Solving the Single Server Semi-Markov Queue with Matrix Exponential Kernel Matrices for Interarrivals and Services

Nail Akar
Electrical and Electronics Eng.
Bilkent University
Bilkent 06800, Ankara, Turkey
akar@ee.bilkent.edu.tr

Khosrow Sohraby
School of Computing and Engineering
University of Missouri - Kansas City
Missouri 64110 USA
sohrabyk@umkc.edu

ABSTRACT

Markov renewal processes with semi-Markov kernel matrices that have matrix-exponential representations form a superset of the well-known phase-type renewal process, Markovian arrival process, and the recently introduced rational arrival process. In this paper, we study the steady-state waiting time distribution in an infinite capacity single server queue with the auto-correlation in interarrival and service times modeled with this general Markov renewal process. Our method relies on the algebraic equivalence between this waiting time distribution and the output of a feedback control system certain parameters of which are to be determined through the solution of a well known numerical linear algebra problem, namely the SDC (Spectral-Divide-and-Conquer) problem. We provide an algorithmic solution to the SDC problem and in turn obtain a simple matrix exponential representation for the waiting time distribution using the ordered Schur decomposition that is known to have numerically stable and efficient implementations in various computing platforms.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probability and Statistics—*Markov processes, Queueing theory*; G.1.3 [**Numerical Analysis**]: Numerical Linear Algebra; C.4 [**Computer Systems Organization**]: Performance of Systems—*Modeling Techniques,Performance attributes*

General Terms

Algorithms, Performance

Keywords

Lindley equation, Markov renewal processes, matrix exponential distribution, ordered Schur decomposition

1. INTRODUCTION

In this paper, we study the steady-state waiting time in a single server queue in which there is auto-correlation both in interarrival and service times. We model such auto-correlations by using a Markov renewal process (MRP) with finite state-space for both interarrivals and services. This queue is known to be the semi-Markov queue or SM/SM/1 queue in short [10]. The reference [33] studies the semi-Markov queue and introduces a non-linear matrix equation for calculating the waiting time distribution; however iterative techniques to solve the nonlinear equation have linear convergence rates. Recently, a spectral decomposition approach is described in [18] to solve the same non-linear equation via the calculation of eigenvalues and eigenvectors of a so-called coupling matrix. However, calculating eigenvectors is known to be error-prone especially for closely located eigenvalues and eigenvalues with multiplicity. Moreover, complex aritmetic is needed for the case of complex eigenvalues. On the other hand, in the matrix analytical approach pioneered by Neuts, the system occupancy is observed at certain embedded epochs and a structured Markov chain is constructed for the queue length for certain subcases of the semi-Markov queue like the (B)MAP/G/1 system [29],[23]. The key to the matrix analytical approach of Neuts is the solution to a nonlinear matrix equation which can be solved by quadratical convergence rate algorithms like the logarithmic reduction algorithm for [19] and the iterative scheme of [27] for QBD type Markov chains, and the cyclic reduction algorithm of [7], the invariant subspace approach of [1], and the technique proposed in [32] for M/G/1 type Markov chains. Given the steady state queue length probabilities, one can then calculate the waiting time distribution although not in a direct and explicit manner. For models like MAP/MAP/1, it has recently been shown that it is not necessary to compute the queue length distribution to compute waiting times. One can turn the waiting time problem to a fluid flow problem [3] using an idea originally due to Asmussen. In this case, the waiting time is phase type and its parameters can be determined by a QBD algorithm as shown in [31].

The goal of the current paper is to provide a numerically efficient and stable solution to the actual (not virtual) waiting times for the semi-Markov queue. The main assumption we make in this paper is that the semi-Markov kernels we deal with, say $F(t)$, are in matrix exponential (ME) form:

$$F(t) = V e^{t\,T} S + F, \tag{1}$$

where $F(t)$ is square of size n and T is of size m; in other

words, n is the number of states and m is the number of modes of the underlying MRP. We call such processes MRP-ME. An MRP-ME is a generalization of some well-known processes like phase-type (PH-type) [28] and ME-type renewal processes [21], Markovian arrival process (MAP) [25], rational arrival process (RAP) [24], and a subcase of MAPs with batch arrivals (BMAP) with a certain structure on the batch size distribution [22]. Assuming that n_A (n_B) and m_A (m_B) denote the number of states and the number of modes of the underlying MRP-ME for interarrivals (services), our main result is that the density of the steady state waiting time is also in matrix exponential form with $n_A m_B$ modes and finding the coefficient matrices of this ME form amounts to a particular spectral divide and conquer (SDC) problem applied on a matrix of size $n_A m_B + n_B m_A$. Given a matrix A, the SDC problem of interest in this paper is finding an orthogonal matrix Q such that

$$Q^T A Q = \begin{bmatrix} A_{++} & A_{+-} \\ 0 & A_{--} \end{bmatrix},$$

where the eigenvalues of A_{++} are exactly the same as the eigenvalues of A with positive or zero real parts. The advantages of the proposed approach are

- The approach benefits from being purely matrix analytical as explained in [28] and [29] and the generality of the model allows one to use a single unifying algorithm for different well-known queueing models.

- In case the interarrival and service MRP-MEs are PH-type or ME-type renewal processes, i.e., $n_A = n_B = 1$, then the SDC problem is to apply on a matrix of additive size $m_A + m_B$ as opposed to multiplicative size, which is a significant advantage for this special case.

- We propose the ordered Schur decomposition as a means of solving the SDC. The ordered Schur decomposition is known to be the standard serial algorithm for SDC in the numerical linear algebra literature [5] due to its numerical stability and computational efficiency.

The remainder of the paper is organized as follows. Section 2 provides preliminaries and notation used throughout the paper. Section 3 describes Markov renewal processes with matrix exponential kernel matrices and how they relate to well-known stochastic models used in queueing literature. We provide our results on the SM/SM/1 queue in Section 4. In Section 5, we present our matrix-analytical algorithm for the SM/SM/1 queue. Section 6 provides numerical examples to validate the effectiveness of the proposed approach. We conclude in the final section.

2. PRELIMINARIES AND NOTATION

We use uppercase (lowercase) letters to denote matrices (vectors or scalars). I and e denote the identity matrix and a column matrix of ones of appropriates sizes, respectively. Let $A = \{A_{ij}\}$ be an $n \times m$ matrix then the vectorized form of A is denoted by $\text{vec}(A)$:

$$\text{vec}(A) = (A_{11}, A_{12}, \ldots, A_{1m}, A_{21}, A_{22}, \ldots, A_{nm})$$

Given also a $p \times q$ matrix B, the Kronecker product of the

matrices A and B is defined as

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1m}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nm}B \end{bmatrix},$$

and the size of $A \otimes B$ is $np \times mq$. A^T denotes A transposed and the matrix A is orthogonal if $A^T A = I$. For a given real, non-symmetric matrix A and a region $\mathcal{D}$ of the complex plane, we want to find an orthogonal matrix Q such that

$$Q^T A Q = \begin{bmatrix} A_{\mathcal{D}\mathcal{D}} & A_{\mathcal{D}\bar{\mathcal{D}}} \\ 0 & A_{\bar{\mathcal{D}}\bar{\mathcal{D}}} \end{bmatrix}, \qquad (2)$$

where the eigenvalues of $A_{\mathcal{D}\mathcal{D}}$ are exactly the same as the eigenvalues of A in $\mathcal{D}$. This problem is called the ordinary spectral divide and conquer (SDC) problem [4]. A real square matrix A of size n can be transformed via an orthogonal transformation U into the so-called real Schur form by writing $U^T A U = R$ where R is quasi-upper triangular, which means that the matrix R has either 1-by-1 or 2-by-2 diagonal blocks on the diagonal corresponding to the real and complex eigenvalues, respectively, of the matrix A [12]. By reordering the blocks by orthogonal transformations, the eigenvalues are made to appear in any order and one can obtain a matrix Q such that the identity (1) is satisfied and the matrices $A_{\mathcal{D}\mathcal{D}}$ and $A_{\bar{\mathcal{D}}\bar{\mathcal{D}}}$ are quasi-upper triangular and the eigenvalues of $A_{\mathcal{D}\mathcal{D}}$ are the same as those of A in $\mathcal{D}$ [6],[8]. This form is called the ordered Schur form and the operation to obtain this form is called the ordered Schur decomposition. We note that obtaining the real Schur form is known to be backward stable and has a complexity of an^3, where a accounts for the iteration in the algorithm and may vary between 10 and 25 [11]. On the other hand, for the computation of the Jordan form that requires calculation of all eigenvectors, there are no results of guaranteed backward stability and the complexity of this decomposition is much higher than that of the Schur decomposition [11]. For this reason, the authors [11] strongly recommend not to use the Jordan decomposition whenever it is possible to use instead the more reliable Schur form which we do in the current paper. We note that this view is also shared by [5] in which the standard serial algorithm for the SDC problem is proposed to be the ordered Schur decomposition due to its well-established numerical stability. The ordered Schur form implementations are available in various platforms in LAPACK [34], MATLAB 7.0 [26], and as a public add-on to MATLAB [8]. We note two highly parallel algorithms, namely the matrix sign function algorithm and the inverse free algorithm that can also be used to solve the same SDC problem [5], but we will not use them in the current paper. Of particular interest to the current paper is when $\mathcal{D}$ is taken as the closed right half plane $\mathbb{C}^+ = \{c \in \mathbb{C} : Re(c) \geq 0\}$. We also define the complementary set $\mathbb{C}^- = \{c \in \mathbb{C} : Re(c) < 0\}$.

Let $x(t)$ be a vector function of the indeterminate variable $t \in (-\infty, +\infty)$. The two sided Laplace transform of $x(t)$ is given by

$$x^*(s) = \int_{-\infty}^{+\infty} e^{-ts} x(t) dt$$

We use the $*$ notation for Laplace transforms throughout this paper. The dirac delta function $\delta(t)$ is a commonly

used tool in engineering and it satisfies

$$\int_{-\epsilon}^{+\epsilon} \delta(t) = 1, \forall\, \epsilon > 0$$

Sometimes, we use $\int_{0^-}^{0^+} \delta(t) = 1$ to refer to the same identity. We note that existence of dirac delta functions in probability density functions is indicative of a probability mass at the origin. The degree of a polynomial $n^*(s)$ in the indeterminate s is denoted by $\deg(n^*)$. A transform is said to be rational if

$$x^*(s) = \frac{n^*(s)}{d^*(s)}$$

for some polynomials $n^*(s)$ and $d^*(s)$. The rational transform $x^*(s)$ is strictly proper if $\deg(n^*) < \deg(d^*)$ and is proper if $\deg(n^*) = \deg(d^*)$. The latter case implies a constant term in the transform and is indicative of a dirac delta function in the time domain. The poles of the rational function $x^*(s)$ are the roots of the denominator polynomial $d^*(s)$. Any strictly proper rational function $x^*(s)$ can additively be decomposed as

$$x^*(s) = x_-^*(s) + x_+^*(s)$$

where the poles of $x_-^*(s)$ and $x_-^*(s)$ reside in $\mathbb{C}^-$ and $\mathbb{C}^+$, respectively. Moreover, this decomposition is unique. If $x_+^*(s) = 0$, then $x^*(s)$ is called stable. If $x_-^*(s) = 0$, then $x^*(s)$ is called anti-stable.

A linear time-invariant dynamical system with p inputs and q outputs is represented by the following set of ordinary differential equations (ODE) [15]:

$$\frac{d}{dt}x(t) \;=\; x(t)T + u(t)V, \qquad (3)$$

$$y(t) \;=\; x(t)H + u(t)D, \qquad (4)$$

where $u(t) = (u_1(t), \ldots, u_p(t))$ and $y(t) = (y_1(t), \ldots, y_q(t))$ denote the input and output vectors, respectively, $x(t) = (x_1(t), \ldots, x_m(t))$ is called the state vector and its components are called the state variables, or simply the states. The matrices V, T, H, and D in the equations (3) and (4) are real matrices of suitable sizes. Considering zero initial state, the transfer matrix $G^*(s)$ between the input and output vectors is written as [15]:

$$y^*(s) = u^*(s)G^*(s) = u^*(s)\left(V(sI - T)^{-1}H + D\right) \qquad (5)$$

where $u^*(s)$ and $y^*(s)$ are the Laplace transforms of the input and output vectors, respectively. The equations of the form (3) and (4) are said to constitute a state-space representation or realization of the given linear time-invariant system with transfer matrix $G^*(s)$ if (5) holds [15]. The number of states (i.e., m) is referred to as the order of of the state-space representation. This representation is said to be minimal if one cannot satisfy the identity (5) with a smaller order. Using similarity transformations, one can obtain infinitely many representations whereas realization theory deals with finding state-space descriptions of linear systems and the properties of these descriptions [15].

3. MARKOV RENEWAL PROCESSES WITH MATRIX EXPONENTIAL SEMI-MARKOV KERNELS

The following is based on [9]: We define, for each $k \in \mathbb{N}$, a random variable X_k taking values in a finite set $E =$ $\{1, 2, \ldots, n\}$ and a random variable T_k taking values in $\mathbb{R}_+ = [0, \infty)$ such that $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$. The stochastic process $(X, T) = \{X_k, T_k; k \in \mathbb{N}\}$ is said to be a *Markov Renewal Process* (MRP) with state space E provided that

$$P\{X_{k+1} = j, T_{k+1} - T_k \leq t \mid X_0, \cdots, X_k; T_0, \ldots, T_k\}$$

$$= P\{X_{k+1} = j, T_{k+1} - T_k \leq t \mid X_k\}$$

for all $k \in \mathbb{N}$, $1 \leq j \leq n$, and $t \in \mathbb{R}_+$. The marginal process X_k of the MRP is called the modulating chain and $i \in E$ is called a state. On the other hand, the other marginal process $\Delta_k, k \in \mathbb{N}$ defined by $\Delta_k = T_{k+1} - T_k$ is called the modulated process. In this study, we focus our attention to the time-homogeneous case for which the probability

$$P\{X_{k+1} = j, \Delta_k \leq t \mid X_k = i\} =: F_{ij}(t) \qquad (6)$$

is independent of the customer number k. The matrix $F(t) = \{F_{ij}(t)\}$ is then called the semi-Markov kernel matrix (or simply the kernel) of the MRP. For each pair $1 \leq i, j \leq n$, the function $F_{ij}(t)$ has all the properties of a distribution function except that the quantities defined by

$$F_{ij} = \lim_{t\to\infty} F_{ij}(t)$$

are not necessarily one but instead satisfy

$$F_{ij} \geq 0, \ \sum_{j=1}^{n} F_{ij} = 1$$

We note that $F_{ij} = P\{X_{n+1} = j \mid X_n = i)$ is the state transition probability from state i to j and we assume $F = \{F_{ij}\}$ is irreducible. Let π be the stationary solution of this Discrete Time Markov Chain (DTMC) such that

$$\pi F = \pi, \pi e = 1.$$

We also note that the quantity

$$F_{ij}(t)/F_{ij} = P\{\Delta_k \leq t \mid X_{k+1} = j, X_k = i\}$$

is the sojourn time distributions in state i when the next state is j. A rich subcase of the MRP is when the kernel matrix $F(t)$ takes a matrix exponential form, i.e.,

$$F(t) = Ve^{t\,T}S + F, t \geq 0 \qquad (7)$$

and equals zero elsewhere. Here, T is square and of size m and V and S are $n \times m$ and $m \times n$, respectively. We call an MRP with a Matrix Exponential kernel matrix an MRP-ME. We also define a kernel density matrix $G(t)$ by differentiating $F(t)$ with respect to t:

$$G(t) = \frac{d}{dt}F(t) \;=\; Ve^{t\,T}TS + (F + VS)\delta(t), t \geq 0 \quad (8)$$

$$= \; Ve^{Tt}H + D\delta(t), t \geq 0 \qquad (9)$$

where $\delta(t)$ is the dirac delta function and $H = TS$ and $D = F + VS$. We also define the Laplace transform $G^*(s)$ of the kernel density matrix:

$$G^*(s) = \int_{0^-}^{\infty} e^{-ts}G(t)dt = V(sI - T)^{-1}H + D. \qquad (10)$$

An MRP-ME is then characterized by the quadruple

$$(V, T, H, D)$$

Let us assume that the sojourn times of the MRP, i.e., $\Delta_k, k \in \mathbb{N}$, are used to model interarrival or service times.

All the well-known probabilistic models described below and used in the analysis of queueing systems fall under the class of MRPs with ME kernels.

3.1 Phase-type Renewal Process

To describe a Phase-type (PH-type) distribution, we define a Markov process on the states $\{1, 2, \ldots, m, m+1\}$ with initial probability vector (v, α) and an infinitesimal generator

$$\begin{bmatrix} T & t \\ 0 & 0 \end{bmatrix}$$

where α is a scalar, v is a row vector of size m, the subgenerator T is an $m \times m$ matrix, and t is a column vector of size m such that $t = -Te$. The distribution of the time till absorption into the absorbing state $m+1$ is called a PH distribution with representation (v, T) [28]. The PH-type renewal process defined by allowing inter-renewal times modeled with a PH-type distribution is a single-state MRP-ME characterized with the quadruple (v, T, t, α).

3.2 Matrix Exponential-type Renewal Process

A renewal process with inter-renewal times having a matrix exponential distribution described in [21],[2] is a single-state MRP-ME similar to the PH-type renewal process and is again characterized by the quadruple (v, T, t, α) but the characterizing matrices do not necessarily possess the probabilistic interpretation for that of PH-type distributions. The case of $\alpha \neq 0$ can also be visualized by allowing batch arrivals at arrival epochs with a batch size distribution of $p_i = \alpha^{i-1}(1 - \alpha), i \geq 1$ and interarrival times thus modeled as an MRP-ME $(v/(1 - \alpha), T, t, 0)$.

3.3 Markovian Arrival Process

The Markovian Arrival Process (MAP) generalizes the Poisson process by allowing non-exponential interarrival times but still maintaining its Markovian structure. The MAP is described by two processes, namely the the count process $N(t)$ and the phase process $X(t)$, assuming values in $\mathbb{N}$ and $\{1, \ldots, n\}$, respectively. The two-dimensional Markov process $\{N(t), X(t)\}$ is then modeled as a Markov process on the state-space $\{(i, j) : i \in \mathbb{N}, 1 \leq j \leq n\}$ whose infinitesimal generator matrix is in block form as

$$\begin{bmatrix} D_0 & D_1 & & \cdots \\ & D_0 & D_1 & \cdots \\ & & D_0 & D_1 & \cdots \\ & & & & \ddots \end{bmatrix}. \qquad (11)$$

In the above generator, D_0 and D_1 are $n \times n$ matrices, D_0 has negative diagonal elements and non-negative off-diagonal elements, D_1 is non-negative, and $D = D_0 + D_1$ is an irreducible infinitesimal generator. When the generator is of the form (11) then the underlying process is called a MAP which is characterized by the matrix pair (D_0, D_1). Note that the MAP has the kernel

$$F(t) = (e^{D_0 t} - I)D_0^{-1}D_1$$

and is characterized by the quadruple $(I, D_0, D_1, 0)$ [14].

3.4 Rational Arrival Process

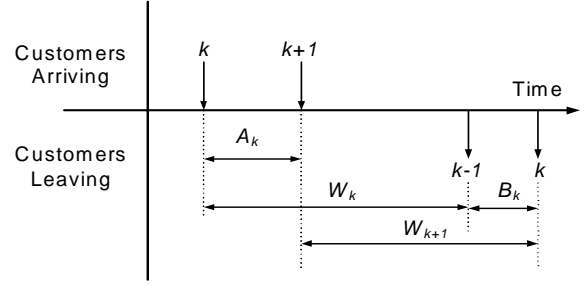The Rational Arrival Process (RAP) introduced in [24] can again be modeled as an MRP-ME characterized by the

quadruple $(I, D_0, D_1, 0)$ similar to a MAP but the matrices D_0 and D_1 do not necessarily possess the probabilistic interpretation available for MAPs.

Figure 1: Successive queue waiting times in a single server queue.

4. THE SINGLE-SERVER SEMI-MARKOV QUEUE

The successive waiting times in a single server queue with infinite capacity using a First Come First Serve (FCFS) service discipline are related through the so-called Lindley recurrence relation [16]:

$$W_{k+1} = (W_k + B_k - A_k)^+ = max(0, W_k + B_k - A_k), k \geq 0, \qquad (12)$$

where B_k and A_k denote the service time of customer k and the interarrival times between customers k and $k+1$, respectively, and W_k denotes the kth customer's waiting time in the queue. This situation is also depicted in Fig. 1 for clarity. We assume in this study that the individual processes A_k and B_k are auto-correlated but not cross-correlated and we use the MRP-ME process to model auto-correlation in interarrival and service times. We use the sojourn times of an MRP-ME to model the processes A_k and B_k, i.e., (X_k^A, T_k^A) and (X_k^B, T_k^B) are the two two-dimensional Markov renewal processes with state spaces $E_A = \{1, 2, \ldots, n_A\}$ and $E_B = \{1, 2, \ldots, n_B\}$, respectively, describing the interarrival and service times. Let G_A and G_A^* denote the kernel density matrix and its Laplace transform for the MRP-ME underlying the arrival process:

$$G_A(t) = V_A e^{t\,T_A} H_A + D_A \delta(t), \qquad (13)$$
$$G_A^*(s) = V_A (sI - T_A)^{-1} H_A + D_A. \qquad (14)$$

We note that the MRP-ME process has n_A states but the matrix T_A is square of size m_A. Let π_A be the steady-state vector of the modulating process X_k^A so that π_A satisfies

$$\pi_A(-V_A T_A^{-1} H_A + D_A) = \pi_A, \ \pi_A e = 1. \qquad (15)$$

Similarly, let G_B and G_B^* denote the kernel density matrix and its Laplace transform for the MRP-ME that models the service times:

$$G_B(t) = V_B e^{t\,T_B} H_B + D_B \delta(t), \qquad (16)$$
$$G_B^*(s) = V_B (sI - T_B)^{-1} H_B + D_B. \qquad (17)$$

We assume the service MRP-ME process has n_B states but the matrix T_B is square of size m_B. Also let π_B be the steady-state vector of the modulating process X_k^B so that π_B satisfies

$$\pi_B(-V_B T_B^{-1} H_B + D_B) = \pi_B, \ \pi_B e = 1. \qquad (18)$$

We also assume that these two state-space representations are irreducible. The queue described by the evolution equation (12) with the MRP-ME interarrival and service times described above is referred to as the semi-Markov queue. It can be shown that the mean interarrival time $E[A_k]$ and the mean service time $E[B_k]$ satisfy

$$E[A_k] = E[A] = \pi_A V_A T_A^{-2} H_A e$$

and

$$E[B_k] = E[B] = \pi_B V_B T_B^{-2} H_B e,$$

respectively. We assume throughout this paper that the load ρ defined

$$\rho = E[B]/E[A]$$

is strictly less than one. Therefore $W_k \to W$ as $k \to \infty$ in distribution, where W is called the steady-state waiting time, $F_W(t)$ and $G_W(t)$ denote its distribution and density, respectively. The Laplace transform of $G_W(t)$ is denoted by $G_W^*(s)$. In this paper, our goal is to calculate $G_W(t), t \in \mathbb{R}_+$. We first need the following theorem.

THEOREM 1. *Let X and Y be two independent non-negative random variables with ME-type densities or equivalently densities with rational Laplace transforms*

$$G_X^*(s) = D_X + \frac{N_X^*(s)}{D_X^*(s)}, \deg(N_X^*) < \deg(D_X^*)$$

and

$$G_Y^*(s) = D_Y + \frac{N_Y^*(s)}{D_Y^*(s)}, \deg(N_Y^*) < \deg(D_Y^*)$$

respectively. Then there exists a polynomial $U^*(s)$ of degree $\deg(D_Y^*)$ with $U^*(0) = 0$ so that the random variable defined by $Z = (X - Y)^+$ has an ME-type density with Laplace transform $G_Z^*(s)$ of the form

$$G_Z^*(s) = G_X^*(s)G_Y^*(-s) - \frac{U^*(s)}{D_Y^*(-s)} \qquad (19)$$

Conversely, if one can find a polynomial $U^*(s)$ with degree $\deg(D_Y^*)$ satisfying $U^*(0) = 0$ and the right hand side of (19) having all its poles in the open left half plane then identity (19) holds and gives the expression for the Laplace transform of the density of the random variable Z.

PROOF. Consider double-sided Laplace transforms [30] and recall that right-sided densities $G(t)$, i.e., $G(t) = 0, t < 0$ possess stable Laplace transforms, i.e., their poles are in the open left half plane. On the other hand, left-sided densities $G(t)$, i.e., $G(t) = 0, t > 0$, have anti-stable Laplace transforms, i.e., their poles are in the open right half plane. Note that the density of the random variable $Z_1 = X - Y$ denoted by $G_{Z_1}(t)$ is double-sided and it has the two-sided Laplace transform

$$G_{Z_1}^*(s) = G_X^*(s)G_Y^*(-s) \qquad (20)$$

but the strictly proper part of G_{Z_1} can be decomposed into its stable and anti-stable components in the following unique way:

$$G_{Z_1}^*(s) = D_X D_Y + \frac{U_1^*(s)}{D_X^*(s)} + \frac{U_2^*(s)}{D_Y^*(-s)}, \qquad (21)$$

where $\deg(U_1^*) < \deg(D_X^*)$ and $\deg(U_2^*) < \deg(D_Y^*)$. Note that the + operator removes the left side of a double-sided

density and places all the corresponding probability mass at $t = 0$. Therefore in the transform domain

$$
\begin{aligned}
G_Z^*(s) &= D_X D_Y + \frac{U_1^*(s)}{D_X^*(s)} + \frac{U_2^*(0)}{D_Y^*(0)} \\
&= G_X^*(s)G_Y^*(-s) - \frac{U_2^*(s)}{D_Y^*(-s)} + \frac{U_2^*(0)}{D_Y^*(0)} \\
&= G_X^*(s)G_Y^*(-s) - \frac{U^*(s)}{D_Y^*(-s)} \qquad (22)
\end{aligned}
$$

where

$$U^*(s) = U_2^*(s) - \frac{U_2^*(0)D_Y^*(-s)}{D_Y^*(0)}$$

Evaluating the identity (22) at $s = 0$ and noting that $D_Y^*(0)$ cannot be zero, we show that $U^*(0) = 0$ and this concludes the if part of the proof. The only if part can be proved by observing the unique spectral decomposition of a rational function into its stable and anti-stable parts and tracing back the proof of the if part. $\square$

We're now ready to study the steady-solution of the Lindley's equation (12). For this purpose, we define for $i = 1, \ldots, n_A$ and $j = 1, \ldots, n_B$

$$
\begin{aligned}
F_{W,ij}(t) &= \lim_{k \to \infty} P\{W_k \le t, X_k^A = i, X_k^B = j\} \\
&= P\{W \le t, X^A = i, X^B = j\}, \qquad (23)
\end{aligned}
$$

and

$$\tilde{F}_W(t) = \mathrm{vec}(\{F_{W,ij}(t)\})$$

Similarly, we define

$$G_{W,ij}(t) = \frac{d}{dt}F_{W,ij}(t) \qquad (24)$$

and

$$\tilde{G}_W(t) = \mathrm{vec}(\{G_{W,ij}(t)\})$$

In our analysis, the following Laplace transforms are crucial:

$$G_{W,ij}^*(s) = \int_{0^-}^{\infty} e^{-st}G_{W,ij}(t)dt, \tilde{G}_W^*(s) = \mathrm{vec}(\{G_{W,ij}^*(s)\}) \qquad (25)$$

From Lindley's equation (12) and Theorem 1, we note the existence of polynomials $U_{klij}^*(s)$, $k, i = 1, \ldots, n_A$, $l, j = 1, \ldots, n_B$ with $U_{klij}^*(0) = 0$ such that the following hold:

$$
\begin{aligned}
G_{W,ij}^*(s) &= \sum_{k=1}^{n_A}\sum_{l=1}^{n_B} G_{W,kl}^*(s)G_{A,ki}^*(-s)G_{B,lj}^*(s) \\
&\quad - \sum_{k=1}^{n_A}\sum_{l=1}^{n_B} \frac{U_{klij}^*(s)}{D_{A,ki}^*(-s)}, \qquad (26)
\end{aligned}
$$

where $D_{A,ki}^*(-s)$ is the denominator of $G_{A,ki}^*(-s)$ and its degree is the same as that of $U_{klij}^*(s)$. This identity can be shown to reduce to the existence of polynomials $U_{ij}^*(s)$ with degree m_A and $U_{ij}^*(0) = 0$ such that the second term on the right hand side of (26) can be simplified as in

$$
\begin{aligned}
G_{W,ij}^*(s) &= \sum_{k=1}^{n_A}\sum_{l=1}^{n_B} G_{W,kl}^*(s)G_{A,ki}^*(-s)G_{B,lj}^*(s) \\
&\quad - \frac{U_{ij}^*(s)}{\det(sI + T_A)} \qquad (27)
\end{aligned}
$$

Conversely, if one can find polynomials $U_{ij}^*(s)$ with degree m_A and $U_{ij}^*(0) = 0$ such that (27) holds with the right hand side of (27) is free of closed right half plane poles, i.e., all poles in the open left half plane, then the identity (27) completely describes $G_{W,ij}^*(s)$ which is what we want to find. The identity (27) can also be put into the following vector form:

$$
\begin{aligned}
\tilde{G}_W^*(s) &= \tilde{G}_W^*(s)\left(I_{n_A} \otimes G_B^*(s)\right)\left(G_A^*(-s) \otimes I_{n_B}\right) \\
&- \frac{\tilde{U}^*(s)}{\det(sI + T_A)}, \tilde{U}^*(s) = \mathrm{vec}\left(\{U_{ij}^*(s)\}\right) \quad (28)
\end{aligned}
$$

Our goal is then to find $\tilde{U}^*(s)$ with $\tilde{U}^*(0) = 0$ such that identity (28) holds for a stable transform vector $\tilde{G}_W^*(s)$. However, the identity (28) does not directly lend itself to a computational procedure for finding the unknown polynomials and even so, doing the calculations in the transform domain is cumbersome and ill-conditioned and additionally one needs to perform transform inversion in the end to calculate $\tilde{G}_W(t)$. In order to avoid transform domain calculations for finding $\tilde{U}^*(s)$, we first introduce the following linear system of differential equations defined for $t \geq 0$ (denoted by $\mathbf{S_A}$) associated with the interarrival times in state-space form but with non-zero initial states:

$$
\begin{aligned}
\frac{d}{dt}x_A(t) &= -x_A(t)\tilde{T}_A + u_A(t)\tilde{V}_A, \ x_A(0^-) = x_0, \quad (29) \\
y_A(t) &= -x_A(t)\tilde{H}_A + u_A(t)\tilde{D}_A + d_0\,\delta(t), \quad (30)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{V}_A &= V_A \otimes I_{n_B}, \\
\tilde{T}_A &= T_A \otimes I_{n_B}, \\
\tilde{H}_A &= H_A \otimes I_{n_B} \\
\tilde{D}_A &= D_A \otimes I_{n_B}, \quad (31)
\end{aligned}
$$

and $\otimes$ denotes the Kronecker product. In the linear system above, $x_A(t)$ is the system state with a non-zero initial value x_0. Moreover, the system $\mathbf{S_A}$ has two input vectors, one being the control input $u_A(t)$, the other being the dirac delta function $\delta(t)$ feeding in through d_0, and one output vector $y_A(t)$. The system parameters d_0 and y_0 are not known yet but they are to be determined. Next consider another linear system of differential equations (denoted by $\mathbf{S_B}$) associated with the service times in state-space form:

$$
\begin{aligned}
\frac{d}{dt}x_B(t) &= x_B(t)\tilde{T}_B + u_B(t)\tilde{V}_B, \ x_B(0^-) = 0, \quad (32) \\
y_B(t) &= x_B(t)\tilde{H}_B + u_B(t)\tilde{D}_B, \quad (33)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{V}_B &= I_{n_A} \otimes V_B, \\
\tilde{T}_B &= I_{n_A} \otimes T_B, \\
\tilde{H}_B &= I_{n_A} \otimes H_B \\
\tilde{D}_B &= I_{n_A} \otimes D_B \quad (34)
\end{aligned}
$$

In the linear system above, $x_B(t)$ is the system state with a zero initial value. On the other hand, the system $\mathbf{S_B}$ has one vector input $u_B(t)$ and one vector output $y_B(t)$. We interconnect these two systems, for reasons to be made clear later, through the following feedback configuration, say $\mathbf{S_F}$, also depicted in Figure 2:

$$
u_A(t) = y_B(t) =: u_F(t), \ u_B(t) = y_A(t) =: y_F(t) \quad (35)
$$
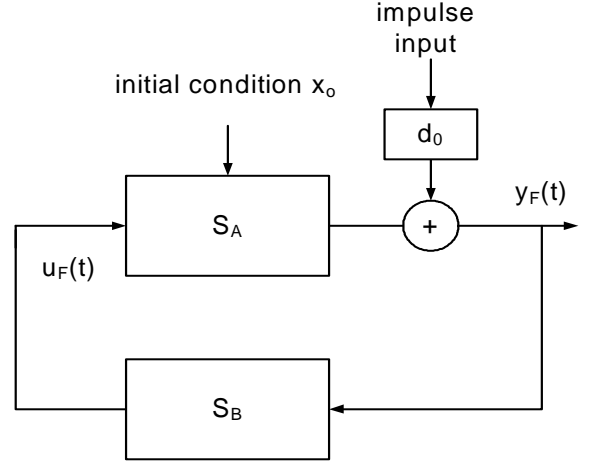
Figure 2: Feedback interconnection diagram of the two linear systems of differential equations $\mathbf{S_A}$ and $\mathbf{S_B}$

where the subscript F is used to refer to the feedback configuration.

We now obtain an expression for the the Laplace transform $y_F^*(s)$ of the output vector $y_F(t)$. Note from (30) and (35) that

$$
y_F^*(s) = -x_A^*(s)\tilde{H}_A + u_F^*(s)\tilde{D}_A + d_0.
$$

Also note from (29) and the derivative rule for Laplace transforms that

$$
x_A^*(s) = \left(u_F^*(s)\tilde{V}_A + x_0\right)(sI + \tilde{T}_A)^{-1}.
$$

Consequently,

$$
\begin{aligned}
y_F^*(s) &= -\left(u_F^*(s)\tilde{V}_A + x_0\right)(sI + \tilde{T}_A)^{-1}\tilde{H}_A \\
&+ u_F^*(s)\tilde{D}_A + d_0 \\
&= u_F^*(s)\left(-\tilde{V}_A(sI + \tilde{T}_A)^{-1}\tilde{H}_A + \tilde{D}_A\right) \\
&- \left(x_0(sI + \tilde{T}_A)^{-1}\tilde{H}_A - d_0\right) \quad (36)
\end{aligned}
$$

From (32) and (33), we know that

$$
u_F^*(s) = y_F^*(s)\left(\tilde{V}_B(sI - \tilde{T}_B)^{-1}\tilde{H}_B + \tilde{D}_B\right)
$$

from which (36) simplifies to

$$
\begin{aligned}
y_F^*(s) &= y_F^*(s)\left(I_{n_A} \otimes G_B^*(s)\right)\left(G_A^*(-s) \otimes I_{n_B}\right) \\
&- \left(x_0(sI + \tilde{T}_A)^{-1}\tilde{H}_A - d_0\right) \quad (37)
\end{aligned}
$$

Note the similarity in the identities (37) and (28). Let us assume that we determine x_0 and d_0 such that the following two conditions are satisfied:

C1) $y_F^*(s)$ is analytic in the closed right half plane

C2) $y_F^*(0) = \tilde{\pi} = \pi_A \otimes \pi_B$

then (37) provides an equivalent expression to (27) and (28) with an appropriate choice of $\tilde{U}^*(s)$. Therefore, in case one has x_0 and d_0 such that the output of the feedback system

$y_F(t)$ satisfies the conditions **C1)** and **C2)** then it is true that

$$y_F^*(s) = \tilde{G}_W^*(s)$$

The next section describes a matrix analytical method to find $\tilde{G}_W(t)$ without using any transform domain calculations but instead uses the transform identities (27) or (37) only for proving the proposed algorithmic method.

5. MATRIX ANALYTICAL METHOD FOR THE SEMI-MARKOV QUEUE

In this section, we present a matrix analytical method for solving for the steady-state waiting time distribution through the calculation of the vector $\tilde{G}_W(t)$. For this purpose, we obtain a state-space representation for the feedback configuration $\mathbf{S_F}$ given through (35). We first try to write the vector input $u_F(t)$ in terms of the individual states $x_A(t)$ and $x_B(t)$ of the systems $\mathbf{S_A}$ and $\mathbf{S_B}$, respectively:

$$
\begin{aligned}
u_F(t) &= x_B(t)\tilde{H}_B + u_B(t)\tilde{D}_B && (38)\\
&= x_B(t)\tilde{H}_B \\
&\quad + \left(-x_A(t)\tilde{H}_A + u_A(t)\tilde{D}_A + d_0\,\delta(t)\right)\tilde{D}_B \\
&= \left(-x_A(t)\tilde{H}_A\tilde{D}_B + x_B(t)\tilde{H}_B + d_0\tilde{D}_B\,\delta(t)\right)\tilde{D}_{AB}
\end{aligned}
$$

where

$$\tilde{D}_{AB} = (I - \tilde{D}_A\tilde{D}_B)^{-1} \qquad (39)$$

Similarly, the output vector $y_F(t)$ is written in terms of the individual states x_A and x_B as follows:

$$
\begin{aligned}
y_F(t) &= -x_A(t)\tilde{H}_A + u_A(t)\tilde{D}_A + d_0\,\delta(t) && (40)\\
&= -x_A(t)\tilde{H}_A \\
&\quad + \left(x_B(t)\tilde{H}_B + u_B(t)\tilde{D}_B\right)\tilde{D}_A + d_0\,\delta(t) \\
&= \left(-x_A(t)\tilde{H}_A + x_B(t)\tilde{H}_B\tilde{D}_A + d_0\,\delta(t)\right)\tilde{D}_{BA}
\end{aligned}
$$

where

$$\tilde{D}_{BA} = (I - \tilde{D}_B\tilde{D}_A)^{-1} \qquad (41)$$

By substituting (38) in the state equations (29), one can easily show that

$$
\begin{aligned}
\frac{d}{dt}x_A(t) &= x_A(t)(-\tilde{T}_A - \tilde{H}_A\tilde{D}_B\tilde{D}_{AB}\tilde{V}_A) \\
&\quad + x_B(t)\tilde{H}_B\tilde{D}_{AB}\tilde{V}_A \\
&\quad + d_0\tilde{D}_B\tilde{D}_{AB}\tilde{V}_A\,\delta(t) && (42)
\end{aligned}
$$

On the other hand, insertion of (40) in the state equations (32) leads to the following modified state equations for $x_B(t)$:

$$
\begin{aligned}
\frac{d}{dt}x_B(t) &= -x_A(t)\tilde{H}_A\tilde{D}_{BA}\tilde{V}_B \\
&\quad + x_B(t)(\tilde{T}_B + \tilde{H}_B\tilde{D}_A\tilde{D}_{BA}\tilde{V}_B) \\
&\quad + d_0\tilde{D}_{BA}\tilde{V}_B\,\delta(t) && (43)
\end{aligned}
$$

Combining the differential equations (42) and (43) we obtain

$$\frac{d}{dt}x_F(t) = (\frac{d}{dt}x_A(t), \frac{d}{dt}x_B(t)) = x_F(t)T_F + d_0 V_F\,\delta(t), \quad (44)$$

with the initial value

$$x_F(0^-) = (x_0, 0),$$

and the matrices T_F and V_F are defined as

$$
T_F = \begin{bmatrix} -\tilde{T}_A - \tilde{H}_A\tilde{D}_B\tilde{D}_{AB}\tilde{V}_A & -\tilde{H}_A\tilde{D}_{BA}\tilde{V}_B \\ \tilde{H}_B\tilde{D}_{AB}\tilde{V}_A & \tilde{T}_B + \tilde{H}_B\tilde{D}_A\tilde{D}_{BA}\tilde{V}_B \end{bmatrix},
$$
$$(45)$$

and

$$V_F = \begin{bmatrix} \tilde{D}_B\tilde{D}_{AB}\tilde{V}_A & \tilde{D}_{BA}\tilde{V}_B \end{bmatrix} \qquad (46)$$

Integrating the state equations (44) from 0^- to 0^+ (or simply 0), we obtain the following state equation

$$\frac{d}{dt}x_F(t) = x_F(t)T_F, \ x_F(0) = (x_0, 0) + d_0 V_F \qquad (47)$$

and the output equation

$$y_F(t) = x_F(t)H_F + d_0\tilde{D}_{BA}\delta(t) \qquad (48)$$

where

$$H_F = \begin{bmatrix} -\tilde{H}_A\tilde{D}_{BA} \\ \tilde{H}_B\tilde{D}_A\tilde{D}_{BA} \end{bmatrix}. \qquad (49)$$

We note that the matrix T_F has $n_B m_A - 1$ eigenvalues with positive real parts, one eigenvalue at the origin, and $n_A m_B$ eigenvalues with negative real parts [17]. Let Q_F be an orthogonal matrix, i.e., $Q_F^T Q_F = I$, such that

$$Q_F^T T_F Q_F = \begin{bmatrix} T_{F,++} & T_{F,+-} \\ 0 & T_{F,--} \end{bmatrix}, \qquad (50)$$

where the eigenvalues of the matrix $T_{F,++}$ and $T_{F,--}$ have nonnegative and negative real parts, and they are of size $n_B m_A$ and $n_A m_B$, respectively. We then define the following transformation

$$\tilde{x}_F(t) = x_F(t)Q_F \qquad (51)$$

and partition

$$\tilde{x}_F(t) = (\tilde{x}_{F,+}(t), \tilde{x}_{F,-}(t))$$

appropriately so that the row vectors $\tilde{x}_{F,+}(t)$ and $\tilde{x}_{F,-}(t)$ are of size $n_B m_A$ and $n_A m_B$, respectively. From (47) and (50) we first obtain

$$\frac{d}{dt}\tilde{x}_{F,+}(t) = \tilde{x}_{F,+}(t)T_{F,++}$$

but since all the eigenvalues of $T_{F,++}$ have nonnegative real parts the only condition for the analyticity of $y_F^*(s)$ in the closed right half plane is

$$\tilde{x}_{F,+}(0) = 0. \qquad (52)$$

Partitioning Q_F as in (50)

$$Q_F = \begin{bmatrix} Q_{F,++} & Q_{F,+-} \\ Q_{F,-+} & Q_{F,--} \end{bmatrix}, \qquad (53)$$

and defining

$$Q_{F,+} = \begin{bmatrix} Q_{F,++} \\ Q_{F,-+} \end{bmatrix}, \ Q_{F,-} = \begin{bmatrix} Q_{F,+-} \\ Q_{F,--} \end{bmatrix}, \qquad (54)$$

the condition (52) can be reduced to a linear matrix equation in the unknowns x_0 and d_0:

$$
\begin{aligned}
\tilde{x}_{F,+}(0) &= x_F(0)Q_{F,+} \\
&= ((x_0, 0) + d_0 V_F)Q_{F,+} \\
&= x_0 Q_{F,++} + d_0 V_F Q_{F,+} = 0 && (55)
\end{aligned}
$$

The equation (55) ensures that the condition **C1)** is satisfied. In order to express the condition **C2)** as a linear matrix equation we first write $\tilde{x}_F(t)$ as

$$
\begin{aligned}
\tilde{x}_{F,-}(t) &= \tilde{x}_{F,-}(0)e^{t\,T_{F,--}} \\
&= (x_0 Q_{F,+-} + d_0 V_F Q_{F,-})\, e^{t\,T_{F,--}}
\end{aligned}
$$

Finally, the expression for the output vector $y_F(t)$ in (48) can further be simplified to

$$
\begin{aligned}
y_F(t) &= (x_0\,Q_{F,+-} + d_0\,V_F\,Q_{F,-})\,e^{t\,T_{F,--}}\,Q_{F,-}^T\,H_F \\
&\quad + \; d_0\,\tilde{D}_{BA}\,\delta(t) \qquad\qquad\qquad\qquad (56)
\end{aligned}
$$

Noting that $\int_{0^-}^{\infty} y_F(t) = \tilde{\pi}$ for condition **C2)**, we obtain another linear matrix equation in x_0 and d_0 so that **C2)** is satisfied:

$$
\begin{aligned}
\tilde{\pi} &= -x_0\,Q_{F,+-}\,T_{F,--}^{-1}\,Q_{F,-}^T\,H_F \\
&\quad + \; d_0\left(-V_F\,Q_{F,-}\,T_{F,--}^{-1}\,Q_{F,-}^T\,H_F + \tilde{D}_{BA}\right) \quad (57)
\end{aligned}
$$

Combination of (55) and (57) give $(m_A + n_A)n_B$ equations with $(m_A + n_A)n_B$ unknowns; x_0 is of size $n_B m_A$ and d_0 is of size $n_A n_B$. Solving for x_0 and d_0 from (55) and (57) leads us to a matrix exponential waiting time distribution:

$$
y_F^*(t) = \tilde{G}_W(t) = v e^{t\,T} H + d\delta(t) \qquad (58)
$$

where

$$
\begin{aligned}
v &:= x_0\,Q_{F,+-} + d_0\,V_F\,Q_{F,-} & (59)\\
T &:= T_{F,--} & (60)\\
H &:= Q_{F,-}^T\,H_F & (61)\\
d &:= d_0\,\tilde{D}_{BA} & (62)
\end{aligned}
$$

Note that the density of the steady-state waiting time is written as

$$
G_W(t) = \tilde{G}_W(t)e = v e^{t\,T} He + de\,\delta(t)
$$

from which one can find the ith moment of the waiting time as follows:

$$
E[W^i] = (-1)^{i+1}\,i!\,vT^{-(i+1)}He.
$$

Although the development of the overall algorithm might be elaborate, the algorithm itself is relatively simple and is given in Table 1 for the sake of reference.

6. NUMERICAL EXAMPLE

As a simple example, we use a simple PH/PH/1 queue studied in [13]. In this example, we study an IPP/E$_k$/1 queue where the IPP (Interrupted Poisson Process) is a PH-type process with two phases, namely the OFF and ON phases, and E$_k$ denotes the Erlangian distribution with k stages [13]. The mean service rate is set to 100 in this numerical example. In an IPP, the arrivals are Poisson with rate λ in the ON phase and there are no arrivals in the OFF phase; the IPP has the following ME representation $(v_A, T_A, h_A, 0)$ given in [13]

$$
v = \begin{bmatrix} 0 & 1 \end{bmatrix}, T_A = \begin{bmatrix} -\gamma_{01} & \gamma_{01} \\ \gamma_{10} & -(\gamma_{10} + \lambda) \end{bmatrix}, h_A = \begin{bmatrix} 0 \\ \lambda \end{bmatrix}.
$$

E$_k$ distributions have natural ME representations given in [28]. The burstiness b of an IPP is defined as the ratio between the arrival rate in a burst and the overall average

Table 1: Algorithm to find $G_W(t)$ given the pair of quadruples (V_A, T_A, H_A, D_A) and (V_B, T_B, H_B, D_B) characterizing the MRP-MEs for interarrival and service times, respectively.

1.	Find the steady-state vectors π_A and π_B using (15) and (18), respectively.
2.	Let $\tilde{\pi} = \pi_A \otimes \pi_B$.
3.	Obtain $\tilde{V}_A$, $\tilde{T}_A$, $\tilde{H}_A$, and $\tilde{D}_A$ by (31).
4.	Obtain $\tilde{V}_B$, $\tilde{T}_B$, $\tilde{H}_B$, and $\tilde{D}_B$ by (34).
5.	Construct $\tilde{D}_{AB}$ and $\tilde{D}_{BA}$ by (39) and (41).
6.	Construct the matrices T_F, V_F, and H_F by using (45), (46), and (49), respectively.
7.	Find an orthogonal matrix Q_F so as to obtain the ordered Schur form of T_F where ordering is done as in (50).
8.	Partition the matrix Q_F as in (53) and (54).
9.	Solve for x_0 and d_0 using the linear equations (55) and (57).
10.	Define v, T, H, and d by (59), (60), (61), and (62).
11.	Write $G_W(t) = v e^{t\,T} He + de\,\delta(t)$.

arrival rate. In this numerical example, we fix $\gamma_{01} = 10$ and choose γ_{10} so as to fix the burstiness $b = 4$. The rate parameter λ is then chosen so as to attain a desired load ρ on the queueing system.

The algorithm of Ref. [20] uses the logarithmic reduction (LR) iterative procedure for the PH/PH/1 queue. The LR procedure was first introduced in [19]. The advantage of the algorithm [20] stems from the reduced size of the matrices that are used within the LR procedure; the order of the matrices are the sum of the phases (i.e., $m = m_A + m_B$) in the arrival and service time distributions in [20]. This is in contrast with matrices of size being their product (i.e., $m = m_A m_B$) in the original matrix-geometric algorithm given in [28]. This order reduction brings a considerable computational advantage. However, we note that calculation of the input matrices to the LR procedure in [20] still requires the construction of a matrix with the order of the product of the number of phases in the arrival and service time distributions and further matrix multiplications involving this product-sized matrix. On the other hand, the Schur decomposition-based algorithm proposed in this paper does not use matrices of multiplicative size in any step of the algorithm and the Schur form of a matrix of additive size is used. In Table 2, we report the probability mass at zero, i.e., the probability that an arriving customer finds the queue empty and hence will not wait in the queue, as well as the mean waiting time in the queue for the IPP/E$_k$/1 queue. These results are obtained using MATLAB 7.0 and the LR procedure of [20] with a stopping criterion $\varepsilon = 10^{-9}$ and the Schur-based algorithm proposed in this paper as a function of ρ and the number of stages, i.e., k, of the Erlangian service time distribution. For the latter, we use the Matlab functions `schur.m` for Schur decomposition and `ordschur.m` for ordering the eigenvalues. We observe that the results agree upto a significant number of digits validating the accuracy of the Schur-based algorithm.

We also consider correlated arrivals in the currentexample. We study the statistical multiplexing of N voice sources with silence detection so that each voice source is modeled

as a two-state IPP with mean off time and the mean on time set to 650 ms and 353 ms, respectively. The mean number of packets generated by each voice source in an on period is set to 22 and the packet sizes are such that voice peak rate in the on state is 32 Kbps. Although packet size are fixed we model them by the E_{20} distribution with the same mean. On the other hand, the multiplexer's service rate is 10 times each voice source's peak rate. In Table 3, we report the probability that an arriving customer finds the queue empty, the mean waiting time in the queue, and the CPU time needed for the voice multiplexing example as a function of the number of voice sources N using the Schur approach proposed in this paper and the matrix geometric (MG) approach outlined in [14] that also derives explicit waiting time expressions for the underlying MAP/PH/1 queue. We also note that we employ the quadratically convergent iterations proposed in [27] for finding the associated rate matrix in the MAP/PH/1 queue. All the values we calculate are identical up to six significant digits so we report them once. We also provide the CPU times needed in the Schur approach and the MG approach. The results are indicative of similar computational complexities that the two approaches possess although we are led to believe that the MG approach is slightly better for low loads and this situation is reversed for increased loads. The real advantage in using the proposed approach is that the expression for the waiting time distribution is matrix exponential and all related moments can be derived algorithmically. In the MG approach, it is the steady state queue length probabilities that have a relatively simple expression (i.e., matrix geometric). On the other hand, in the MG approach it is generally difficult to calculate the waiting times and their moments. As an example, Heindl was able to derive the first two moments for a MAP/PH/1 queue in [14] which were already cumbersome.

7. CONCLUSIONS

In this paper, we introduce a stochastic model, namely the MRP-ME, which is a Markov renewal process with a matrix exponential kernel matrix. We then study the steady-state waiting time in a semi-Markov queue with infinite capacity in which the interarrivals and services are both modeled with MRP-MEs. Without having to solve for the steady-state queue lengths by matrix-geometric techniques, we introduce an algorithm that directly finds the waiting time distribution which is in matrix exponential form. The algorithm to obtain the parameters of the matrix exponential form is relatively easy to implement and its numerical engine is the ordered Schur decomposition whose various stable implementations exist in the literature. The numerical examples we present lead us to believe that the proposed algorithm is a promising candidate for a wide range of Markov renewal queueing problems. As future work, we list the possibility of allowing inter-dependence between arrivals and services and the finite capacity case.

Acknowledgment

8. REFERENCES

[1] N. Akar and K. Sohraby. An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Commun. Statist. - Stochastic Models*, 13(3):381–416, 1997.

[2] S. Asmussen and M. Bladt. Renewal theory and queueing algorithms for matrix exponential distributions. In S. R. Chakravarthy and S. Alfa, editors, *Matrix Analytic Methods in Stochastic Models*, pages 313–342. Marcel Dekker, Inc., 1997.

[3] A. Badescu, L. Breuer, A. da Silva Soares, G. Latouche, M.-A. Remiche, and D. Stanford. Risk processes analyzed as fluid queues. *Scand. Actuarial J.*, 2:127–141, 2005.

[4] Z. Bai and J. Demmel. Inverse free parallel spectral divide and conquer algorithms for nonsymmetric eigenproblems. Computer Science Division Report CSD-94-793, University of California at Berkeley, Feb. 1994.

[5] Z. Bai, J. Demmel, and M. Gu. Inverse free parallel spectral divide and conquer algorithms for nonsymmetric eigenproblems. *Numer. Math.*, 76:389–396, 1997.

[6] Z. Bai and J. W. Demmel. On swapping diagonal blocks in real Schur form. *Lin. Alg. Appl.*, 186:73–95, 1993.

[7] D. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM Jour. on Matrix. Analy. Appl.*, 17:906–926, 1996.

[8] J. H. Brandts. Matlab code for sorting real Schur forms. *Numerical Linear Algebra with Applications*, 9(3):249–261, 2002.

[9] E. Cinlar. *Introduction to Stochastic Processes*. Prentice Hall, 1975.

[10] J. H. A. de Smit. Explicit Wiener-Hopf factorizations for the analysis of multidimensional queues. In J. Dshalalow, editor, *Advances in Queueing: Theory, Methods and Open Problems*, pages 293–311. CRC Press, Boca Raton, FL, 1995.

[11] P. M. V. Dooren. *Numerical linear algebra for signals, systems, and control*. Universite Catholique de Louvain, Belgium, 2003. Draft notes prepared for the Graduate School in Systems and Control.

[12] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[13] B. R. Haverkort. *Performance of computer communication systems: A model-based approach*. John Wiley and Sons, 1998.

[14] A. Heindl. *Traffic-Based Decomposition of General Queueing Networks with Correlated Input Processes*. Shaker Verlag, Aachen, Germany, 2001. Ph. D. Thesis.

[15] T. Kailath. *Linear Systems*. Prentice Hall, 1980.

[16] L. Kleinrock. *Queuing Systems, Vol. 1, Theory*. John Wiley, New York, 1975.

[17] J. Kumaran, K. Mitchell, and V. de Liefvoort. An efficient solution to the waiting time distribution in a correlated single server queue. In *Proc. ITC'19*, pages 2327–2336, Beijing, China, 2005.

[18] J. Kumaran, K. Mitchell, and A. van de Liefvoort. A spectral approach to compute performance measures in a correlated single server queue. *Performance*

ρ	k	probability mass at zero		mean waiting time (sec.)	
		LR	Schur	LR	Schur
0.6	4	6.863976e-02	6.863976e-02	1.011650e-01	1.011650e-01
	16	6.473452e-02	6.473452e-02	9.951466e-02	9.951466e-02
	64	6.381361e-02	6.381361e-02	9.910554e-02	9.910554e-02
	256	6.358851e-02	6.358851e-02	9.900358e-02	9.900358e-02
0.9	4	1.099612e-02	1.099612e-02	7.089990e-01	7.089990e-01
	16	1.055977e-02	1.055977e-02	7.004524e-01	7.004524e-01
	64	1.047128e-02	1.047128e-02	6.983209e-01	6.983209e-01
	256	1.045084e-02	1.045084e-02	6.977885e-01	6.977885e-01
0.9999	4	9.839703e-06	9.839678e-06	8.123942e+02	8.123962e+02
	16	9.497196e-06	9.497179e-06	8.030204e+02	8.030219e+02
	64	9.431793e-06	9.431779e-06	8.006772e+02	8.006784e+02
	256	9.417017e-06	9.417052e-06	8.000949e+02	8.000919e+02

Table 2: The probability mass at zero and the mean waiting time for the IPP/E$_k$/1 queue as a function of ρ and the number of stages of the Erlangian service time distribution for the LR algorithm and the proposed Schur algorithm

N	CPU time in sec.		probability mass at zero	mean waiting time (sec.)
	MG	Schur		
8	0.369	0.530	6.541755e-1	5.009292e-4
16	2.96	3.09	3.748969e-1	2.558805e-3
24	10.6	9.92	1.171852e-1	4.383876e-2
26	14.5	12.8	6.139214e-2	1.201707e-1
28	21.4	15.7	1.005555e-2	1.004764

Table 3: The probability mass at zero and the mean waiting time for the MAP/E$_k$/1 queue as a function of N obtained both by the Schur approach and the MG approach and the corresponding CPU times for the two algorithms

Evalaution Review, 33(2):12–14, 2005.

[19] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi-birth-death processes. *J. Appl. Prob.*, 30:650–674, 1993.

[20] G. Latouche and V. Ramaswami. The PH/PH/1 queue at epochs of queue size change. *Queueing Systems*, 25:97–114, 1997.

[21] L. R. Lipsky. *Queueing Theory: A Linear Algebraic Approach*. MacMillan, 1992.

[22] D. M. Lucantoni. New results for the single server queue with a batch Markovian arrival process. *Stoch. Mod.*, 7:1–46, 1991.

[23] 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.

[24] D. M. Lucantoni. New results for the single server queue with a batch Markovian arrival process. *Stochastic Processes and Their Applications*, 82:127–142, 1999.

[25] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts. A single-server queue with server vacations and a class of non-renewal arrival processes. *Adv. Appl. Prob.*, 22:676–705, 1990.

[26] The MATH WORKS Inc. *MATLAB 7.0.0.19901 (R14)*, 2005.

[27] V. Naoumov, U. Krieger, and D. Wagner. Analysis of a multi-server delay-loss system with a general markovian arrival process. In S. Chakravarthy and A. Alfa, editors, *Matrix-analytical methods in Stochastic models*, pages 43–66. Marcel Dekker, 1997.

[28] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981.

[29] M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, Inc., New York, 1989.

[30] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & Systems (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[31] V. Ramaswami. Matrix analytic methods for stochastic fluid flows. In *Proceedings of the 16th International Teletraffic Congress*, pages 1019–1030, Edinburgh, UK, 1999.

[32] A. Riska and E. Smirni. Exact aggregate solutions for M/G/1-type Markov processes. In *SIGMETRICS '02: Proceedings of the 2002 ACM SIGMETRICS international conference on measurement and modeling of computer systems*, pages 86–96, New York, NY, USA, 2002. ACM Press.

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

[34] *LAPACK Users's Guide*, 1995. Second edition.