat{\mathbf{F}}^{(i-1)} & \widehat{\mathbf{F}}^{(i)} \\ \widehat{\mathbf{B}} & \mathbf{L} & \cdots & \widehat{\mathbf{F}}^{(i-2)} & \widehat{\mathbf{F}}^{(i-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} & \mathbf{L} \end{bmatrix}, \mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}] = \begin{bmatrix} \widehat{\mathbf{F}}^{(i+1)} & \widehat{\mathbf{F}}^{(i+2)} & \widehat{\mathbf{F}}^{(i+3)} & \cdots \\ \mathbf{F}^{(i)} & \mathbf{F}^{(i+1)} & \mathbf{F}^{(i+2)} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \mathbf{F}^{(4)} & \cdots \\ \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots \end{bmatrix},$$

$$\mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}] = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \qquad \mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}] = \begin{bmatrix} \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots \\ \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The stochastic complement for states in $\mathcal{A}$ has an infinitesimal generator defined as follows

$$\overline{\mathbf{Q}} = \mathbf{Q}[\mathcal{A},\mathcal{A}] + \mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}] \cdot (-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1} \cdot \mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}].$$

[4] Recall that $\overline{\boldsymbol{\pi}}$ is the stationary probability vector of the stochastic complement of states in $\mathcal{A}$ and $\boldsymbol{\pi}[\mathcal{A}]$ is the stationary probability vector of states in $\mathcal{A}$ in the original M/G/1 process. They relate to each other based on the equation $\overline{\boldsymbol{\pi}} = \boldsymbol{\pi}[\mathcal{A}]/(\boldsymbol{\pi}[\mathcal{A}]\mathbf{1}^T)$, which implies that any relation that holds among subvectors $\overline{\boldsymbol{\pi}}^{(j)}$ for $j \leq i$ would hold for subvectors $\boldsymbol{\pi}^{(j)}$ for $j \leq i$ as well

Observe that $\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}]$ is the same matrix for any $i \geq 1$. We define its inverse to be as follows

$$(-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} & \mathbf{A}_{0,2} & \mathbf{A}_{0,3} & \cdots \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \cdots \\ \mathbf{A}_{2,0} & \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} & \cdots \\ \mathbf{A}_{3,0} & \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{14}$$

From the special structure of $\mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}]$ we conclude that the second term of the above summation is a matrix with all block entries equal to zero except the very last block column, whose block entries $\mathbf{X}_j$ are of the form:

$$\mathbf{X}_i = \sum_{k=0}^{\infty} \widehat{\mathbf{F}}^{(i+1+k)} \cdot \mathbf{A}_{k,0} \cdot \mathbf{B}$$

and

$$\mathbf{X}_j = \sum_{k=0}^{\infty} \mathbf{F}^{(j+1+k)} \cdot \mathbf{A}_{k,0} \cdot \mathbf{B}, \quad 0 \leq j < i.$$

The infinitesimal generator $\overline{\mathbf{Q}}$ of the stochastic complement of states in $\mathcal{A}$ is determined as

$$\overline{\mathbf{Q}} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \cdots & \widehat{\mathbf{F}}^{(i-1)} & \widehat{\mathbf{F}}^{(i)} + \mathbf{X}_i \\ \widehat{\mathbf{B}} & \mathbf{L} & \cdots & \mathbf{F}^{(i-2)} & \mathbf{F}^{(i-1)} + \mathbf{X}_{i-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L} & \mathbf{F}^{(1)} + \mathbf{X}_1 \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} & \mathbf{L} + \mathbf{X}_0 \end{bmatrix}. \tag{15}$$

We define $\overline{\boldsymbol{\pi}}$ to be the steady-state probability vector of the CTMC with infinitesimal generator $\overline{\mathbf{Q}}$ and $\boldsymbol{\pi}_{\mathcal{A}}$ the steady-state probability vector of the CTMC with infinitesimal generator $\mathbf{Q}_{M/G/1}$ corresponding to the states in $\mathcal{A}$. There is a linear relation between $\overline{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}_{\mathcal{A}}$:

$$\overline{\boldsymbol{\pi}} = \frac{\boldsymbol{\pi}_{\mathcal{A}}}{\boldsymbol{\pi}_{\mathcal{A}} \cdot \mathbf{1}^T}. \tag{16}$$

From the relation $\overline{\boldsymbol{\pi}}\overline{\mathbf{Q}} = \mathbf{0}$, it follows that

$$\overline{\boldsymbol{\pi}}^{(i)} \cdot (\mathbf{L} + \mathbf{X}_0) = -(\overline{\boldsymbol{\pi}}^{(0)} \cdot (\widehat{\mathbf{F}}^{(i)} + \mathbf{X}_i) + \sum_{j=1}^{i-1} \overline{\boldsymbol{\pi}}^{(j)} \cdot (\mathbf{F}^{(i-j)} + \mathbf{X}_{i-j})) \quad \forall i \geq 1$$

and

$$\boldsymbol{\pi}^{(i)} \cdot (\mathbf{L} + \mathbf{X}_0) = -(\boldsymbol{\pi}^{(0)} \cdot (\widehat{\mathbf{F}}^{(i)} + \mathbf{X}_i) + \sum_{j=1}^{i-1} \boldsymbol{\pi}^{(j)} \cdot (\mathbf{F}^{(i-j)} + \mathbf{X}_{i-j})) \quad \forall i \geq 1. \tag{17}$$

The above equation shows that there in no geometric relation between vectors $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$, however it provides a recursive relation for the computation of the steady-state probability vector for M/G/1 Markov chains. In the following, we further work on simplifying the expression of matrices $\mathbf{X}_j$ for $0 \leq j \leq i$.

From the definition of the stochastic complementation (see Appendix A) we know that an entry $[r, c]$ in $(-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}]^{-1} \cdot \mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}])$ [5] represents the probability that starting from state $r \in \overline{\mathcal{A}}$ the process enters $\mathcal{A}$ through state c. Since $\mathcal{A}$ is entered from $\overline{\mathcal{A}}$ only through states in $\mathcal{S}^{(i)}$ we can use the probabilistic interpretation of matrix $\mathbf{G}$ to figure out the entries in $(-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}]^{-1}) \cdot \mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}]$. An entry $[r, c]$ in $\mathbf{G}^j$ for $j > 0$ represents the probability that starting from state $r \in \mathcal{S}^{(i+j)}$ for $i > 0$ the process enters set $\mathcal{S}^{(i)}$ through state c. It is straightforward now to define

$$(-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}]^{-1}) \cdot \mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}] = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}^1 \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}^2 \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{18}$$

The above result simplifies the expression of $\mathbf{X}_j$ as follows

$$\mathbf{X}_i = \sum_{k=1}^{\infty} \widehat{\mathbf{F}}^{(i+k)} \cdot \mathbf{G}^k \quad \text{and} \quad \mathbf{X}_j = \sum_{k=1}^{\infty} \mathbf{F}^{(j+k)} \cdot \mathbf{G}^k, \quad 0 \leq j < i. \tag{19}$$

This is in essence Ramaswami's recursive formula. We will return to this in the following section after we elaborate on matrix $\mathbf{G}$, its implications, and its probabilistic interpretation.

4 General Solution Method for M/G/1

For the solution of M/G/1-type processes, several algorithms exist [2, 20, 26]. These algorithms first compute matrix $\mathbf{G}$ as the solution of the following equation:

$$\mathbf{B} + \mathbf{L}\mathbf{G} + \sum_{i=1}^{\infty} \mathbf{F}^{(i)} \mathbf{G}^{i+1} = \mathbf{0}. \tag{20}$$

The matrix $\mathbf{G}$ has an important probabilistic interpretation: an entry (r, c) in $\mathbf{G}$ expresses the conditional probability of the process first entering $\mathcal{S}^{(i-1)}$ through state c, given that it starts from state r of $\mathcal{S}^{(i)}$ [26, page 81][6]. Figure 5 illustrates the relation of entries in $\mathbf{G}$ for different paths of the process. From the probabilistic interpretation of $\mathbf{G}$ the following structural properties hold [26]

– if the M/G/1 process with infinitesimal generator $\mathbf{Q}_{M/G/1}$ is recurrent then $\mathbf{G}$ is row-stochastic,

[5] Only the entries of the last block column of $(-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}]^{-1}) \cdot \mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}]$ are different from zero.

[6] The probabilistic interpretation of $\mathbf{G}$ is the same for both DTMCs and CTMCs.

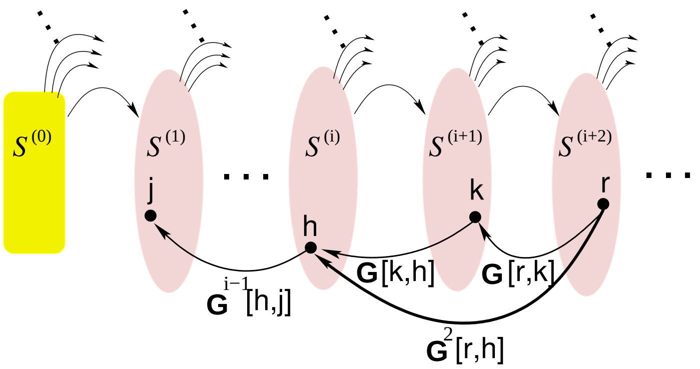

Fig. 5. Probabilistic interpretation of **G**.

– to any zero column in matrix **B** of the infinitesimal generator $\mathbf{Q}_{M/G/1}$, there is a corresponding zero column in matrix **G**.

The **G** matrix is obtained by solving iteratively Eq.(20). However, recent advances show that the computation of **G** is more efficient when displacement structures are used based on the representation of M/G/1-type processes by means of QBD processes [20, 2, 1, 16]. The most efficient algorithm for the computation of **G** is the cyclic reduction algorithm [2].

4.1 Ramaswami's Formula

From Eqs.(17) and (19) and the aid of matrix **G**, we derive Ramaswami's recursive formula [29], which is numerically stable because it entails only additions and multiplications[7]. Ramaswami's formula defines the following recursive relation among stationary probability vectors $\boldsymbol{\pi}^{(i)}$ for $i \geq 0$:

$$\boldsymbol{\pi}^{(i)} = -\left(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(i)} + \sum_{k=1}^{i-1} \boldsymbol{\pi}^{(k)}\mathbf{S}^{(i-k)}\right) \mathbf{S}^{(0)^{-1}} \quad \forall i \geq 1, \tag{21}$$

where, letting $\mathbf{F}^{(0)} \equiv \mathbf{L}$, matrices $\widehat{\mathbf{S}}^{(i)}$ and $\mathbf{S}^{(i)}$ are defined as follows:

$$\widehat{\mathbf{S}}^{(i)} = \sum_{l=i}^{\infty} \widehat{\mathbf{F}}^{(l)}\mathbf{G}^{l-i},\ i \geq 1, \quad \mathbf{S}^{(i)} = \sum_{l=i}^{\infty} \mathbf{F}^{(l)}\mathbf{G}^{l-i},\ i \geq 0. \tag{22}$$

Observe that the above auxiliary sums represent the last column in the infinitesimal generator $\overline{\mathbf{Q}}$ defined in Eq.(15). We can express them in terms of matrices $\mathbf{X}_i$ defined in Eq.(19) as follows:

$$\widehat{\mathbf{S}}^{(i)} = \widehat{\mathbf{F}}^{(i)} + \mathbf{X}_i,\ i \geq 1 \quad \mathbf{S}^{(i)} = \mathbf{F}^{(i)} + \mathbf{X}_i,\ i \geq 0.$$

[7] Subtractions on these type of formulas present the possibility of numerical instability [26, 29].

Given the above definition of $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$ and the normalization condition, a unique vector $\boldsymbol{\pi}^{(0)}$ can be obtained by solving the following system of m linear equations, i.e., the cardinality of set $\mathcal{S}^{(0)}$:

$$\boldsymbol{\pi}^{(0)} \left[\left(\widehat{\mathbf{L}}^{(0)} - \widehat{\mathbf{S}}^{(1)} \mathbf{S}^{(0)^{-1}} \widehat{\mathbf{B}} \right)^{\diamond} \mid \mathbf{1}^T - \sum_{i=1}^{\infty} \widehat{\mathbf{S}}^{(i)} \left(\sum_{j=0}^{\infty} \mathbf{S}^{(j)} \right)^{-1} \mathbf{1}^T \right] = [\mathbf{0} \mid 1], \quad (23)$$

where the symbol "$\diamond$" indicates that we discard one (any) column of the corresponding matrix, since we added a column representing the normalization condition. Once $\boldsymbol{\pi}^{(0)}$ is known, we can then iteratively compute $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$, stopping when the accumulated probability mass is close to one. After this point, measures of interest can be computed. Since the relation between $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$ is not straightforward, computation of measures of interest requires generation of the whole stationary probability vector.

4.2 Special Case: Explicit computation of G

A special case of M/G/1-type processes occurs when $\mathbf{B}$ is a product of two vectors, i.e., $\mathbf{B} = \boldsymbol{\alpha} \cdot \boldsymbol{\beta}$. Assuming, without loss of generality, that $\boldsymbol{\beta}$ is normalized, then $\mathbf{G} = \mathbf{1}^T \cdot \boldsymbol{\beta}$, i.e., it is derived explicitly [28, 30].

For this special case, $\mathbf{G} = \mathbf{G}^n$, for $n \geq 1$. This special structure of matrix $\mathbf{G}$ simplifies the form of matrices $\widehat{\mathbf{S}}^{(i)}$ for $i \geq 1$, and $\mathbf{S}^{(i)}$ for $i \geq 0$ defined in Eq.(22):

$$\begin{aligned} \widehat{\mathbf{S}}^{(i)} &= \widehat{\mathbf{F}}^{(i)} + (\textstyle\sum_{j=i+1}^{\infty} \widehat{\mathbf{F}}^{(j)}) \cdot \mathbf{G}, \quad i \geq 1 \\ \mathbf{S}^{(i)} &= \widehat{\mathbf{F}}^{(i)} + (\textstyle\sum_{j=i+1}^{\infty} \mathbf{F}^{(j)}) \cdot \mathbf{G}, \quad i \geq 0, \mathbf{F}^{(0)} \equiv \mathbf{L}. \end{aligned} \quad (24)$$

The major gain of this special case is the fact that $\mathbf{G}$ does not need to be either computed or fully stored.

4.3 Fast FFT Ramaswami's Formula

[19] gives an improved version of Ramaswami's formula. Once $\boldsymbol{\pi}^{(0)}$ is known using Eq.(23), the stationary probability vector is computed using matrix-generating functions associated with block triangular Toeplitz matrices[8]. These matrix-generating functions are computed efficiently by using fast Fourier transforms (FFT).

The algorithm of the Fast FFT Ramaswami's formula is based on the fact that in practice it is not possible to store an infinite number of matrices to express the M/G/1-type process. Assuming that only p matrices can be stored

[8] A Toeplitz matrix has equal elements in each of its diagonals allowing the use of computationally efficient methods.

then the infinitesimal generator $\mathbf{Q}_{M/G/1}$ has the following structure

$$\mathbf{Q}_{M/G/1} = \begin{bmatrix} \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} & \widehat{\mathbf{F}}^{(2)} & \widehat{\mathbf{F}}^{(3)} & \widehat{\mathbf{F}}^{(4)} & \cdots & \widehat{\mathbf{F}}^{(p)} & \mathbf{0} & \mathbf{0} & \cdots \\ \widehat{\mathbf{B}} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots & \mathbf{F}^{(p-1)} & \mathbf{F}^{(p)} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots & \mathbf{F}^{(p-2)} & \mathbf{F}^{(p-1)} & \mathbf{F}^{(p)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots & \mathbf{F}^{(p-3)} & \mathbf{F}^{(p-2)} & \mathbf{F}^{(p-1)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \mathbf{F}^{(3)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{L} & \mathbf{F}^{(1)} & \mathbf{F}^{(2)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} & \mathbf{L} & \mathbf{F}^{(1)} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B} & \mathbf{L} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (25)$$

Because there are only p matrices of type $\widehat{\mathbf{F}}^{(i)}$ and $\mathbf{F}^{(i)}$, there are only p sums of type $\widehat{\mathbf{S}}^{(i)}$ and $\mathbf{S}^{(i)}$ to be computed. Therefore, the computation of $\boldsymbol{\pi}^{(i)}$ for $i > 0$ using Ramaswami's formula, i.e., Eq.(21), depends only on p vectors $\boldsymbol{\pi}^{(j)}$ for $max(0, i-p) \leq j < i$. Define

$$\tilde{\boldsymbol{\pi}}^{(1)} = [\boldsymbol{\pi}^{(1)}, ..., \boldsymbol{\pi}^{(p)}] \quad \text{and} \quad \tilde{\boldsymbol{\pi}}^{(i)} = [\boldsymbol{\pi}^{(p(i-1)+1)}, ..., \boldsymbol{\pi}^{(pi)}] \quad \text{for} \quad i \geq 2. \quad (26)$$

The above definition simplifies the formalization of Ramaswami's formula since $\tilde{\boldsymbol{\pi}}^{(i)}$ depends only on the values of $\tilde{\boldsymbol{\pi}}^{(i-1)}$ for $i > 1$. If we apply Ramaswami's formula for vectors $\boldsymbol{\pi}^{(1)}$ to $\boldsymbol{\pi}^{(p)}$, we obtain the following equations

$$\begin{array}{l} \boldsymbol{\pi}^{(1)} = -\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(1)}(\mathbf{S}^{(0)})^{-1} \\ \boldsymbol{\pi}^{(2)} = -(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(2)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(1)})(\mathbf{S}^{(0)})^{-1} \\ \boldsymbol{\pi}^{(3)} = -(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(3)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(2)} + \boldsymbol{\pi}^{(2)}\mathbf{S}^{(1)})(\mathbf{S}^{(0)})^{-1} \\ \vdots \\ \boldsymbol{\pi}^{(p)} = -(\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(p)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(p-1)} + ... + \boldsymbol{\pi}^{(p-1)}\mathbf{S}^{(1)})(\mathbf{S}^{(0)})^{-1} \end{array}. \quad (27)$$

We rewrite the above equations in the following form:

$$\begin{array}{ll} \boldsymbol{\pi}^{(1)}\mathbf{S}^{(0)} & = -\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(1)} \\ \boldsymbol{\pi}^{(2)}\mathbf{S}^{(0)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(1)} & = -\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(2)} \\ \boldsymbol{\pi}^{(3)}\mathbf{S}^{(0)} + \boldsymbol{\pi}^{(2)}\mathbf{S}^{(1)} + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(2)} & = -\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(3)} \\ \vdots & \\ \boldsymbol{\pi}^{(p)}\mathbf{S}^{(0)} + \boldsymbol{\pi}^{(p-1)}\mathbf{S}^{(1)} + ... + \boldsymbol{\pi}^{(1)}\mathbf{S}^{(p-1)} & = -\boldsymbol{\pi}^{(0)}\widehat{\mathbf{S}}^{(p)} \end{array}. \quad (28)$$

Define

$$\mathbf{Y} = \begin{bmatrix} \mathbf{S}^{(0)} & \mathbf{S}^{(1)} & \mathbf{S}^{(2)} & \cdots & \mathbf{S}^{(p-1)} \\ \mathbf{0} & \mathbf{S}^{(0)} & \mathbf{S}^{(1)} & \cdots & \mathbf{S}^{(p-2)} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}^{(0)} & \cdots & \mathbf{S}^{(p-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}^{(0)} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \left[\widehat{\mathbf{S}}^{(1)},\ \widehat{\mathbf{S}}^{(2)},\ \widehat{\mathbf{S}}^{(3)}, \cdots,\ \widehat{\mathbf{S}}^{(p)}\right]. \quad (29)$$

The set of equations in Eq.(28) can be written in a compact way by using the definitions in Eq.(26) and Eq.(29).

$$\tilde{\boldsymbol{\pi}}^{(1)} = -\boldsymbol{\pi}^{(0)} \cdot \mathbf{b} \cdot \mathbf{Y}^{-1}. \qquad (30)$$

We apply Ramswami's formula for all vectors $\boldsymbol{\pi}^{(j)}$, $p(i-1)+1 \le j \le pi$ in $\tilde{\boldsymbol{\pi}}^{(i)}$ for $i > 1$.

$$\begin{array}{l}
\boldsymbol{\pi}^{(p(i-1)+1)} = -(\boldsymbol{\pi}^{(p(i-2)+1)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1))}\mathbf{S}^{(1)}) \quad (\mathbf{S}^{(0)})^{-1} \\
\boldsymbol{\pi}^{(p(i-1)+2)} = -(\boldsymbol{\pi}^{(p(i-2)+2)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1)+1)}\mathbf{S}^{(1)}) \quad (\mathbf{S}^{(0)})^{-1} \\
\boldsymbol{\pi}^{(p(i-1)+3)} = -(\boldsymbol{\pi}^{(p(i-2)+3)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1)+2)}\mathbf{S}^{(1)}) \quad (\mathbf{S}^{(0)})^{-1} \\
\vdots \\
\boldsymbol{\pi}^{(p(i-1)+p)} = -(\boldsymbol{\pi}^{(p(i-2)+p)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1)+p-1)}\mathbf{S}^{(1)}) \quad (\mathbf{S}^{(0)})^{-1}
\end{array}.$$

These equations can be rewritten in the following form

$$\begin{array}{ll}
\boldsymbol{\pi}^{(p(i-1)+1)}\mathbf{S}^{(0)} & = -(\boldsymbol{\pi}^{(p(i-2)+1)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1))}\mathbf{S}^{(1)}) \\
\boldsymbol{\pi}^{(p(i-1)+2)}\mathbf{S}^{(0)} + \boldsymbol{\pi}^{(p(i-1)+1)}\mathbf{S}^{(1)} & = -(\boldsymbol{\pi}^{(p(i-2)+2)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1))}\mathbf{S}^{(2)}) \\
\boldsymbol{\pi}^{(p(i-1)+3)}\mathbf{S}^{(0)} + ... + \boldsymbol{\pi}^{(p(i-1)+1)}\mathbf{S}^{(2)} & = -(\boldsymbol{\pi}^{(p(i-2)+3)}\mathbf{S}^{(p)} + ... + \boldsymbol{\pi}^{(p(i-1))}\mathbf{S}^{(3)}) \\
\vdots & \\
\boldsymbol{\pi}^{(p(i-1)+p)}\mathbf{S}^{(0)} + ... + \boldsymbol{\pi}^{(p(i-1)+1)}\mathbf{S}^{(p-1)} & = -\boldsymbol{\pi}^{(p(i-1))}\mathbf{S}^{(p)}
\end{array}.$$

The above set of equations can be written in a matrix form as

$$\tilde{\boldsymbol{\pi}}^{(i)} = -\tilde{\boldsymbol{\pi}}^{(i-1)} \cdot \mathbf{Z}\mathbf{Y}^{-1} \qquad i \ge 2, \qquad (31)$$

where matrix $\mathbf{Y}$ is defined in Eq.(29) and the definition of matrix $\mathbf{Z}$ is given by the following

$$\mathbf{Z} = \begin{bmatrix}
\mathbf{S}^{(p)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{S}^{(p-1)} & \mathbf{S}^{(p)} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\mathbf{S}^{(2)} & \mathbf{S}^{(3)} & \cdots & \mathbf{S}^{(p)} & \mathbf{0} \\
\mathbf{S}^{(1)} & \mathbf{S}^{(2)} & \cdots & \mathbf{S}^{(p-1)} & \mathbf{S}^{(p)}
\end{bmatrix}. \qquad (32)$$

The Fast Ramaswami's Formula consists of the set of equations defined in Eq.(30) and Eq.(31). The effectiveness of the representation of the Ramaswami's formula in the form of Eq.(30) and Eq.(31) comes from the special structure of matrices $\mathbf{Y}$ and $\mathbf{Z}$. The matrix $\mathbf{Y}$ is an upper block triangular Toeplitz matrix and the matrix $\mathbf{Z}$ is a lower block triangular Toeplitz matrix. Using fast Fourier transforms one can compute efficiently the inverse of a Toeplitz matrix or the multiplication of a vector with a Toeplitz matrix [19]. Although the use of Fourier transforms for matrix operations may result in numerical instabilities, in numerous test cases the above algorithm has not experienced instability [19, 21].

4.4 ETAQA-M/G/1

ETAQA[31] is an aggregation-based technique that computes only $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$ and $\sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$ for an M/G/1-type processes by solving a finite system of linear equations. Distinctively from the classic techniques of solving M/G/1-type processes, this method recasts the problem into that of solving a finite linear system in $m + 2n$ unknowns, where m is the number of states in the boundary portion of the process and n is the number of states in each of the repetitive "levels" of the state space, and are able to obtain exact results. The ETAQA methodology uses basic, well-known results for Markov chains. Assuming that the state space $\mathcal{S}$ is partitioned into sets $\mathcal{S}^{(j)}$, $j \geq 0$, instead of evaluating the probability distribution of *all* states in each $\mathcal{S}^{(j)}$, we calculate the *aggregate* probability distribution of n classes of states $\mathcal{T}^{(i)}$, $1 \leq i \leq n$, appropriately defined (see Figure 6). The

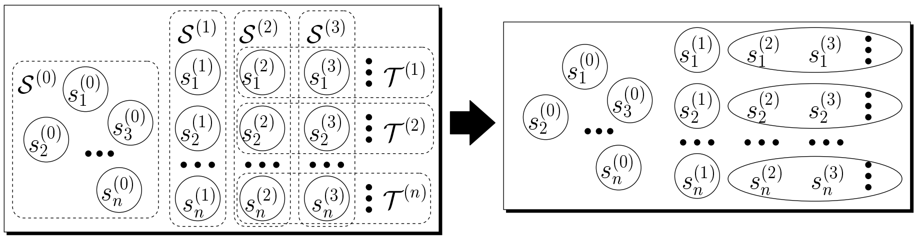

Fig. 6. Aggregation of an infinite $\mathcal{S}$ into a finite number of states.

following theorem formalizes the basic ETAQA result.

Theorem 1. [ETAQA] *Given an ergodic CTMC with infinitesimal generator* $\mathbf{Q}_{M/G/1}$ *having the structure shown in Eq.(1), with stationary probability vector* $\boldsymbol{\pi} = [\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, ...]$ *the system of linear equations*

$$\mathbf{x} \cdot \mathbf{X} = [1, \mathbf{0}] \tag{33}$$

where $\mathbf{X} \in \mathbb{R}^{(m+2n)\times(m+2n)}$ *is defined as follows*

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}^T & \widehat{\mathbf{L}} & \widehat{\mathbf{F}}^{(1)} - \sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G} & (\sum_{i=2}^{\infty} \widehat{\mathbf{F}}^{(i)} + (\sum_{i=3}^{\infty} \widehat{\mathbf{S}}^{(i)} \cdot \mathbf{G})^{\diamond} \\ \mathbf{1}^T & \widehat{\mathbf{B}} & \mathbf{L} - \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} & (\sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \sum_{i=2}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G})^{\diamond} \\ \mathbf{1}^T & \mathbf{0} & \mathbf{B} - \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G} & (\sum_{i=1}^{\infty} \mathbf{F}^{(i)} + \mathbf{L} + \sum_{i=1}^{\infty} \mathbf{S}^{(i)} \cdot \mathbf{G})^{\diamond} \end{bmatrix} \tag{34}$$

admits a unique solution $\mathbf{x} = [\boldsymbol{\pi}^{(0)},\ \boldsymbol{\pi}^{(1)},\ \boldsymbol{\pi}^{(*)}]$, *where* $\boldsymbol{\pi}^{(*)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$.

ETAQA approach is in the same spirit as the one presented in [5, 4] for the exact solution of a very limited class of QBD and M/G/1-type Markov chains, but in distinct contrast to these works, the above theorem does not require any restriction on the form of the chain's repeating pattern, thus can be applied to *any* type of M/G/1 chain.

ETAQA provides a recursive approach to compute metrics of interest once $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\boldsymbol{\pi}^{(*)}$ have been computed. We consider measures that can be expressed as the expected reward rate:

$$r = \sum_{j=0}^{\infty} \sum_{i \in \mathcal{S}^{(j)}} \boldsymbol{\rho}_i^{(j)} \boldsymbol{\pi}_i^{(j)},$$

where $\boldsymbol{\rho}_i^{(j)}$ is the *reward rate* of state $s_i^{(j)}$. For example, to compute the expected queue length in steady state, where $\mathcal{S}^{(j)}$ represents the system states with j customers in the queue, we let $\boldsymbol{\rho}_i^{(j)} = j$. To compute the second moment of the queue length, we let $\boldsymbol{\rho}_i^{(j)} = j^2$.

Since our solution approach obtains $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)}$, we rewrite r as

$$r = \boldsymbol{\pi}^{(0)} \boldsymbol{\rho}^{(0)T} + \boldsymbol{\pi}^{(1)} \boldsymbol{\rho}^{(1)T} + \sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)} \boldsymbol{\rho}^{(j)T},$$

where $\boldsymbol{\rho}^{(0)} = [\boldsymbol{\rho}_1^{(0)}, \ldots, \boldsymbol{\rho}_m^{(0)}]$ and $\boldsymbol{\rho}^{(j)} = [\boldsymbol{\rho}_1^{(j)}, \ldots, \boldsymbol{\rho}_n^{(j)}]$, for $j \geq 1$. Then, we must show how to compute the above summation without explicitly using the values of $\boldsymbol{\pi}^{(j)}$ for $j \geq 2$. We can do so if the reward rate of state $s_i^{(j)}$, for $j \geq 2$ and $i = 1, \ldots, n$, is a polynomial of degree k in j with arbitrary coefficients $\mathbf{a}_i^{[0]}, \mathbf{a}_i^{[1]}, \ldots, \mathbf{a}_i^{[k]}$:

$$\forall j \geq 2,\ \forall i \in \{1, 2, \ldots, n\}, \qquad \boldsymbol{\rho}_i^{(j)} = \mathbf{a}_i^{[0]} + \mathbf{a}_i^{[1]} j + \cdots + \mathbf{a}_i^{[k]} j^k. \tag{35}$$

The definition of $\boldsymbol{\rho}_i^{(j)}$ illustrates that the set of measures of interest that one can compute includes any moment of the probability vector $\boldsymbol{\pi}$ as long as the reward rate of the i^{th} state in each set $\mathcal{S}^{(j)}$ has the same polynomial coefficients for all $j \geq 2$.

We compute $\sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)} \boldsymbol{\rho}^{(j)T}$ as follows

$$\begin{aligned} \textstyle\sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)} \boldsymbol{\rho}^{(j)^T} &= \textstyle\sum_{j=2}^{\infty} \boldsymbol{\pi}^{(j)} \left(\mathbf{a}^{[0]} + \mathbf{a}^{[1]} j + \cdots + \mathbf{a}^{[k]} j^k \right)^T \\ &= \mathbf{r}^{[0]} \mathbf{a}^{[0]T} + \mathbf{r}^{[1]} \mathbf{a}^{[1]T} + \cdots + \mathbf{r}^{[k]} \mathbf{a}^{[k]T}, \end{aligned} \tag{36}$$

and the problem is reduced to the computation of $\mathbf{r}^{[l]} = \sum_{j=2}^{\infty} j^l \boldsymbol{\pi}^{(j)}$, for $l = 0, \ldots, k$. We show how $\mathbf{r}^{[k]}$, $k > 0$, can be computed recursively, starting from $\mathbf{r}^{[0]}$, which is simply $\boldsymbol{\pi}^{(*)}$.

$$\mathbf{r}^{[k]} \cdot [(\mathbf{B} + \mathbf{L} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)})^{\diamond} \mid (\sum_{j=1}^{\infty} j^k \mathbf{F}^{(j)} - \mathbf{B}) \mathbf{1}^T] = [\mathbf{b}^{[k]} \mid c^{[k]}], \tag{37}$$

where the definitions of $\mathbf{b}^{[k]}$, $c^{[k]}$, $\widehat{\mathbf{F}}_{[k,j,h]}$, and $\mathbf{F}_{[k,j,h]}$ are as follows

$$\mathbf{b}^{[k]} = -\left(\boldsymbol{\pi}^{(0)} \widehat{\mathbf{F}}_{[k,1,1]} + \boldsymbol{\pi}^{(1)} (2^k \mathbf{L} + \mathbf{F}_{[k,1,2]}) + \sum_{l=1}^{k} \binom{k}{l} \mathbf{r}^{[k-l]} (\mathbf{L} + \mathbf{F}_{[l,1,1]}) \right),$$

$$c^{[k]} = -(\boldsymbol{\pi}^{(0)} \sum_{j=2}^{\infty} j^k \widehat{\mathbf{F}}_{[0,j,0]} \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \sum_{j=1}^{\infty} (j+1)^k \mathbf{F}_{[0,j,0]} \mathbf{1}^T + \sum_{l=1}^{k} \binom{k}{l} \mathbf{r}^{[k-l]} \sum_{j=1}^{\infty} j^l \mathbf{F}_{[0,j,0]} \mathbf{1}^T)$$

$$\widehat{\mathbf{F}}_{[k,j,h]} = \sum_{l=j}^{\infty} (l+h)^k \cdot \widehat{\mathbf{F}}^{(l)}, \qquad \mathbf{F}_{[k,j,h]} = \sum_{l=j}^{\infty} (l+h)^k \cdot \mathbf{F}^{(l)}, \qquad j \geq 1.$$

As an example, we consider $\mathbf{r}^{[1]}$, which is used to compute measures such as the first moment of the queue length. In this case,

$$\mathbf{b}^{[1]} = -\boldsymbol{\pi}^{(0)} \sum_{j=1}^{\infty} (j+1) \widehat{\mathbf{F}}^{(j)} - \boldsymbol{\pi}^{(1)} (2\mathbf{L} + \sum_{j=1}^{\infty} (j+2) \mathbf{F}^{(j)}) - \boldsymbol{\pi}^{(*)} (\mathbf{L} + \sum_{j=1}^{\infty} (j+1) \mathbf{F}^{(j)})$$

$$c^{[1]} = -\boldsymbol{\pi}^{(0)} \sum_{j=2}^{\infty} j \widehat{\mathbf{F}}_{[0,j]} \mathbf{1}^T - \boldsymbol{\pi}^{(1)} \sum_{j=1}^{\infty} (j+1) \mathbf{F}_{[0,j]} \mathbf{1}^T - \boldsymbol{\pi}^{(*)} \sum_{j=1}^{\infty} j \mathbf{F}_{[0,j]} \mathbf{1}^T$$

5 Conditions for stability

We briefly review the conditions that enable us to assert that the CTMC described by the infinitesimal generator $\mathbf{Q}_{M/G/1}$ in Eq.(1) is stable, that is, admits a probability vector satisfying $\boldsymbol{\pi}\mathbf{Q}_{M/G/1} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{1}^T = 1$.

First observe that the matrix $\widetilde{\mathbf{Q}} = \mathbf{B} + \mathbf{L} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}$ is an infinitesimal generator, since it has zero row sums and non-negative off-diagonal entries. The conditions for stability depend on the irreducibility of matrix $\widetilde{\mathbf{Q}}$.

If $\widetilde{\mathbf{Q}}$ is irreducible then there exists a unique positive vector $\widetilde{\boldsymbol{\pi}}$ that satisfies the equations $\widetilde{\boldsymbol{\pi}}\widetilde{\mathbf{Q}} = \mathbf{0}$ and $\widetilde{\boldsymbol{\pi}}\mathbf{1}^T = 1$. In this case, the stability condition for the M/G/1-type process with infinitesimal generator $\mathbf{Q}_{M/G/1}$ [26] is given by the following inequality

$$\widetilde{\boldsymbol{\pi}}(\mathbf{L} + \sum_{j=1}^{\infty} (j+1) \mathbf{F}^{(j)}) \mathbf{1}^T = \widetilde{\boldsymbol{\pi}}(\mathbf{L} + \sum_{j=1}^{\infty} \mathbf{F}^{(j)}) \mathbf{1}^T + \widetilde{\boldsymbol{\pi}} \sum_{j=1}^{\infty} j \mathbf{F}^{(j)} \mathbf{1}^T < 0.$$

Since $\mathbf{B}+\mathbf{L}+\sum_{j=1}^{\infty} \mathbf{F}^{(j)}$ is an infinitesimal generator, then $(\mathbf{B}+\mathbf{L}+\sum_{j=1}^{\infty} \mathbf{F}^{(j)})\mathbf{1}^T = 0$. By substituting in the condition for stability the term $(\mathbf{L}+\sum_{j=1}^{\infty} \mathbf{F}^{(j)})\mathbf{1}^T$ with its equal $(-\mathbf{B}\mathbf{1}^T)$, the condition for stability can be re-written as:

$$\widetilde{\boldsymbol{\pi}} \sum_{j=1}^{\infty} j \mathbf{F}^{(j)} \mathbf{1}^T < \widetilde{\boldsymbol{\pi}} \mathbf{B} \mathbf{1}^T. \tag{38}$$

As in the scalar case, the equality in the above relation results in a null-recurrent CTMC.

In the example of the $BMAP_1/Cox_2/1$ queue depicted in Figure 3 the infinitesimal generator $\widetilde{\mathbf{Q}}$ and its stationary probability vector are

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} -0.8\mu & 0.8\mu \\ \gamma & -\gamma \end{bmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{\pi}} = \left[\frac{\gamma}{\gamma+0.8\mu}, \; \frac{0.8\mu}{\gamma+0.8\mu}\right],$$

while

$$\sum_{j=1}^{\infty} j\mathbf{F}^{(j)} = \begin{bmatrix} 4\lambda & 0 \\ 0 & 4\lambda \end{bmatrix}.$$

The stability condition is expressed as

$$\left[\frac{\gamma}{\gamma+0.8\mu}, \; \frac{0.8\mu}{\gamma+0.8\mu}\right] \cdot \begin{bmatrix} 4\lambda & 0 \\ 0 & 4\lambda \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} < \left[\frac{\gamma}{\gamma+0.8\mu}, \; \frac{0.8\mu}{\gamma+0.8\mu}\right] \cdot \begin{bmatrix} 0.2\mu & 0 \\ \gamma & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

which can be written in the following compact form

$$4\lambda < \frac{\mu \cdot \gamma}{0.8\mu + \gamma}.$$

If $\widetilde{\mathbf{Q}}$ is reducible, then the stability condition is different. By identifying the absorbing states in $\widetilde{\mathbf{Q}}$, its state space can be rearranged as follows

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_k & \mathbf{0} \\ \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_k & \mathbf{D}_0 \end{bmatrix}, \tag{39}$$

where the blocks $\mathbf{C}_h$ for $1 \leq h \leq k$ are irreducible and infinitesimal generators. Since the matrices $\mathbf{B}$, $\mathbf{L}$ and $\mathbf{F}^{(i)}$ for $i \geq 1$ have non-negative off-diagonal elements, they can be restructured similarly and have block components $\mathbf{B}_{\mathbf{C}_h}$, $\mathbf{B}_{\mathbf{D}_l}$, $\mathbf{L}_{C_h}$, $\mathbf{L}_{D_l}$, and $\mathbf{F}^{(i)}_{C_h}$, $\mathbf{F}^{(i)}_{D_l}$ for $1 \leq h \leq k$, $0 \leq l \leq k$, and $i \geq 1$.

This implies that each of the sets $\mathcal{S}^{(i)}$ for $i \geq 1$ is partitioned into subsets that communicate only through the boundary portion of the process, i.e., states in $\mathcal{S}^{(0)}$. The stability condition in Eq.(38) should be satisfied by all the irreducible blocks identified in Eq.(39) in order for the M/G/1-type process to be stable as summarized below:

$$\widetilde{\boldsymbol{\pi}}^{(h)} \sum_{j=1}^{\infty} j\mathbf{F}^{(i)}_{C_h}\mathbf{1}^T < \widetilde{\boldsymbol{\pi}}^{(h)}\mathbf{B}_{\mathbf{C}_h}\mathbf{1}^T \qquad \forall 1 \leq h \leq k. \tag{40}$$

6 MAMSolver: a matrix-analytic methods tool

In this section we briefly describe the MAMSolver [32] software tool[9] for the solution of M/G/1-type, GI/M/1-type, and QBD processes. MAMSolver is a

[9] Available at `http://www.cs.wm.edu/MAMSolver/`

collection of the most efficient solution methodologies for M/G/1-type, GI/M/1-type, and QBD processes. In contrast to existing tools such as MAGIC [36] and MGMTool [10], which provide solutions only for QBD processes, MAMSolver provides implementations for the classical and the most efficient algorithms that solve M/G/1-type, GI/M/1-type, and QBD processes. The solution provided by MAMSolver consists of the stationary probability vector for the processes under study, the queue length and queue length distribution, as well as probabilistic indicators for the queueing model such as the caudal characteristic [16].

MAMSolver provides solutions for both DTMCs and CTMCs. The matrix-analytic algorithms are defined in terms of matrices, making matrix manipulations and operations the basic elements of the tool. The input to MAMSolver, in the form of a structured text file, indicates the method to be used for the solution and the finite set of matrices that accurately describe the process to be solved. Several tests are performed within the tool to insure that special cases are treated separately and therefore efficiently.

To address possible numerical problems that may arise during matrix operations, we use well known and heavily tested routines provided by the Lapack and BLAS packages[10]. Methods such as LU-decomposition, GMRES, and BiCGSTAB are used for the solution of systems of linear equations.

The solution of **QBD** processes starts with the computation of the matrix $\mathbf{R}$ using the logarithmic reduction algorithm [16]. However for completeness we provide also the classical iterative algorithm. There are cases when $\mathbf{G}$ (and $\mathbf{R}$) can be computed explicitly [28]. We check if the conditions for the explicit computation hold in order to simplify and speedup the solution. The available solution methods for QBD processes are matrix-geometric and Etaqa.

The classic matrix geometric solution is implemented to solve **GI/M/1** processes. The algorithm goes first through the classic iterative procedure to compute $\mathbf{R}$ (to our knowledge, there is no alternative more efficient one). Then, it computes the boundary part of the stationary probability vector. Since there exists a geometric relation between vectors $\boldsymbol{\pi}^{(i)}$ for $i \geq 1$, there is no need to compute the whole stationary probability vector.

M/G/1 processes require the computation of matrix $\mathbf{G}$. More effort has been placed on efficient solution of M/G/1 processes. MAMSolver provides the classic iterative algorithm, the cyclic-reduction algorithm, and the explicit one for special cases. The stationary probability vector is computed recursively using either Ramaswami's formula or its fast FFT version. Etaqa is the other available alternative for the solution of M/G/1 processes.

For a set of input and output examples and the source code of MAMSolver, we point the interested reader to the tool's website `http://www.cs.wm.edu/MAMSolver/`.

7 Concluding Remarks

In this tutorial, we derived the basic matrix analytic results for the solution of M/G/1-type Markov processes. Via simple examples and from first principles,

[10] Available from http://www.netlib.org.

we illustrated why the solution of QBD and GI/M/1-type processes is simpler than the solution of M/G/1-type processes. We direct the interested reader in the two books of Neuts [25, 26] for further reading as well as to the book of Latouche and Ramaswami [16]. Our target was to present enough material for a modeler to solve performance models with embedded Markov chains of the M/G/1 form.

References

1. D. A. Bini and B. Meini. Using displacement structure for solving non-skip-free M/G/1 type Markov chains. In A. Alfa and S. Chakravarthy, editors, *Advances in Matrix Analytic Methods for Stochastic Models*, pages 17–37, Notable Publications Inc. NJ, 1998.
2. D. A. Bini, B. Meini, and V. Ramaswami. Analyzing M/G/1 paradigms through QBDs: the role of the block structure in computing the matrix G. In G. Latouche and P. Taylor, editors, *Advances in Matrix-Analytic Methods for Stochastic Models*, pages 73–86, Notable Publications Inc. NJ, 2000.
3. L. Breuer. Parameter estimation for a class of BMAPs. In G. Latouche and P. Taylor, editors, *Advances in Matrix-Analytic Methods for Stochastic Models*, pages 87–97, Notable Publications Inc. NJ, 2000.
4. G. Ciardo, A. Riska, and E. Smirni. An aggregation-based solution method for M/G/1-type processes. In B. Plateau, W. J. Stewart, and M. Silva, editors, *Numerical Solution of Markov Chains*, pages 21–40. Prensas Universitarias de Zaragoza, Zaragoza, Spain, 1999.
5. G. Ciardo and E. Smirni. ETAQA: an efficient technique for the analysis of QBD processes by aggregation. *Performance Evaluation*, vol. 36-37, pages 71–93, 1999.
6. J. N. Daige and D. M. Lucantoni. Queueing systems having phase-dependent arrival and service rates. In J. W. Stewart, editor, *Numerical Solution of Markov Chains*, pages 179–215, Marcel Dekker, New York, 1991.
7. H. R. Gail, S. L. Hantler, and B. A. Taylor. Use of characteristic roots for solving infinite state Markov chains. In W. K. Grassmann, editor, *Computational Probability*, pages 205–255, Kluwer Academic Publishers, Boston, MA, 2000.
8. W. K. Grassmann and D. A. Stanford. Matrix analytic methods. In W. K. Grassmann, editor, *Computational Probability*, pages 153–204, Kluwer Academic Publishers, Boston, MA, 2000.
9. D. Green. Lag correlation of approximating departure process for MAP/PH/1 queues. In G. Latouche and P. Taylor, editors, *Advances in Matrix-Analytic Methods for Stochastic Models*, pages 135–151, Notable Publications Inc. NJ, 2000.
10. B. Haverkort, A. Van Moorsel, and A. Dijkstra. MGMtool: A Performance Analysis Tool Based on Matrix Geometric Methods. In R. Pooley, and J. Hillston, editors, *Modelling Techniques and Tools*, pages 312–316, Edinburgh University Press, 1993.
11. D. Heyman and A. Reeves. Numerical solutions of linear equations arising in Markov chain models. *ORSA Journal on Computing*, vol. 1 pages 52–60, 1989.
12. D. Heyman and D. Lucantoni. Modeling multiple IP traffic streams with rate limits. In Proceedings of the 17^{th} International Teletraffic Congress, Brazil, Dec. 2001.
13. L. Kleinrock. Queueing systems. Volume I: Theory, Wiley, 1975.
14. G. Latouche. A simple proof for the matrix-geometric theorem. *Applied Stochastic Models and Data Analysis*, vol. 8, pages 25–29, 1992.

15. G. Latouche and G. W. Stewart. Numerical methods for M/G/1 type queues. In G. W. Stewart, editor, *Computations with Markov chains*, pages 571–581, Kluwer Academic Publishers, Boston, MA, 1995.
16. G. Latouche and V. Ramaswami. *Introduction to Matrix Geometric Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA, 1999.
17. D. M. Lucantoni. The BMAP/G/1 queue: A tutorial. In L. Donatiello and R. Nelson, editors, *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, pages 330–358. Springer-Verlag, 1993.
18. D. M. Lucantoni. *An algorithmic analysis of a communication model with retransmission of flawed messages*. Pitman, Boston, 1983.
19. B. Meini. An improved FFT-based version of Ramaswami's formula. *Comm. Statist. Stochastic Models*, vol. 13, pages 223–238, 1997.
20. B. Meini. Solving M/G/1 type Markov chains: Recent advances and applications. *Comm. Statist. Stochastic Models*, vol. 14(1&2), pages 479–496, 1998.
21. B. Meini. *Fast algorithms for the numerical solution of structured Markov chains*. Ph.D. Thesis, Department of Mathematics, University of Pisa, 1998.
22. C. D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, vol. 31(2) pages 240–271, June 1989.
23. R. Nelson. Matrix geometric solutions in Markov models: a mathematical tutorial. Research Report RC 16777 (#742931), IBM T.J. Watson Res. Center, Yorktown Heights, NY, Apr. 1991.
24. R. Nelson. *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag, 1995.
25. M. F. Neuts. *Matrix-geometric solutions in stochastic models*. Johns Hopkins University Press, Baltimore, MD, 1981.
26. M. F. Neuts. *Structured stochastic matrices of M/G/1 type and their applications*. Marcel Dekker, New York, NY, 1989.
27. B. F. Nielsen. Modeling long-range dependent and heavy-tailed phenomena by matrix analytic methods. In *Advances in Matrix-Analytic Methods for Stochastic Models,* G. Latouche and P. Taylor, editors, Notable Publications, pages 265–278, 2000.
28. V. Ramaswami and G. Latouche. A general class of Markov processes with explicit matrix-geometric solutions. *OR Spektrum*, vol. 8, pages 209–218, Aug. 1986.
29. V. Ramaswami. A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Comm. Statist. Stochastic Models*, vol. 4, pages 183–263, 1988.
30. V. Ramaswami and J. L. Wang. A hybrid analysis/simulation for ATM performance with application to quality-of-service of CBR traffic. *Telecommunication Systems*, vol. 5, pages 25–48, 1996.
31. A. Riska and E. Smirni. An exact aggregation approach for M/G/1-type Markov chains. In the *Proceedings of the ACM International Conference on Measurement and Modeling of Computer Systems (ACM SIGMETRICS '02)*, pages 86–96, Marina Del Rey, CA, 2002.
32. A. Riska and E. Smirni. MAMSolver: a Matrix-analytic methods tools. In T. Field et al. (editors), *TOOLS 2002, LNCS 2324*, pages 205–211, Springer-Verlag, 2002.
33. A. Riska, M. S. Squillante, S.-Z. Yu, Z. Liu, and L. Zhang. Matrix-analytic analysis of a MAP/PH/1 queue fitted to web server data. 4^{th} *Conference on Matrix-Analytic Methods* (to appear), Adelaide, Australia, July 2002.

34. H. Schellhaas. On Ramaswami's algorithm for the computation of the steady state vector in Markov chains of M/G/1 type. *Comm. Statist. Stochastic Models*, vol. 6, pages 541–550, 1990.
35. M. S. Squillante. Matrix-analytic methods: Applications, results and software tools. In G. Latouche and P. Taylor, editors, *Advances in Matrix-Analytic Methods for Stochastic Models*, Notable Publications Inc. NJ, 2000.
36. M. S. Squillante. MAGIC: A computer performance modeling tool based on matrix-geometric techniques. In G. Balbo and G. Serazzi, editors, Computer Performance Evaluation: Modeling Techniques and Tools, North-Holland, Amsterdam, pages 411–425, 1992.

Appendix A: Stochastic Complementation

Here, we briefly outline the concept of stochastic complementation [22]. While [22] introduces the concept of stochastic complementation for DTMCs with finite state spaces we define it instead for the infinite case, a straightforward extension, and state the results in terms of CTMCs.

Partition the state space $\mathcal{S}$ of an ergodic CTMC with infinitesimal generator matrix $\mathbf{Q}$ and stationary probability vector $\boldsymbol{\pi}$, satisfying $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, into two disjoint subsets, $\mathcal{A}$ and $\overline{\mathcal{A}}$.

Definition 1. [22] (Stochastic complement) The stochastic complement of $\mathcal{A}$ is

$$\overline{\mathbf{Q}} = \mathbf{Q}[\mathcal{A},\mathcal{A}] + \mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}](-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1}\mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}], \tag{41}$$

where $(-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1}[r,c]$ represents the mean time spent in state $c \in \overline{\mathcal{A}}$, starting from state $r \in \overline{\mathcal{A}}$, before reaching any state in $\mathcal{A}$, and $((-\mathbf{Q}[\overline{\mathcal{A}},\overline{\mathcal{A}}])^{-1}\mathbf{Q}[\overline{\mathcal{A}},\mathcal{A}])[r,c']$ represents the probability that, starting from $r \in \overline{\mathcal{A}}$, we enter $\mathcal{A}$ through state c'. □

The stochastic complement $\overline{\mathbf{Q}}$ is the infinitesimal generator of a new CTMC which mimics the original CTMC but "skips over" states in $\overline{\mathcal{A}}$. The following theorem formalizes this concept.

Theorem 2. [22] *The stochastic complement $\overline{\mathbf{Q}}$ of $\mathcal{A}$ is an infinitesimal generator and is irreducible if $\mathbf{Q}$ is. If $\boldsymbol{\alpha}$ is its stationary probability vector satisfying $\boldsymbol{\alpha}\overline{\mathbf{Q}} = \mathbf{0}$, then $\boldsymbol{\alpha} = Norm(\boldsymbol{\pi}[\mathcal{A}])$.* □

This implies that the stationary probability distribution $\boldsymbol{\alpha}$ of the stochastic complement differs from the corresponding portion of the stationary distribution of the original CTMC $\boldsymbol{\pi}[\mathcal{A}]$ only by the constant $\boldsymbol{\pi}[\mathcal{A}]\mathbf{1}^T$, which represents the probability of being in $\mathcal{A}$ in the original CTMC.

There are cases where we can take advantage of the special structure of the CTMC and explicitly generate the stochastic complement of a set of states $\mathcal{A}$. To consider these cases, rewrite the definition of stochastic complement in Eq.(41) as

$$\overline{\mathbf{Q}} = \mathbf{Q}[\mathcal{A},\mathcal{A}] + RowSum(\mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}])\mathbf{Z}, \tag{42}$$

where $\mathbf{Z} = Norm(\mathbf{Q}[\mathcal{A}, \overline{\mathcal{A}}])\ (-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}])^{-1}\mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}]$. The r^{th} diagonal element of $RowSum(\mathbf{Q}[\mathcal{A}, \overline{\mathcal{A}}])$ represents the rate at which the set $\mathcal{A}$ is left from its r^{th} state to reach any of the states in $\overline{\mathcal{A}}$, while the r^{th} row of $\mathbf{Z}$, which sums to one, specifies how this rate should be redistributed over the states in $\mathcal{A}$ when the process eventually reenters it.

Lemma 1. **(Single entry)** If $\mathcal{A}$ can be entered from $\overline{\mathcal{A}}$ only through a single state $c \in \mathcal{A}$, the matrix $\mathbf{Z}$ defined in Eq. (42) is trivially computable: it is a matrix of zeros except for its c^{th} column, which contains all ones. □

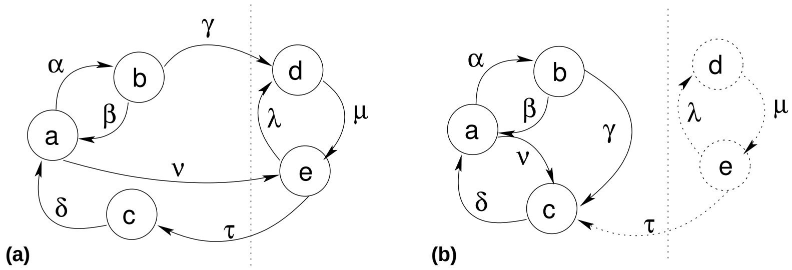

Fig. 7. Stochastic Complementation for a finite Markov chain.

We choose the simple finite Markov chain depicted in Figure 7(a) to explain the concept of stochastic complementation. The state space of this Markov chain is $\mathcal{S} = \{a, b, c, d, e\}$. We construct the stochastic complement of the states in set $\mathcal{A} = \{a, b, c\}$ $(\overline{\mathcal{A}} = \{d, e\})$, as shown in Figure 7(b). The matrices used in Eq.(41) for this example are:

$$\mathbf{Q}[\mathcal{A}, \mathcal{A}] = \begin{bmatrix} -(\alpha+\nu) & \alpha & 0 \\ \beta & -(\gamma+\beta) & 0 \\ \delta & 0 & -\delta \end{bmatrix}, \qquad \mathbf{Q}[\mathcal{A}, \overline{\mathcal{A}}] = \begin{bmatrix} 0 & \nu \\ \gamma & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \tau \end{bmatrix}, \qquad \mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}] = \begin{bmatrix} -\mu & \mu \\ \lambda & -(\lambda+\tau) \end{bmatrix}.$$

Observe that in this case one can apply Lemma 1 to trivially construct the stochastic complement, since $\mathcal{A}$ is entered from states in $\overline{\mathcal{A}}$ only through state c. There are only two transitions from states in $\mathcal{A}$ to states in $\overline{\mathcal{A}}$; the transition with rate γ from state b to state d and the transition with rate ν from state a to state e. These two transitions are folded back into $\mathcal{A}$ through state c, which is the single entry in $\mathcal{A}$. The following derivation shows that because of the special single entry state the two folded transitions have the original rates, γ and ν respectively.

$$\mathbf{Z} = Norm(\mathbf{Q}[\mathcal{A}, \overline{\mathcal{A}}])(-\mathbf{Q}[\overline{\mathcal{A}}, \overline{\mathcal{A}}])^{-1}\mathbf{Q}[\overline{\mathcal{A}}, \mathcal{A}] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{\lambda+\tau}{\mu\tau} & \frac{1}{\tau} \\ \frac{\lambda}{\mu\tau} & \frac{1}{\tau} \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \tau \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

which further results in:

$$RowSum(\mathbf{Q}[\mathcal{A},\overline{\mathcal{A}}]) \cdot \mathbf{Z} = \begin{bmatrix} \nu & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \nu \\ 0 & 0 & \gamma \\ 0 & 0 & 0 \end{bmatrix}.$$

Appendix B: Explicit computation of R

QBD processes are defined as the intersection of M/G/1 and GI/M/1-type processes. Hence, both matrix $\mathbf{G}$ (characteristic for M/G/1) and matrix $\mathbf{R}$ (characteristic for GI/M/1) can be defined for a QBD process as solutions of the following quadratic equations [16]:

$$\mathbf{B} + \mathbf{LG} + \mathbf{FG}^2 = \mathbf{0}, \qquad \mathbf{F} + \mathbf{RL} + \mathbf{R}^2\mathbf{B} = \mathbf{0}.$$

If matrix-geometric is used to solve a QBD process then the relation between $\boldsymbol{\pi}^{(i)}$ and $\boldsymbol{\pi}^{(i-1)}$ for $i > 1$ is expressed in terms of $\mathbf{R}$

$$\boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}^{(i-1}\mathbf{R},$$

If matrix-analytic is the solution method then the relation between $\boldsymbol{\pi}^{(i)}$ and $\boldsymbol{\pi}^{(i-1)}$ is based on Ramaswami's recursive formula:

$$\boldsymbol{\pi}^{(i)} = -\boldsymbol{\pi}^{(i-1)}\mathbf{S}^{(1)}(\mathbf{S}^{(0)})^{-1},$$

where $\mathbf{S}^{(1)} = \mathbf{F}$ and $\mathbf{S}^{(0)} = (\mathbf{L} + \mathbf{FG})$, i.e., the only auxiliary sums (see Subsection 4.1) used in the solution of M/G/1 processes that are defined for a QBD process. The above equations allow the derivation of the fundamental relation between $\mathbf{R}$ and $\mathbf{G}$ [16, pages 137-8],

$$\mathbf{R} = -\mathbf{F}(\mathbf{L} + \mathbf{FG})^{-1}. \tag{43}$$

Obviously, for the case of QBD processes, knowing $\mathbf{G}$ (or $\mathbf{R}$) implies a direct computation of $\mathbf{R}$ (or $\mathbf{G}$). Computing $\mathbf{G}$ is usually easier than computing $\mathbf{R}$: $\mathbf{G}$'s computation is a prerequisite to the computation of $\mathbf{R}$ in the logarithmic reduction algorithm, the most efficient algorithm to compute $\mathbf{R}$ [16]. If $\mathbf{B}$ can be expressed as a product of two vectors

$$\mathbf{B} = \boldsymbol{\alpha} \cdot \boldsymbol{\beta},$$

where, without loss of generality $\boldsymbol{\beta}$ is assumed to be a normalized vector, then $\mathbf{G}$ and $\mathbf{R}$ can be explicitly obtained as

$$\mathbf{G} = \mathbf{1} \cdot \boldsymbol{\beta}, \qquad \mathbf{R} = -\mathbf{F}(\mathbf{L} + \mathbf{F1} \cdot \boldsymbol{\beta})^{-1}.$$

Representative examples, where the above condition holds, are the queues $M/Cox/1$, $M/Hr/1$, and $M/Er/1$, whose service process is Coxian, Hyperexponential, and Erlang distribution respectively.