STOCHASTIC ANALYSES ARISING FROM A NEW APPROACH FOR CLOSED QUEUEING NETWORKS

A Dissertation

by

FENG SUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2006

Major Subject: Industrial Engineering

STOCHASTIC ANALYSES ARISING FROM A NEW APPROACH FOR CLOSED QUEUEING NETWORKS

A Dissertation

by

FENG SUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Richard M. Feldman
Committee Members,	Guy L. Curry
	David R. Larson
	Brett A. Peters
Head of Department,	Brett A. Peters

May 2006

Major Subject: Industrial Engineering

ABSTRACT

Stochastic Analyses Arising from a New Approach for
Closed Queueing Networks. (May 2006)

Feng Sun, B.S., Beijing University of Aeronautics and Astronautics;
M.S., Beijing University of Aeronautics and Astronautics

Chair of Advisory Committee: Dr. Richard M. Feldman

Analyses are addressed for a number of problems in queueing systems and stochastic modeling that arose due to an investigation into techniques that could be used to approximate general closed networks.

In Chapter II, a method is presented to calculate the system size distribution at an arbitrary point in time and at departures for a $\lambda(n)$/G/1/N queue. The analysis is carried out using an embedded Markov chain approach. An algorithm is also developed that combines our analysis with the recursive method of Gupta and Rao. This algorithm compares favorably with that of Gupta and Rao and will solve some situations when Gupta and Rao's method fails or becomes intractable.

In Chapter III, an approach is developed for generating exact solutions of the time-dependent conditional joint probability distributions for a phase-type renewal process. Closed-form expressions are derived when a class of Coxian distributions are used for the inter-renewal distribution. The class of Coxian distributions was chosen so that solutions could be obtained for any mean and variance desired in the inter-renewal times.

In Chapter IV, an algorithm is developed to generate numerical solutions for the steady-state system size probabilities and waiting time distribution functions of the SM/PH/1/N queue by using the matrix-analytic method. Closed form results

are also obtained for particular situations of the preceding queue. In addition, it is demonstrated that the SM/PH/1/N model can be implemented to the analysis of a sequential two-queue system. This is an extension to the work by Neuts and Chakravarthy.

In Chapter V, principal results developed in the preceding chapters are employed for approximate analysis of the closed network of queues with arbitrary service times. Specifically, the $\lambda(n)$/G/1/N queue is applied to closed networks of a general topology, and a sequential two-queue model consisting of the $\lambda(n)$/G/1/N and SM/PH/1/N queues is proposed for tandem queueing networks.

To my Parents and Sister

ACKNOWLEDGMENTS

I express gratitude to my committee chair, Dr. Richard Feldman, for being an excellent advisor and continually providing his help and encouragement over the years. I am also appreciative for his significant contributions and improvements on the manuscripts that we submitted for refereed journals as part of this dissertation. I also thank Dr. Guy Curry and Dr. Brett Peters for their meticulous comments and suggestions. In addition, I am grateful to Dr. Feldman and Dr. Curry for the financial support granted in their capacities as departmental graduate directors. I also express my thanks to Dr. David Larson who taught me the analysis theory and participated on the committee during this dissertation process.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

We are in a world of queueing systems. Phenomena of waiting lines or, by the British version, queues, can be observed frequently in almost everyone's daily life. People are said to spend in average about five years of their lives waiting in lines and six months at traffic lights [31, pp.469]. Stochastic processes arise from the two streams of entities entering the system for service and exiting after completing service. The interaction of the stochastic arrival and departure processes results in the complexity of queueing systems. In terms of networks of queues, we generally refer to open or closed, depending on if arrival streams to and departure streams from the network may take place. Typically, the area of closed queueing networks is more challenging, due to the fact that the total number of customers in the network is constrained to be a constant.

In Chapters II, III and IV of this dissertation, both theoretical developments and computational algorithms are presented for a number of problems involved in queueing theory and stochastic processes. I am convinced by Tijms [46] that, "theory is better understood when the algorithms that solve the problems the theory addresses are presented at the same time". Finally, the solutions proposed in the preceding chapters are implemented in Chapter V for investigation on approximate analytical methods for closed queueing networks.

The journal model is *IEEE Transactions on Automatic Control*.

A. The $\lambda(n)$/G/1/N Queue

The $\lambda(n)$/G/1/N queue has wide application in modeling logistic and manufacturing systems. Depending on the function $\lambda(n)$, the model covers many special cases, such as the M/G/1 system with finite waiting space or with balking probabilities. Courtois and Georges [12] investigated a generalized model of which both the arrival and the service rates are allowed to be arbitrary functions of the current queue size using an embedded Markov chain approach. However, their results are difficult to be applied in computation, as pointed out by Gupta and Rao [18]. By approximating the general service distribution using the Cox-r (C_r) representation, a numerical solution was obtained by Marie [29] and Marie and Pellaumail [30] and provides a particularly efficient algorithm for Cox-2 and Erlang-r service processes. In addition, Marie [28] proposed an approximate solution to general closed queueing networks, where each node is analyzed iteratively as a $\lambda(n)/C_r/1/N$ model. Stewart and Marie [44] and Willits [49] extended the previous model to include multiple servers. See Appendix A for a brief description of Gupta and Rao's method.

Gupta and Rao [17, 18] developed a recursive algorithm to calculate the steady-state probabilities at departures and at an arbitrary point in time for the $\lambda(n)$/G/1/N queue. All the load-dependent arrival rates $\lambda(n)$ must be either identical or all distinct from each other. However, their approach can be modified to accommodate the situation where some of the $\lambda(n)$ values are identical and some are distinct. For example, in [17], Gupta and Rao calculated the system size probabilities of a M/G/1/N machine interference problem with Y spares ($Y < N$). The running times of machines between breakdowns follow an exponential distribution with a constant rate λ, and the repair time distribution is arbitrary. This is essentially a $\lambda(n)$/G/1/$(N+Y)$ model, with $\lambda(n) = N\lambda$ for $0 \leq n \leq Y$ and $\lambda(n) = (N+Y-n)\lambda$ for $Y+1 \leq n \leq N+Y-1$. A difficulty with the procedure in [17, 18] is that whenever a new $\lambda(n)$/G/1/$(N+Y)$ system is analyzed with a different combination of identical and distinct values of $\lambda(n)$, their basic equations will be rewritten and the procedure to solve the equations will be repeated.

In Chapter II, we develop an efficient algorithm for computing the steady-state probabilities of the system size for the $\lambda(n)$/G/1/N queue. We apply an embedded Markov chain approach and provide equations for obtaining numerical results. Combining our analysis with that in [17, 18], we determine a unified algorithm that applies to a broad range of $\lambda(n)$ values, where an operation of partial fraction expansion is easily performed with regard to whether or not the $\lambda(n)$ values are distinct or identical. The proposed algorithm compares favorably with that in [17, 18] in both accuracy and stability.

B. Phase-Type Renewal Processes

A Phase-Type (PH) Renewal Process is a renewal process that has a PH distribution for the inter-renewal times. For a PH renewal process, in Chapter III we propose a procedure to generate the exact solution for the matrices of time-dependent conditional joint probability distributions, denoted by $P(n,t)$, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$. (The definition of $P(n,t)$ will be given later in Chapter III. Throughout this dissertation, $\mathcal{N}$ will denote the nonnegative integers and $\mathcal{R}^+$ will denote the nonnegative reals.) The computation of the matrices $P(n,t)$, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$, is essential in many queueing and stochastic process analyses (e.g., see [35] for the GI/PH/1 queue and [37] for the SM/PH/1 system). One of the reasons for the popularity of PH-renewal processes is the availability of computationally tractable algorithms for obtaining approximate values for the $P(n,t)$ matrices [36]; however, if the algorithmic methods could be replaced by explicit forms for the $P(n,t)$ terms, computational speed and accuracy would become even more enhanced.

In particular, we shall choose a Coxian distribution [13] to match inter-renewal times of arbitrary mean and variance and then derive the associated forms for $P(n,t)$. In addition, we compute the distribution for the counting process $\{N(t); t \geq 0\}$ of PH renewals.

C. The SM/PH/1/N Queue

The SM/PH/1 queueing system (i.e., semi-Markov arrival process, PH service system, single server) is very general, and the associated infinity capacity model has been studied by Neuts and Chakravarthy [37] using the matrix-analytic method. Specifying various arrival and service distributions of the preceding model, Neuts [35] provided solutions for the GI/PH/1, SM/M/1, and PH/PH/1 queues.

The underlying processes of a variety of stochastic models that are rich in structure and of extensive application can be conveniently expressed in terms of matrix solutions. As the principal pioneer in the discipline of matrix-analytic methods, Neuts [36, 39] examined two primary classes of structured queueing systems, which are said to be of the GI/M/1 type and M/G/1 type, respectively. (Also see [38] for an introductory description of matrix-analytic methods in queueing theory.) Matrix-analytic methods have been widely used in the stochastic modeling and analysis of engineering and commercial problems, and theoretical developments fruitfully proposed along with the applications (e.g., see [9, 27]).

In Chapter IV, we consider the finite-buffer SM/PH/1/N queueing system and present numerically tractable solutions for its steady-state system size probabilities and waiting time distribution functions. We also obtain closed form results for some types of semi-Markov arrival processes (those in which the Laplace-Stieltjes transform of the semi-Markovian kernel can be obtained in closed form) with a Cox-2 service process. In addition, we show that the SM/PH/1/N model can be used to study a sequential two-queue system with finite waiting space in both queues. Upon completing service in the first queue, a customer leaves the whole system if observing that the second queue is full. Literature concerned with two queues in series, but assuming that a customer can not leave the first queue and a block occurs when the second queue reaches full capacity, can be found in [25, 33, 34].

D. Closed Queueing Networks

Closed queueing networks are of increasing use in modeling and analysis of modern manufacturing and computer systems, in particular, of such production control processes as Kanban and CONWIP (e.g., see [21]) in which the amount of material being processed remains constant. For the class of closed networks with exponentially distributed service times implementing a first-come first-served (FCFS) discipline, the steady-state joint probabilities have a product-form solution and can be solved exactly [8, 42]. However, the assumption of exponential service time distributions is often too restrictive to appropriately model real-life systems. Therefore, the approximate analysis of non-product-form networks is historically an important research field, to which much work has been devoted.

The literature on approximations for closed queueing networks can be divided into two categories, depending on whether or not aggregation is employed. We first review several techniques not using aggregation and then discuss those that utilize aggregation techniques.

A number of approaches have been proposed in which the topology of the targeted network is not simplified. Kobayashi [22, 23] discussed both stationary and transient solutions for open and closed networks via the diffusion process approximation. Whitt [48] investigated the relationship between open and closed queueing networks and provided the fixed-population-mean (FPM) method, i.e., an approach to use open models with specified expected equilibrium populations, to approximate closed models. It is pointed out in [48] that the FPM approximation for the throughput will be good if either the total population or the number of stations is sufficiently large. Based on the principle of maximum entropy, Kouvatsos [24] and Walstra [47] presented a product-form solution for open and closed queueing networks with FCFS discipline and multiple job classes. By applying the Brownian system theory to closed queueing networks, analytical approaches were proposed by Harrison et al. [19] and Dai and Harrison [16]. However, the computational charge is very demanding and thus, their approaches are limited to networks of very small size. A special situation of the approaches of Brownian models was addressed by Schwerer and Van Mieghem [43] for balanced three-station systems.

The principle of the other category of approaches is to substitute approximately a subnetwork of the original network by a flow equivalent single queue. It is shown by Dallery and Cao [15] and Baynat and Dallery [4] that, under moderate assumptions, a closed network with a complex configuration can be suitably partitioned into a set of subnetworks, each of which is aggregated into a load-dependent exponential single queue. Thus, the original network is approximated by an equivalent product-form network. The methods proposed by Chandy et al. [11] and Marie [28] are the typical aggregation methods for single class networks. As pointed out by Baynat and Dallery [5], these methods are distinguished essentially in the way the load-dependent rates are estimated. Extensions of Marie's method to load-dependent general networks was presented by Akyildiz and Sieber [1] and to multiclass networks by Neuse and Chandy

[32] and Baynat and Dallery [5].

Almost all of the approaches in the literature are developed by means of two-moment approximations for the service times of the original closed network. However, it has been empirically shown that the effect of the third moment of the service time distribution becomes considerable for single queues such as the M/G/1, GI/G/1 and $\lambda(n)$/G/1/N if the squared coefficient of variation (SCV) is greater than 1 [2]. Even though the effect on queueing networks is more complicated, it is worthwhile to find a way to implement the third and higher moments.

In Chapter V, we will demonstrate a unique approach by using exact forms of service time distributions. In addition, we attempt a sequential two-queue model for the approximate analysis of tandem closed networks.

CHAPTER II

STEADY-STATE PROBABILITIES OF $\lambda(n)$/G/1/N QUEUE

We apply an embedded Markov chain approach to the analysis of a $\lambda(n)$/G/1/N queue and develop an algorithm to calculate the system length distribution at an arbitrary point in time and at departures. For the $\lambda(n)$/G/1/N queue to be considered, it is assumed that there exists an integer $N \geq 2$ such that $\lambda(n) > 0$ for all $0 \leq n \leq N-1$, and $\lambda(n) = 0$ for all $n \geq N$. (At times we shall represent $\lambda(n)$ by λ_n for ease of notation.) The service time distribution is arbitrary, where the only requirement is its Laplace-Stieltjes transform can be obtained in closed form and its mean value is finite. The proposed algorithm can be implemented for approximate analysis of closed queueing networks with arbitrary service times and topology, which will be demonstrated in Chapter V.

A. Analysis

Our analysis starts with the load-dependent distribution of interarrival times, denoted by

$$\alpha_n(t) = 1 - e^{-\lambda(n)t}, \qquad \text{for } 0 \leq n \leq N-1, t \in \mathcal{R}^+. \tag{2.1}$$

Let $B_{m,n}(t)$ denote the probability that the total interarrival time of m successive arrivals is less than or equal to t, i.e., there are at least m arrivals in $(0,t]$, given the initial state or customer population n, and there are no services during the time interval. In other words, $B_{m,n}$ is the generalized Erlang distribution, and is given by

$$B_{m,n} = \alpha_n * \alpha_{n+1} * \cdots * \alpha_{n+m-1}, \qquad \text{for } m \geq 1, 0 \leq m+n \leq N, \tag{2.2}$$

where "$*$" represents the convolution operation. Obviously we have $B_{0,n}(t) = 1$ for $0 \leq n \leq N$.

Throughout this dissertation, the Laplace transform of a function will be denoted by a superscript L; in other words,

$$F^L(s) \equiv \int_0^\infty e^{-st} F(t)dt, \qquad \text{for } s \in \mathcal{R}^+, \tag{2.3}$$

where F is a function with support in the nonnegative reals. We will denote by

$$F^*(s) \equiv \int_0^\infty e^{-st}\, dF(t), \qquad \text{for } s \in \mathcal{R}^+, \tag{2.4}$$

the Laplace-Stieltjes transform if in addition F is a non-decreasing function F. We shall also need the factorial derivative and we denote these by $F^{(n)}(s)$ for $n \in \mathcal{N}$; in other words,

$$F^{(n)}(s_0) \equiv F^{(n)}(s)|_{s=s_0} = \frac{1}{n!}\frac{d^n}{ds^n}F(s)|_{s=s_0}, \tag{2.5}$$

where $F^{(0)} = F$.

We have

$$\begin{aligned} B^*_{m,n}(s) &= \int_0^\infty e^{-st}\, dB_{m,n}(t) \\ &= \alpha^*_n(s)\alpha^*_{n+1}(s)\cdots\alpha^*_{n+m-1}(s) \\ &= \frac{\prod_{i=1}^m \lambda_{n+i-1}}{B(s)}, \qquad \text{for } m \geq 1, 1 \leq m+n \leq N, \end{aligned} \tag{2.6}$$

where $B(s) = \prod_{i=1}^m (s + \lambda_{n+i-1})$. With regard to its repeating roots, we rewrite $B(s)$ as

$$B(s) = \prod_{i=1}^{u(m,n)} (s + b_i)^{k_i(m,n)}, \tag{2.7}$$

where $b_1 \neq b_2 \neq \cdots \neq b_{u(m,n)}$ and $\sum_{i=1}^{u(m,n)} k_i(m,n) = m$. Instead of using the Laplace-Stieltjes transform, it is straightforward for us to obtain the expression of $B_{m,n}$ from its Laplace transform

$$B^L_{m,n}(s) = \frac{1}{s} B^*_{m,n}(s). \tag{2.8}$$

By taking the partial fraction expansion then inverse Laplace transformation of $B^L_{m,n}(s)$, and noting that $B^*_{0,n}(s) = 1$ for $0 \leq n \leq N$, we have

$$B_{m,n}(t) = \begin{cases} 1, & \text{for } m = 0, 0 \leq n \leq N; \\ 1 + \sum_{i=1}^{u(m,n)} \sum_{v=1}^{k_i(m,n)} \frac{a_{i,v}(m,n)}{(v-1)!} t^{v-1} e^{-b_i t}, & \text{for } m \geq 1, 1 \leq m+n \leq N; \\ 0, & \text{otherwise,} \end{cases} \tag{2.9}$$

where

$$a_{i,v}(m,n) = \Big[(s+b_i)^{k_i(m,n)} \frac{1}{s} B^*_{m,n}(s)\Big]^{(k_i(m,n)-v)}_{s=-b_i}. \tag{2.10}$$

The operation of the partial fraction expansion is given in algebra textbooks, see for example [45], and is provided by computer softwares for symbolic mathematical calculation, such as $Maple^{®}$.

Then we denote by

$$\psi_{m,n}(t) = B_{m,n}(t) - B_{m+1,n}(t) \tag{2.11}$$

the probability that there are exactly m arrivals in $(0, t]$, given the initial population n, and yield its Laplace transform

$$\begin{aligned} \psi^L_{m,n}(s) &= \frac{1}{s}\Big(B^*_{m,n}(s) - B^*_{m+1,n}(s)\Big) \\ &= \begin{cases} \frac{1}{s+\lambda_n}, & \text{for } m = 0, 0 \leq n \leq N-1; \\ \frac{\prod_{i=1}^{m} \lambda_{n+i-1}}{D(s)}, & \text{for } m \geq 1, 1 \leq m+n \leq N, \end{cases} \end{aligned} \tag{2.12}$$

where $D(s) = \prod_{i=1}^{m+1}(s + \lambda_{n+i-1})$. Re-arranging according to repeating roots yields

$$D(s) = \prod_{i=1}^{u(m+1,n)} (s + d_i)^{k_i(m+1,n)}. \tag{2.13}$$

Therefore we have

$$\psi_{m,n}(t) = \begin{cases} 1, & \text{for } m=0, n=N; \\ e^{-\lambda(n)t}, & \text{for } m=0, 0 \le n \le N-1; \\ \displaystyle\sum_{i=1}^{u(m+1,n)} \sum_{v=1}^{k_i(m+1,n)} \frac{c_{i,v}(m,n)}{(v-1)!} t^{v-1} e^{-d_i t}, & \text{for } m \ge 1, 1 \le m+n \le N; \\ 0, & \text{otherwise,} \end{cases} \tag{2.14}$$

where

$$c_{i,v}(m,n) = \left[(s+d_i)^{k_i(m+1,n)} \psi^L_{m,n}(s) \right]^{(k_i(m+1,n)-v)}_{s=-d_i}. \tag{2.15}$$

For the $\lambda(n)/\mathrm{G}/1/N$ queue, let $T_0 = 0, T_1, T_2, \cdots$ be the instants of successive departures, and denote by X_i the number of customers left behind by the i^{th} departure. Then the stochastic process $(X,T) = \{X_i, T_i; i \in \mathcal{N}\}$ is a Markov renewal process with state space $E_N = \{0, 1, \cdots, N-1\}$. As conventional, the process is assumed time-homogeneous. That is,

$$Q_{kj}(t) = P\{X_{i+1} = j, T_{i+1} - T_i \le t | X_i = k\} \qquad \text{for } j,k \in E_N, t \in \mathcal{R}^+, \tag{2.16}$$

which is independent of i. The service time distribution $\varphi(\cdot)$ is arbitrary, with its Laplace-Stieltjes transform given by $\varphi^*(s)$. Denote $Q(t)$ to be the matrix whose (k,j)th entry is $Q_{kj}(t)$; that yields

$$Q(t) = \begin{pmatrix} p_0(t) & p_1(t) & p_2(t) & \cdots & p_{N-1}(t) \\ q_{0,1}(t) & q_{1,1}(t) & q_{2,1}(t) & \cdots & q_{N-1,1}(t) \\ & q_{0,2}(t) & q_{1,2}(t) & \cdots & q_{N-2,2}(t) \\ & & \ddots & \vdots & \vdots \\ \mathbf{0} & & & q_{0,N-1}(t) & q_{1,N-1}(t) \end{pmatrix}, \tag{2.17}$$

where

$$q_{m,n}(t) = \int_0^t \varphi(dx)\psi_{m,n}(x), \qquad \text{for } 1 \le n \le N-1, 1 \le m+n \le N, \tag{2.18}$$

$$p_m(t) = \int_0^t \lambda_0 e^{-\lambda(0)(t-x)} q_{m,1}(x)\,dx, \qquad \text{for } 0 \le m \le N-1. \tag{2.19}$$

Here we also have

$$q_{N-n,n}(t) = \varphi(t) - \sum_{i=0}^{N-n-1} q_{i,1}(t), \qquad \text{for } 1 \le n \le N-1, \tag{2.20}$$

$$p_{N-1}(t) = \int_0^t \lambda_0 e^{-\lambda(0)x} \varphi(t-x)\,dx - \sum_{i=0}^{N-2} p_i(t). \tag{2.21}$$

The single-step probability transition matrix of the embedded Markov chain $\{X_i; i \in \mathcal{N}\}$ is

$$\tilde{Q} = \begin{pmatrix} q_{0,1} & q_{1,1} & q_{2,1} & \cdots & q_{N-1,1} \\ q_{0,1} & q_{1,1} & q_{2,1} & \cdots & q_{N-1,1} \\ & q_{0,2} & q_{1,2} & \cdots & q_{N-2,2} \\ & & \ddots & \vdots & \vdots \\ \mathbf{0} & & & q_{0,N-1} & q_{1,N-1} \end{pmatrix}, \tag{2.22}$$

where

$$q_{m,n} = \int_0^\infty \varphi(dt)\psi_{m,n}(t), \qquad \text{for } 1 \le n \le N-1, 1 \le m+n \le N. \tag{2.23}$$

By (2.23) and substituting $\psi_{m,n}(t)$ by the right-hand side of (2.14), then using the final value property of the Laplace-Stieltjes transformation yields

$$q_{m,n} = \begin{cases} \sum_{i=1}^{u(m+1,n)} \sum_{v=1}^{k_i(m+1,n)} (-1)^{v-1} c_{i,v}(m,n) \varphi^{*(v-1)}(d_i), & \text{for } m \ge 1, 1 \le n \le N-m; \\ \varphi^*(\lambda_n), & \text{for } m = 0, 1 \le n \le N-1; \\ 0, & \text{otherwise,} \end{cases} \tag{2.24}$$

Let Y_t denote the number of customers in the system at time t. Then $Y = \{Y_t; t \ge 0\}$ is a semi-regenerative process, with state space $E_{N+1} = \{0, 1, \cdots, N\}$ and embedded Markov renewal process (X,T). Let $K(t,k,j) = P\{Y_t = j, T_1 > t | X_0 = k\}$, for $0 \le j \le N$, $0 \le k \le N$, and $t \ge 0$. We have

$$K(t,k,j) = \begin{cases} e^{-\lambda(0)t}, & \text{for } k = j = 0; \\ \int_0^t \lambda_0 e^{-\lambda(0)(t-x)}[1-\varphi(x)]B_{N-1,0}(x)\,dx, & \text{for } k = 0, j = N; \\ \int_0^t \lambda_0 e^{-\lambda(0)(t-x)}[1-\varphi(x)]\psi_{j-1,0}(x)\,dx, & \text{for } k = 0, N-1 \ge j \ge 1; \\ [1-\varphi(t)]B_{N-k,k}(t), & \text{for } N = j \ge k \ge 1; \\ [1-\varphi(t)]\psi_{j-k,k}(t), & \text{for } N-1 \ge j \ge k \ge 1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.25}$$

This is because, for $k = j = 0$, $e^{-\lambda(0)t}$ is the probability that there is no arrival in $(0,t]$; for $k = 0$ and $j = N$, after the first arrival at some time $t - x$ before t, during the time interval $(t-x,t]$ there are $N-1$ or more arrivals with probability $B_{N-1,0}(x)$, and satisfying $T_1 > x$ with probability $1-\varphi(x)$; for $k = 0$ and $N-1 \ge j \ge 1$, after the first arrival at $t-x$, the probability that there are $j-1$ arrivals during the interval $(t-x,t]$ is $\psi_{j-1,0}(x)$; if there are already $k \ge 1$ in the system, then T_1 is the same as the first service time. In addition, during the time interval $(0,t]$, the probability that there are $j-k$ arrivals is $B_{N-k,k}(t)$ for $N = j \ge k$; and is $\psi_{j-k,k}(t)$ for $N-1 \ge j \ge k$.

From (2.25) and introducing the notation

$$g(s) = \frac{1-\varphi^*(s)}{s}, \tag{2.26}$$

we then have,

$$\int_0^\infty K(t,k,j)dt =$$

$$
\begin{cases}
\lim_{s\to 0} \dfrac{1}{s+\lambda_0}, & \text{for } k=j=0;\\
\lim_{s\to 0} \dfrac{\lambda_0}{s+\lambda_0}\Big[g(s)+\sum_{i=1}^{u(N-1,0)}\sum_{v=1}^{k_i(N-1,0)}(-1)^{v-1}a_{i,v}(N-1,0)g^{(v-1)}(s+b_i)\Big], & \\
 & \text{for } k=0, j=N;\\
\lim_{s\to 0} \dfrac{\lambda_0}{s+\lambda_0}g(s+\lambda_0), & \text{for } k=0, j=1;\\
\lim_{s\to 0} \dfrac{\lambda_0}{s+\lambda_0}\sum_{i=1}^{u(j,0)}\sum_{v=1}^{k_i(j,0)}(-1)^{v-1}c_{i,v}(j-1,0)g^{(v-1)}(s+d_i), & \text{for } k=0, N-1\geq j\geq 2;\\
\lim_{s\to 0} g(s), & \text{for } j=k=N;\\
\lim_{s\to 0} \Big[g(s)+\sum_{i=1}^{u(N-k,k)}\sum_{v=1}^{k_i(N-k,k)}(-1)^{v-1}a_{i,v}(N-k,k)g^{(v-1)}(s+b_i)\Big], & \\
 & \text{for } N=j>k\geq 1;\\
\lim_{s\to 0} g(s+\lambda_k),, & \text{for } N-1\geq j=k\geq 1;\\
\lim_{s\to 0} \sum_{i=1}^{u(j-k+1,k)}\sum_{v=1}^{k_i(j-k+1,k)}(-1)^{v-1}c_{i,v}(j-k,k)g^{(v-1)}(s+d_i), & \text{for } N-1\geq j>k\geq 1;\\
0, & \text{otherwise}
\end{cases}
$$

$$
= \tag{2.27}
$$

$$
\begin{cases}
\dfrac{1}{\lambda_0}, & \text{for } k=j=0;\\
g(0)+\displaystyle\sum_{i=1}^{u(N-1,0)}\sum_{v=1}^{k_i(N-1,0)}(-1)^{v-1}a_{i,v}(N-1,0)g^{(v-1)}(b_i), & \text{for } k=0, j=N;\\
g(\lambda_0), & \text{for } k=0, j=1;\\
\displaystyle\sum_{i=1}^{u(j,0)}\sum_{v=1}^{k_i(j,0)}(-1)^{v-1}c_{i,v}(j-1,0)g^{(v-1)}(d_i), & \text{for } k=0, N-1\ge j\ge 2;\\
g(0), & \text{for } k=j=N;\\
g(0)+\displaystyle\sum_{i=1}^{u(N-k,k)}\sum_{v=1}^{k_i(N-k,k)}(-1)^{v-1}a_{i,v}(N-k,k)g^{(v-1)}(b_i), & \text{for } N=j>k\ge 1;\\
g(\lambda_k), & \text{for } N-1\ge j=k\ge 1;\\
\displaystyle\sum_{i=1}^{u(j-k+1,k)}\sum_{v=1}^{k_i(j-k+1,k)}(-1)^{v-1}c_{i,v}(j-k,k)g^{(v-1)}(d_i), & \text{for } N-1\ge j>k\ge 1;\\
0, & \text{otherwise,}
\end{cases}
$$

where $g(0)$ can be obtained by observing that

$$
g(0) \;\equiv\; \lim_{s\to 0} g(s) = -\lim_{s\to 0}\frac{d}{ds}\,\varphi^*(s) \;\equiv\; -\varphi^{*(1)}(0)\,. \tag{2.28}
$$

The benefit of this last equality is that if one of the logic software systems that handle symbolic computation is used, then it is possible that $g(0)$ cannot be evaluated directly since it involves a division by zero; however, $\varphi^{*(1)}(0)$ may be possible if φ^* is not too complex.

Now we are ready to determine $y(j)=\lim_{t\to\infty}\mathrm{P}\{Y_t=j|X_0\}$; that is, the steady-state probability of j customers in the system at an arbitrary point of time. By [14, Theorems 10.4.3 and 10.6.12], we have

$$
y(j)=\sum_{k=0}^{j}\eta(k)\int_0^\infty K(t,k,j)\,dt, \tag{2.29}
$$

where

$$\eta(k) = \frac{\nu(k)}{\sum_{i=0}^{N-1} \nu(i)\tau(i)}, \tag{2.30}$$

which is the inverse of the mean recurrence time of state k in (X, T). The vector $\nu = (\nu(0), \nu(1), \cdots, \nu(N-1))$ satisfies

$$\nu\tilde{Q} = \nu, \tag{2.31}$$

and $\nu(k) = 0$, for $k \geq N$. The expected sojourn time for each visit to state i is given by

$$\tau(i) = \begin{cases} b + \dfrac{1}{\lambda}, & \text{for } i = 0; \\ b, & \text{for } i \geq 1, \end{cases} \tag{2.32}$$

where $b = \int_0^\infty [1 - \varphi(t)]\, dt$ is the mean service time. Therefore we have

$$y(j) = \sum_{k=0}^{j} \frac{\nu(k)}{\dfrac{\nu(0)}{\lambda_0} + b\displaystyle\sum_{i=0}^{N-1} \nu(i)} \int_0^\infty K(t, k, j)\, dt, \qquad \text{for } 0 \leq j \leq N. \tag{2.33}$$

Denote by $\pi(j)$ the steady-state probability of j customers remaining in the system immediately after a service completion, for $N-1 \geq j \geq 0$. The normalizing condition yields

$$\pi(j) = \frac{\nu(j)}{\displaystyle\sum_{i=0}^{N-1} \nu(i)}, \qquad \text{for } N - 1 \geq j \geq 0. \tag{2.34}$$

Before ending this section, we give explicit equations for the computation of $q_{m,n}$ and $\int_0^\infty K(t, k, j)dt$ for two particular cases, according to whether all the $\lambda(n)$ values are identical or distinct from each other.

Case 1. Assume $\lambda_0 = \lambda_1 = \cdots = \lambda_{N-1} = \lambda$. Then for each given $n = 0, 1, \cdots, N-1$, and m such that $m \geq 1$ and $1 \leq m+n \leq N$, $B_{m,n}$ is an Erlang (m, λ)

distribution and $\psi_{m,n}$ follows Poisson distribution. Now we have

$$q_{m,n} = \begin{cases} (-\lambda)^m \varphi^{*(m)}(\lambda), & \text{for } m \geq 1, 1 \leq m+n \leq N-1; \\ \varphi^*(0) - \sum_{i=0}^{m-1} (-\lambda)^i \varphi^{*(i)}(\lambda), & \text{for } m \geq 1, m+n = N; \\ \varphi^*(\lambda), & \text{for } m = 0, N-1 \geq n \geq 0; \\ 0, & \text{otherwise,} \end{cases} \tag{2.35}$$

$$\int_0^\infty K(t,k,j)dt = \begin{cases} \frac{1}{\lambda}, & \text{for } k = j = 0; \\ g(0) - \sum_{i=0}^{N-2} (-\lambda)^i g^{(i)}(\lambda), & \text{for } k = 0, j = N; \\ g(\lambda), & \text{for } k = 0, j = 1; \\ (-\lambda)^{j-1} g^{(j-1)}(\lambda), & \text{for } k = 0, N-1 \geq j \geq 2; \\ g(0), & \text{for } k = j = N; \\ g(0) - \sum_{i=0}^{N-k-1} (-\lambda)^i g^{(i)}(\lambda), & \text{for } N = j > k \geq 1; \\ g(\lambda), & \text{for } N-1 \geq j = k \geq 1; \\ (-\lambda)^{j-k} g^{(j-k)}(\lambda), & \text{for } N-1 \geq j > k \geq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.36}$$

Case 2. On the other hand, if $\lambda_0 \neq \lambda_1 \neq \cdots \neq \lambda_{N-1}$, then for each pair of m and n we have

$$q_{m,n} = \begin{cases} \sum_{v=1}^{m+1} c_v(m,n) \varphi^*(\lambda_{n+v-1}), & \text{for } m \geq 1, 1 \leq n \leq N-m; \\ \varphi^*(\lambda_n), & \text{for } m = 0, 1 \leq n \leq N-1; \\ 0, & \text{otherwise,} \end{cases} \tag{2.37}$$

$$\int_0^\infty K(t,k,j)\, dt = \tag{2.38}$$

$$
\begin{cases}
\frac{1}{\lambda_0}, & \text{for } k = j = 0; \\
g(0) + \sum_{v=1}^{N-1} a_v(N-1, 0) g(\lambda_{v-1}), & \text{for } k = 0, j = N; \\
g(\lambda_0), & \text{for } k = 0; j = 1; \\
\sum_{v=1}^{j} c_v(j-1, 0) g(\lambda_{v-1}), & \text{for } k = 0, N-1 \geq j \geq 2; \\
g(0), & \text{for } j = k = N; \\
g(0) + \sum_{v=1}^{N-k} a_v(N-k, k) g(\lambda_{k+v-1}), & \text{for } N = j > k \geq 1; \\
g(\lambda_k), & \text{for } N-1 \geq j = k \geq 1; \\
\sum_{v=1}^{j-k+1} c_v(j-k, k) g(\lambda_{k+v-1}), & \text{for } N-1 \geq j > k \geq 1; \\
0, & \text{otherwise,}
\end{cases}
$$

where

$$
a_v(m, n) = - \prod_{k=1,2,\cdots,m}^{k \neq v} \frac{\lambda_{n+k-1}}{\lambda_{n+k-1} - \lambda_{n+v-1}}, \tag{2.39}
$$

$$
c_v(m, n) = \frac{\prod_{k=1}^{m} \lambda_{n+k-1}}{\prod_{k=1,2,\cdots,m+1}^{k \neq v} (\lambda_{n+k-1} - \lambda_{n+v-1})}. \tag{2.40}
$$

The values of $a_v(m, n)$ and $c_v(m, n)$ satisfy

$$
\sum_{v=1}^{m} a_v(m, n) = -1 \tag{2.41}
$$

$$
\sum_{v=1}^{m+1} c_v(m, n) = 0. \tag{2.42}
$$

B. Algorithm

We observe that, for Case 1, our computation of $\pi(j)$ and $y(j)$ is exact. For Case 2 (i.e., the λ_{n+v-1} values are distinct for $1 \leq v \leq m$), as m goes large or some of the λ_{n+v-1} values are close to each other, $a_v(m, n)$ and $c_v(m, n)$ may become extremely large, and their computed values no longer satisfy (2.41) and (2.42) due to round-off errors. This brings incremental errors in succeeding computations. By substituting

$$c_u(m,n) = -\sum_{v=1,2,\cdots,m+1}^{v\neq u} c_v(m,n) \tag{2.43}$$

for (2.37), we obtain

$$q_{m,n} = \tag{2.44}$$
$$\begin{cases} \sum\limits_{v=1,2,\cdots,m+1}^{v\neq u} c_v(m,n)[\varphi^*(\lambda_{n+v-1}) - \varphi^*(\lambda_{n+u-1})], & \text{for } m \geq 1, 1 \leq n \leq N-m; \\ \varphi^*(\lambda_n), & \text{for } m = 0, 1 \leq n \leq N-1; \\ 0, & \text{otherwise.} \end{cases}$$

To assure smaller computational errors, given m and n, the index u may be determined by picking from $c_v(m,n)$ computed by (2.40) for all $v = 1, 2, \cdots, m+1$ the one with greatest absolute value.

We observe also in computational experiments that, as m goes larger, the transition probabilities $q_{m,n}$ may appear to be negative or clearly unreasonable after their row cumulative value in the matrix $\tilde{Q}$ approaches one. This causes inaccuracy of $\pi(j)$, obtained by (2.31) and (2.34). The inaccuracy of the computed $\pi(j)$ and $\int_0^\infty K(t,k,j)\,dt$ further deteriorates that of $y(j)$ computed by (2.33).

Expand (2.31) as a system of equations, then for each $j = 0, 1, \cdots, N-1$, add the equations for $\nu(0), \cdots, \nu(j)$ side by side and solve for $\nu(j+1)$. This gives

$$\nu(k) = \begin{cases} \dfrac{1}{q_{0,1}}(1 - q_{0,1})\nu(0), & \text{for } k = 1; \\ \dfrac{1}{q_{0,k}}\Big[(1 - \sum\limits_{i=0}^{k-1} q_{i,1})\nu(0) + \sum\limits_{j=1}^{k-1}(1 - \sum\limits_{i=0}^{k-j} q_{i,j})\nu(j)\Big], & \text{for } N-1 \geq k \geq 2. \end{cases} \tag{2.45}$$

For a given $j \leq N-2$, if we meet some $u \leq N-j-1$ such that, $\sum_{i=0}^{u-1} q_{i,j} \leq 1$, and $\sum_{i=0}^{u} q_{i,j} > 1$, this indicates the computed value of $q_{u,j}$ is not correct. Except for the situation that a number of $\lambda(n)$ have very close values, usually the summation of the remaining probabilities, $\sum_{i=u}^{N-j-1} q_{i,j}$, accounts for only a small portion of the total cumulative which equals one. Assuming the computed values $q_{i,j}$, for $i = 0, 1, \cdots, u-1$, are trustable, we approximate $\sum_{i=0}^{k} q_{i,j}$, for $k = u, u+1, \cdots, N-j-1$, by curve fitting. In practice, we let $(x_0, f_0) = (0, q_{0,j})$, $(x_1, f_1) = (1, \sum_{i=0}^{1} q_{i,j})$, $\cdots$, $(x_{u-1}, f_{u-1}) = (u-1, \sum_{i=0}^{u-1} q_{i,j})$, and the equation to fit is

$$f(x) = 1 + (f_{u-1} - 1)(\frac{x_{N-j} - x}{x_{N-j} - x_{u-1}})^a, \tag{2.46}$$

where a is the parameter to be determined by using the least square method. For $a > 0$, the equation is strictly increasing and satisfies $f(x_{u-1}) = f_{u-1}$ and $f(x_{N-j}) = 1$.

Though (2.33) provides a numerically tractable solution for the system size probability at an arbitrary point in time, we choose an alternative approach that not only requires less computational work but also improves outcome accuracy. This is accomplished by combining our results with that of Gupta and Rao [18]. By (19) and (20) in [18], respectively,

$$p_0 = \frac{\nu(1)\varphi^*(\lambda_1)}{\lambda_0[1 - \varphi^*(\lambda_1)]}, \tag{2.47}$$

$$p_1 = \frac{\nu(1)}{\lambda_1}, \tag{2.48}$$

and recursively by (24) in [18],

$$p_j = \frac{1}{\lambda_j}[\lambda_{j-1}\, p_{j-1} + \nu(j) - \nu(j-1)], \qquad \text{for } N-1 \geq j \geq 2. \tag{2.49}$$

The last unknown quantity p_N is given by (25) in [18],

$$p_N = -\lambda_{N-1}\, \hat{p}_{N-1}, \tag{2.50}$$

where $\hat{p}_{N-1}$ is determined by (26) and (27) in [18]. That is,

$$\hat{p}_1 = \frac{1}{\lambda_1}[-\lambda_0\, p_0\, b - \nu(1)\, b + p_1], \tag{2.51}$$

$$\hat{p}_j = \frac{1}{\lambda_j}[\lambda_{j-1}\, \hat{p}_{j-1} - \nu(j)\, b + p_j], \qquad \text{for } N-1 \geq j \geq 2, \tag{2.52}$$

where b is the mean service time, specified in Section II.A. Finally we have

$$y(j) = \frac{p_j}{\sum_{i=0}^{N} p_i}, \qquad \text{for } N \geq j \geq 0. \tag{2.53}$$

We summarize the proposed algorithm in the following:

Step 1. For each $n = 0, 1, \cdots, N-1$, and each m such that $m \geq 0$ and $1 \leq m + n \leq N$, compute $q_{m,n}$ using (2.24). In particular, if all the λ_{n+i-1} values for $1 \leq i \leq m$ are identical, then compute $q_{m,n}$ using (2.35); if all the λ_{n+i-1} for $1 \leq i \leq m$ are distinct from each other, then compute $q_{m,n}$ using (2.44).

Step 2. For each $j = 1, 2, \cdots, N-2$, and for k from 1 up to $N-j-1$, compute $\sum_{i=0}^{k} q_{i,j}$. If we meet some $k = u$ such that, $\sum_{i=0}^{u-1} q_{i,j} \leq 1$, and $\sum_{i=0}^{u} q_{i,j} > 1$, then approximate $\sum_{i=0}^{u} q_{i,j} = f(u), \sum_{i=0}^{u+1} q_{i,j} = f(u+1), \cdots, \sum_{i=0}^{N-j-1} q_{i,j} = f(N-j-1)$, where $f(\cdot)$ is given by (2.46). Note that $\sum_{i=0}^{N-j} q_{i,j} = 1$, for all $j = 1, 2, \cdots, N-1$.

Step 3. Set $\nu(0) = 1$, then compute $\nu(j)$, for $j = 1, 2, \cdots, N-1$ using (2.45).

Step 4. For all $j = 0, 1, \cdots, N-1$, compute $\pi(j)$ using (2.34).

Step 5. For all $j = 0, 1, \cdots, N$, compute $y(j)$ using (2.47) through (2.53).

C. Numerical Examples

We compare results generated by [18], the proposed algorithm, and simulation, for various $\lambda(n)$ functions, and service time distributions with various effective traffic intensities and squared coefficients of variation (SCV), and different forms such as exponential (M), deterministic (D), Cox-2 (C_2) and Gamma. For almost all the simulation outcomes given in this section, we have a 95% confidence interval with the half-width of the interval being less than 2% of the mean estimate, or 0.00001 for very small probabilities. In both the proposed algorithm and that in [18], the system length probabilities $\pi(j)$ and $y(j)$ are normalized in the calculations and, thus, the corresponding cumulative probabilities always equal one.

If all the arrival rates $\lambda(n) = \lambda$, for $N - 1 \geq n \geq 0$, this is a M/G/1/N model. Both the proposed method and that of [18] generate exact results, as the example shown in Table I. The columns 2 and 5 of Table I were given in [18] and we recalculated them. The results in column 4 were obtained by applying the M/M/1/N model given in queueing theory textbooks.

Table I. Steady-state probabilities at arbitrary and at departures for $\lambda(n) = \lambda$, $N = 11$

	M $\lambda = 1, \mu = 5$			D $\lambda = 1, \mu = 2$			C_2 $\lambda = 2, \mu \approx 1.24, SCV \approx 3.42$ $(\mu_1 = 7, \mu_2 = 0.5, p = 1/3)$		
	Gupta's	New	Exact	Gupta's	New	Simu.	Gupta's	New	Simu.
$\pi(0)$	0.80000	0.80000	0.80000	0.50000	0.50000	0.49996	0.05054	0.05054	0.05052
$\pi(1)$	0.16000	0.16000	0.16000	0.32436	0.32436	0.32438	0.03807	0.03807	0.03799
$\pi(2)$	0.03200	0.03200	0.03200	0.12260	0.12260	0.12263	0.04061	0.04061	0.04049
$\pi(3)$	0.00640	0.00640	0.00640	0.03779	0.03779	0.03780	0.04826	0.04826	0.04820
$\pi(4)$	0.00128	0.00128	0.00128	0.01091	0.01091	0.01091	0.05879	0.05879	0.05883
$\pi(5)$	0.00026	0.00026	0.00026	0.00311	0.00311	0.00310	0.07200	0.07200	0.07198
$\pi(6)$	0.00005	0.00005	0.00005	0.00088	0.00088	0.00088	0.08827	0.08827	0.08828
$\pi(7)$	0.00001	0.00001	0.00001	0.00025	0.00025	0.00025	0.10825	0.10825	0.10839
$\pi(8)$	0.00000	0.00000	0.00000	0.00007	0.00007	0.00007	0.13275	0.13275	0.13284
$\pi(9)$	0.00000	0.00000	0.00000	0.00002	0.00002	0.00002	0.16280	0.16280	0.16268
$\pi(10)$	0.00000	0.00000	0.00000	0.00001	0.00001	0.00001	0.19965	0.19965	0.19979
y(0)	0.80000	0.80000	0.80000	0.50000	0.50000	0.49997	0.03027	0.03027	0.03021
y(1)	0.16000	0.16000	0.16000	0.32436	0.32436	0.32435	0.02280	0.02280	0.02274
y(2)	0.03200	0.03200	0.03200	0.12260	0.12260	0.12263	0.02432	0.02432	0.02430
y(3)	0.00640	0.00640	0.00640	0.03779	0.03779	0.03780	0.02891	0.02891	0.02886
y(4)	0.00128	0.00128	0.00128	0.01091	0.01091	0.01092	0.03521	0.03521	0.03519
y(5)	0.00026	0.00026	0.00026	0.00311	0.00311	0.00310	0.04312	0.04312	0.04310
y(6)	0.00005	0.00005	0.00005	0.00088	0.00088	0.00088	0.05287	0.05287	0.05284
y(7)	0.00001	0.00001	0.00001	0.00025	0.00025	0.00025	0.06484	0.06484	0.06488
y(8)	0.00000	0.00000	0.00000	0.00007	0.00007	0.00007	0.07951	0.07951	0.07962
y(9)	0.00000	0.00000	0.00000	0.00002	0.00002	0.00002	0.09751	0.09751	0.09752
y(10)	0.00000	0.00000	0.00000	0.00001	0.00001	0.00001	0.11958	0.11958	0.11961
y(11)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.40105	0.40105	0.40113

The tables that follow demonstrate that the proposed algorithm is stable for distinct $\lambda(n)$ as well. This includes situations for which the method of [17, 18] produces inaccurate results.

When all $\lambda(n)$ are different from each other, we give comparisons for large N in Table II and for close $\lambda(n)$ values in Table III, respectively. As shown in Table II, the proposed algorithm effectively eliminates both the number and the value of negative probabilities in those situations where the small probabilities computed by [18] become negative. Table III further tests the proposed algorithm for three specific situations, in which the probabilities computed by [18] are all positive but with a number of inaccurate values, shown in column 2; mostly positive but rather inaccurate, with a few negative quantities which may have either small or large absolute

Table II. Steady-state probabilities at arbitrary and at departures for $\lambda(n) = (1 - n/N)\lambda$, $N = 20$

	M $\lambda = 2, \mu = 5$			D $\lambda = 2, \mu = 2$			$Gamma$ $\lambda = 15, \mu \approx 1.17, SCV = 3.5$ $(\alpha = 2/7, \beta = 3)$		
	Gupta's	New	Simu.	Gupta's	New	Simu.	Gupta's	New	Simu.
$\pi(0)$	0.63118	0.63109	0.63118	0.14581	0.14581	0.14589	0.00000	0.00000	0.00000
$\pi(1)$	0.23985	0.23981	0.23985	0.23122	0.23121	0.23135	0.00000	0.00000	0.00000
$\pi(2)$	0.08635	0.08633	0.08635	0.21933	0.21933	0.21948	0.00001	0.00001	0.00000
$\pi(3)$	0.02936	0.02935	0.02935	0.16726	0.16726	0.16734	0.00002	0.00002	0.00000
$\pi(4)$	0.00939	0.00939	0.00939	0.11137	0.11137	0.11128	0.00005	0.00005	0.00000
$\pi(5)$	0.00282	0.00282	0.00282	0.06573	0.06573	0.06556	0.00012	0.00012	0.00001
$\pi(6)$	0.00077	0.00077	0.00079	0.03429	0.03429	0.03416	0.00029	0.00029	0.00004
$\pi(7)$	0.00032	0.00032	0.00021	0.01575	0.01572	0.01569	0.00064	0.00064	0.00012
$\pi(8)$	-0.00041	0.00008	0.00005	0.00606	0.00628	0.00627	0.00141	0.00141	0.00035
$\pi(9)$	0.00142	0.00002	0.00001	0.00291	0.00217	0.00216	0.00299	0.00299	0.00096
$\pi(10)$	-0.00344	0.00000	0.00000	-0.00113	0.00064	0.00064	0.00612	0.00612	0.00248
$\pi(11)$	0.00675	0.00000	0.00000	0.00317	0.00016	0.00016	0.01206	0.01206	0.00614
$\pi(12)$	-0.01067	0.00000	0.00000	-0.00374	0.00003	0.00004	0.02279	0.02279	0.01419
$\pi(13)$	0.01342	0.00000	0.00000	0.00353	0.00001	0.00001	0.04097	0.04097	0.03045
$\pi(14)$	-0.01314	0.00000	0.00000	-0.00247	0.00000	0.00000	0.06947	0.06947	0.05993
$\pi(15)$	0.00972	0.00000	0.00000	0.00134	0.00000	0.00000	0.10971	0.10971	0.10642
$\pi(16)$	-0.00523	0.00000	0.00000	-0.00057	0.00000	0.00000	0.15837	0.15837	0.16642
$\pi(17)$	0.00192	0.00000	0.00000	0.00019	0.00000	0.00000	0.20234	0.20234	0.22108
$\pi(18)$	-0.00043	0.00000	0.00000	-0.00005	0.00000	0.00000	0.21469	0.21469	0.23199
$\pi(19)$	0.00005	0.00000	0.00000	0.00001	0.00000	0.00000	0.15795	0.15795	0.15941
y(0)	0.61209	0.61206	0.61341	0.12726	0.12726	0.12728	0.00000	0.00000	0.00000
y(1)	0.24484	0.24482	0.24097	0.21241	0.21241	0.21249	0.00000	0.00000	0.00000
y(2)	0.09304	0.09303	0.09156	0.21269	0.21269	0.21283	0.00000	0.00000	0.00000
y(3)	0.03349	0.03349	0.03296	0.17174	0.17173	0.17182	0.00000	0.00000	0.00000
y(4)	0.01139	0.01139	0.01121	0.12150	0.12150	0.12153	0.00001	0.00001	0.00000
y(5)	0.00365	0.00365	0.00359	0.07648	0.07649	0.07635	0.00001	0.00001	0.00000
y(6)	0.00106	0.00106	0.00108	0.04275	0.04275	0.04261	0.00003	0.00003	0.00000
y(7)	0.00048	0.00048	0.00030	0.02114	0.02111	0.02103	0.00008	0.00008	0.00001
y(8)	-0.00067	0.00013	0.00008	0.00881	0.00913	0.00912	0.00018	0.00018	0.00005
y(9)	0.00251	0.00004	0.00002	0.00462	0.00344	0.00344	0.00042	0.00042	0.00014
y(10)	-0.00667	0.00001	0.00000	-0.00197	0.00112	0.00111	0.00095	0.00095	0.00039
y(11)	0.01455	0.00000	0.00000	0.00615	0.00031	0.00031	0.00209	0.00209	0.00106
y(12)	-0.02588	0.00000	0.00000	-0.00816	0.00007	0.00007	0.00443	0.00443	0.00276
y(13)	0.03719	0.00000	0.00000	0.00879	0.00001	0.00001	0.00910	0.00910	0.00682
y(14)	-0.04247	0.00000	0.00000	-0.00720	0.00000	0.00000	0.01801	0.01801	0.01559
y(15)	0.03771	0.00000	0.00000	0.00467	0.00000	0.00000	0.03413	0.03413	0.03326
y(16)	-0.02534	0.00000	0.00000	-0.00250	0.00000	0.00000	0.06159	0.06159	0.06493
y(17)	0.01242	-0.00000	0.00000	0.00113	0.00000	0.00000	0.10492	0.10492	0.11512
y(18)	-0.00420	-0.00000	0.00000	-0.00040	0.00000	0.00000	0.16698	0.16698	0.18123
y(19)	0.00088	-0.00000	0.00000	0.00010	0.00000	0.00000	0.24570	0.24570	0.24907
y(20)	-0.00009	-0.00017	0.00482	-0.00001	-0.00001	0.00000	0.35137	0.35137	0.32957

Table III. Steady-state probabilities at arbitrary and at departures for $\lambda(n) = 3/2 + 1/2^{1+n/8}, N = 10$
That is, $\lambda(0) = 2.0$, $\lambda(1) \approx 1.958502022$, $\lambda(2) \approx 1.920448208$, $\lambda(3) \approx 1.885552706$, $\lambda(4) \approx 1.853553391$, $\lambda(5) \approx 1.824209889$, $\lambda(6) \approx 1.797301779$, $\lambda(7) \approx 1.772626933$, $\lambda(8) = 1.75$, $\lambda(9) \approx 1.729251011$.

	M $\mu = 1$			*D* $\mu \approx 3.85$ ($mean = 0.26$)			*Gamma* $\mu \approx 1.17, SCV = 3.5$ ($\alpha = 2/7, \beta = 3$)		
	Gupta's	New	Simu.	Gupta's	New	Simu.	Gupta's	New	Simu.
$\pi(0)$	0.00142	0.00187	0.00187	0.04366	0.49554	0.49546	-0.02361	0.06628	0.04836
$\pi(1)$	0.00278	0.00366	0.00366	0.02899	0.32903	0.32904	-0.01735	0.04870	0.04355
$\pi(2)$	0.00533	0.00703	0.00703	0.01095	0.12430	0.12436	-0.01989	0.05583	0.05107
$\pi(3)$	0.01005	0.01325	0.01326	0.00329	0.03731	0.03731	-0.02377	0.06674	0.06206
$\pi(4)$	0.01863	0.02456	0.02457	0.00090	0.01020	0.01022	-0.02861	0.08032	0.07643
$\pi(5)$	0.03398	0.04479	0.04482	0.00023	0.00272	0.00269	-0.03439	0.09654	0.09380
$\pi(6)$	0.06110	0.08053	0.08053	0.00019	0.00068	0.00070	-0.04112	0.11571	0.11470
$\pi(7)$	0.10825	0.14318	0.14270	-0.00298	0.00017	0.00018	-0.05041	0.13603	0.13933
$\pi(8)$	0.17971	0.24976	0.24974	0.02579	0.00004	0.00004	-0.06882	0.15714	0.16828
$\pi(9)$	0.57875	0.43137	0.43183	0.88898	0.00001	0.00001	1.30798	0.17672	0.20242
y(0)	0.00071	0.00093	0.01287	0.07746	0.48796	0.48793	-0.01397	0.03722	0.02755
y(1)	0.00142	0.00187	0.00184	0.05252	0.33086	0.33083	-0.01048	0.02793	0.02536
y(2)	0.00277	0.00366	0.00361	0.02023	0.12747	0.12749	-0.01225	0.03265	0.03029
y(3)	0.00533	0.00702	0.00694	0.00619	0.03896	0.03899	-0.01492	0.03976	0.03764
y(4)	0.01004	0.01324	0.01309	0.00172	0.01084	0.01084	-0.01826	0.04867	0.04687
y(5)	0.01862	0.02453	0.02426	0.00045	0.00293	0.00290	-0.02230	0.05944	0.05854
y(6)	0.03397	0.04476	0.04423	0.00037	0.00075	0.00076	-0.02706	0.07231	0.07266
y(7)	0.06102	0.08070	0.07947	-0.00596	0.00019	0.00020	-0.03364	0.08620	0.08964
y(8)	0.10262	0.14259	0.14087	0.05229	0.00004	0.00005	-0.04652	0.10086	0.10948
y(9)	0.33445	0.24922	0.24653	1.82409	0.00001	0.00001	0.89477	0.11479	0.13312
y(10)	0.42905	0.43148	0.42629	-1.02936	-0.00002	0.00000	0.30463	0.38016	0.36885

values, as illustrated in column 5 for $\pi(j)$ and $y(j)$, respectively; mostly negative thus not acceptable, as shown in column 8.

When some but not all of the $\lambda(n)$ are distinct, we compare in Table IV the proposed algorithm with simulation. Note that here we set $\lambda(2) = \lambda(4) = \lambda(5)$, and $\lambda(3) = \lambda(8)$, the situation different from the machine interference model in [17] in which the $\lambda(n)$ values are identical for the first several, and distinct from each other for the remaining rates.

Table IV. Steady-state probabilities at arbitrary and at departures for $N = 9$, $\lambda(0) = 1.5$, $\lambda(1) = 0.5$, $\lambda(2) = 0.7$, $\lambda(3) = 1.0$, $\lambda(4) = 0.7$, $\lambda(5) = 0.7$, $\lambda(6) = 0.8$, $\lambda(7) = 0.3$, $\lambda(8) = 1.0$

	M $\mu = 5$		D $\mu = 0.5$ ($mean = 2$)		C_2 $\mu \approx 1.24$ $SCV \approx 3.42$ $(\mu_1 = 7, \mu_2 = 0.5)$ $(p = 1/3)$		C_2 $\mu \approx 0.69$ $SCV \approx 0.67$ $(\mu_1 = 1, \mu_2 = 1.5)$ $(p = 2/3)$	
	New	Simu.	New	Simu.	New	Simu.	New	Simu.
$\pi(0)$	0.89505	0.89501	0.00115	0.00116	0.49612	0.49610	0.10249	0.10257
$\pi(1)$	0.08950	0.08954	0.00198	0.00198	0.14175	0.14170	0.08199	0.08209
$\pi(2)$	0.01253	0.01253	0.00418	0.00419	0.09458	0.09457	0.08527	0.08537
$\pi(3)$	0.00251	0.00251	0.01903	0.01907	0.08629	0.08630	0.13147	0.13148
$\pi(4)$	0.00035	0.00035	0.04231	0.04237	0.06341	0.06341	0.13547	0.13541
$\pi(5)$	0.00005	0.00005	0.08746	0.08749	0.04754	0.04756	0.13790	0.13786
$\pi(6)$	0.00001	0.00001	0.23654	0.23636	0.03874	0.03877	0.16313	0.16303
$\pi(7)$	0.00000	0.00000	0.12254	0.12251	0.01617	0.01618	0.06496	0.06493
$\pi(8)$	0.00000	0.00000	0.48481	0.48487	0.01541	0.01542	0.09732	0.09727
y(0)	0.74896	0.74893	0.00038	0.00511	0.29006	0.29005	0.04517	0.04522
y(1)	0.22469	0.22471	0.00198	0.00197	0.24862	0.24865	0.10840	0.10851
y(2)	0.02247	0.02248	0.00298	0.00297	0.11849	0.11848	0.08052	0.08062
y(3)	0.00315	0.00315	0.00951	0.00949	0.07568	0.07565	0.08691	0.08694
y(4)	0.00063	0.00063	0.03021	0.03012	0.07945	0.07942	0.12793	0.12788
y(5)	0.00009	0.00009	0.06245	0.06218	0.05956	0.05954	0.13023	0.13016
y(6)	0.00001	0.00001	0.14778	0.14701	0.04246	0.04249	0.13479	0.13476
y(7)	0.00001	0.00000	0.20416	0.20310	0.04727	0.04732	0.14313	0.14303
y(8)	0.00000	0.00000	0.24231	0.24111	0.01351	0.01350	0.06433	0.06430
y(9)	-0.00000	0.00000	0.29823	0.29692	0.02490	0.02491	0.07859	0.07859

D. Conclusion

Our numerical examples illustrate that the algorithm proposed in this chapter not only is stable for M/G/1/N queue, but also for other queueing models where some or all of the $\lambda(n)$ values are distinct. In these models, the $\lambda(n)$ may have relatively close values, thus making the method that Gupta and Rao developed in [18] fail or not work properly; or some of the $\lambda(n)$ may be identical to a number of different values, thus making the analogous technique by Gupta and Rao in [17] not convenient to apply in that the basic equations in [17] will be rewritten and a different model will be solved for each implementation. According to our experiments, if $N \leq 10$ and most of the $\lambda(n)$ values have the relative difference greater than 1%, and if $\varphi^{*(i)}$ can be easily obtained, then the proposed algorithm is fairly accurate, see Table III for example. For unusual situations that some $\lambda(n)$ values are extremely close to each other, the proposed algorithm may not work accurately, depending on the complexity of the Laplace-Stieltjes transform of the service time distribution and if N is large.

The proposed algorithm does not require a high demand for computer facilities. Our numerical results in Section II.C were generated by a program coded using *Maple V* and run on a Pentium 4 PC.

CHAPTER III

TIME-DEPENDENT PROBABILITIES FOR PHASE-TYPE RENEWAL PROCESSES

We assume that the PH distribution has representation given by $(\boldsymbol{\alpha}, T)$ where T is a $r \times r$ matrix and $\boldsymbol{\alpha}$ is a row vector of dimension r. (Note that T is a matrix with negative diagonal elements and nonnegative off-diagonal elements such that $T\boldsymbol{e} \leq \mathbf{0}$ with at least one row sum being strictly less than zero and where $\boldsymbol{e}$ is a vector of all 1s. Also, $\boldsymbol{\alpha}$ is nonnegative such that $\boldsymbol{\alpha}\boldsymbol{e} = 1$. The convention of using Greek letters for row vectors and non-Greek letters for column vectors will be followed in Chapter III and IV.) For an $(\boldsymbol{\alpha}, T)$ PH distribution, there exists a Markov process, $\{Y(t); t \geq 0\}$ that eventually dies whose infinitesimal generator is given by T; in other words, we assume that all states of the Y process are transient, which implies that T is nonsingular. There is also an associated Markov process, $\{J(t); t \geq 0\}$, with state space $\{1, \cdots, r\}$ that is formed by "restarting" the Y process whenever it dies according to the probabilities given by $\boldsymbol{\alpha}$. The infinitesimal generator for the J process is given by $Q^\circ = T + \boldsymbol{t}^\circ \cdot \boldsymbol{\alpha}$, where $\boldsymbol{t}^\circ$ is a nonnegative column vector satisfying $T\boldsymbol{e} + \boldsymbol{t}^\circ = \mathbf{0}$. (Note that $\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha}$ is a $r \times r$ matrix.) Denote by $U(t)$ the number of renewals in $(0, t]$, which is equivalent to the number of times that the Y process was restarted. The conditional joint probability distributions of the PH renewal process are defined by

$$P_{ij}(n,t) = P\{U(t) = n, J(t) = j | U(0) = 0, J(0) = i\}, \tag{3.1}$$

for $n \in \mathcal{N},\ t \in \mathcal{R}^+,\ i, j \in \{1, \cdots, r\}$. (See Neuts [36] for a more complete description of PH renewal processes.) In this chapter, we propose a procedure to generate the exact solution for the matrices $P(n,t)$, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$. Explicit expressions of $P(n,t)$ will be essential for the SM/PH/1/N queue addressed in Chapter IV.

A. Analysis

As given in Neuts [36], the matrices $P(n,t)$, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$, satisfy

$$P(0,t) = e^{Tt}, \quad \text{and} \tag{3.2}$$

$$P(n,t) = \int_0^t e^{T(t-x)}(\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha})P(n-1,x)dx, \qquad \text{for } n \geq 1. \tag{3.3}$$

Since the Laplace transform $P^L(0,s) = (sI - T)^{-1}$ and the transform of a convolution is the product of the transforms, it follows that

$$P^L(n,s) = (sI - T)^{-1}\left[(\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha})(sI - T)^{-1}\right]^n, \quad \text{for } n \in \mathcal{N}. \tag{3.4}$$

The probability mass function for the counting process will be denoted, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$, by

$$p_n(t) = P\{U(t) = n | U(0) = 0\} = \boldsymbol{\alpha} P(n,t)\boldsymbol{e}. \tag{3.5}$$

Let $\beta_n(t)$ denote the probability that there is at least n renewals in $(0,t]$, and by noting that the Laplace transform of the PH distribution $(\boldsymbol{\alpha}, T)$ is $\boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}^\circ/s$, we have

$$\beta_0^L(s) = \frac{1}{s}; \tag{3.6}$$

$$\beta_n^L(s) = \left[\boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}^\circ\right]^{n-1}\frac{1}{s}\boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}^\circ, \qquad \text{for } n \geq 1. \tag{3.7}$$

Since $p_n(t)$ is given by $\beta_n(t) - \beta_{n+1}(t)$, its Laplace transform is given, for $n \in \mathcal{N}$, by

$$p_n^L(s) = \frac{1}{s}\left[1 - \boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}^\circ\right]\left[\boldsymbol{\alpha}(sI - T)^{-1}\boldsymbol{t}^\circ\right]^n, \; s \in \mathcal{R}^+. \tag{3.8}$$

B. Example

The availability of software that performs symbolic mathematics has the potential to make these computations straight forward. For example, consider the PH distribution with representation of $\boldsymbol{\alpha} = (3/5, 2/5)$ and

$$T = \begin{pmatrix} -3 & 2 \\ 1 & -4 \end{pmatrix}.$$

The *Maple* software package has the capability of performing the matrix operations given by Eq. (3.4) using symbolic logic. It also has an inverse Laplace transform function so that the matrix of Eq. (3.1) can be obtained from (3.4) with very few programming statements. Applying these Maple functions to the above example yields

$$P(0,t) = \begin{pmatrix} \frac{1}{3}e^{-5t} + \frac{2}{3}e^{-2t} & -\frac{2}{3}e^{-5t} + \frac{2}{3}e^{-2t} \\ -\frac{1}{3}e^{-5t} + \frac{1}{3}e^{-2t} & \frac{2}{3}e^{-5t} + \frac{1}{3}e^{-2t} \end{pmatrix} \quad \text{and}$$

$$P(1,t) = \begin{pmatrix} -\frac{1}{9}t\,e^{-5t} + \frac{32}{135}e^{-5t} + \frac{64}{45}t\,e^{-2t} - \frac{32}{135}e^{-2t} & \frac{2}{9}t\,e^{-5t} + \frac{56}{135}e^{-5t} + \frac{64}{45}t\,e^{-2t} - \frac{56}{135}e^{-2t} \\ \frac{1}{9}t\,e^{-5t} - \frac{44}{135}e^{-5t} + \frac{32}{45}t\,e^{-2t} + \frac{44}{135}e^{-2t} & -\frac{2}{9}t\,e^{-5t} - \frac{32}{135}e^{-5t} + \frac{32}{45}t\,e^{-2t} + \frac{32}{135}e^{-2t} \end{pmatrix}.$$

It is equally simple to obtain the time-dependent probabilities for the counting process by applying these built-in Maple functions to Eq. (3.8) thus yielding

$$p_0(t) = -\frac{1}{15}e^{-5t} + \frac{16}{15}e^{-2t},$$
$$p_1(t) = \frac{1}{45}t\,e^{-5t} + \frac{112}{675}e^{-5t} + \frac{512}{225}t\,e^{-2t} - \frac{112}{675}e^{-2t}.$$

As n increases in (3.8) the expressions become increasingly complex but the Maple operations are straight-forward.

C. Computational Considerations and Special Cases

The difficulty with using symbolic logic software is that numerical accuracy can quickly become problematic. As the number of phases increases, round-off errors will begin to accumulate; however, even with two phases, round-off errors will occur when the eigenvalues are unequal but close to each other. For this reason, we obtain closed-form expressions (i.e., computationally attractive expressions that do not include the Laplace transform) for two cases involving the Coxian distribution. These cases were chosen so that a distribution that fits any given mean and variance can be used.

A Coxian distribution [13] is a PH distribution where the transitions after each phase are either to the next phase or out, and the initial probability vector, $\boldsymbol{\alpha}$, has a one in the first component and zeros elsewhere. (In other words, a Coxian distribution always starts with the first phase and then either moves to the next phase or is finished. Figure 1 illustrates the general Coxian model.) A Coxian distribution with two phases can be used to construct a distribution with any given positive mean EX and any given SCV that is greater than or equal to 1/2, with parameters determined by

$$p = \frac{1}{2\,SCV}; \qquad \mu_1 = \frac{2}{EX}; \qquad \mu_2 = \frac{1}{EX\,SCV}. \tag{3.9}$$

A Coxian distribution with r phases, called a Cox-r model, with equal means for each phase and transition probabilities after the first stage equal to one (i.e., $p_2 = p_3 = \cdots = p_{r-1} = 1$) can be used to represent a distribution with an arbitrary mean and with SCV greater than or equal to $1/r$ and less than $1/(r-1)$. The parameters are given by

$$p = 1 - \frac{2\,r\,SCV + r - 2 - \sqrt{r^2 - 4\,rSCV + 4}}{2(SCV+1)(r-1)}; \tag{3.10}$$

$$\mu = \frac{1 + p(r-1)}{EX}. \tag{3.11}$$

(See Altiok [2] for a detailed description.) Thus, Case 1 below can be used for a renewal process with SCV greater than 1/2, and Case 2 below can be used for SCV less than 1.0. (There is an overlap in model choices when the SCV is between 1/2 and 1.)

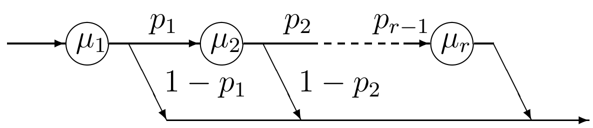

Fig. 1. A General Cox-r Model

Case 1. A Coxian Model with $SCV > 1/2$

We begin with the Cox-2 distribution with $\boldsymbol{\alpha} = (1, 0)$ and

$$T = \begin{pmatrix} -\mu_1 & \mu_1 p \\ 0 & -\mu_2 \end{pmatrix}, \tag{3.12}$$

where $\mu_1 \neq \mu_2$. If $\mu_1 = \mu_2$, Case 2 can be used. Using Eq. (3.4), we obtain an expression for the Laplace transform:

$$P^L(0, s) = \begin{pmatrix} \frac{1}{s+\mu_1} & \frac{\mu_1 p}{(s+\mu_1)(s+\mu_2)} \\ 0 & \frac{1}{s+\mu_2} \end{pmatrix}; \tag{3.13}$$

$$P^L(n, s) = \begin{pmatrix} \frac{\mu_1^n[(1-p)s+\mu_2]^n}{(s+\mu_1)^{n+1}(s+\mu_2)^n} & \frac{\mu_1^{n+1}p[(1-p)s+\mu_2]^n}{(s+\mu_1)^{n+1}(s+\mu_2)^{n+1}} \\ \frac{\mu_1^{n-1}\mu_2[(1-p)s+\mu_2]^{n-1}}{(s+\mu_1)^n(s+\mu_2)^n} & \frac{\mu_1^n\mu_2 p[(1-p)s+\mu_2]^{n-1}}{(s+\mu_1)^n(s+\mu_2)^{n+1}} \end{pmatrix}, \quad \text{for } n \geq 1. \tag{3.14}$$

If symbolic logic software is available and μ_1 is not close to μ_2, the inverse Laplace transform function could be used with (3.13) and (3.14) to obtain the expressions for the conditional joint probability distributions in t. If symbolic logic software is not available or if μ_1 and μ_2 are close, then the method of partial fraction expansion could be used to obtain the inverse. For an $(\boldsymbol{\alpha}, T)$ PH distribution, each component of the matrix function $(sI - T)^{-1}$ is of the form of a rational function $A(s)/B(s)$, where both $A(s)$ and $B(s)$ are polynomials in s and $A(s)$ is of degree less than that of $B(s)$. Thus, $P^L(n, s)$, for fixed $n \in \mathcal{N}$, is a matrix of proper rational functions of s, and the method of partial fraction expansion can be used. This leads to the following:

$$P(0,t) = \begin{pmatrix} e^{-\mu_1 t} & \dfrac{\mu_1 p}{\mu_1 - \mu_2}(-e^{-\mu_1 t} + e^{-\mu_2 t}) \\ 0 & e^{-\mu_2 t} \end{pmatrix}; \tag{3.15}$$

$$P(n,t) = \begin{pmatrix} \mu_1^n f_{1,1,n}(t) & \mu_1^{n+1} p\, f_{1,2,n}(t) \\ \mu_1^{n-1} \mu_2 f_{2,1,n}(t) & \mu_1^n \mu_2 p\, f_{2,2,n}(t) \end{pmatrix}, \qquad \text{for } n \geq 1, \tag{3.16}$$

where

$$f_{m_1,m_2,n}(t) = \sum_{h=1}^{2} \sum_{j=1}^{a(m_1,m_2,n,h)} b(m_1, m_2, n, h, j) t^{j-1} e^{-\mu_h t}. \tag{3.17}$$

The parameters are determined by

$$a(m_1, m_2, n, h) = \begin{cases} n + 2 - h, & \text{for } m_1 = 1, m_2 = 1; \\ n + 1, & \text{for } m_1 = 1, m_2 = 2; \\ n, & \text{for } m_1 = 2, m_2 = 1; \\ n + h - 1, & \text{for } m_1 = 2, m_2 = 2, \end{cases} \tag{3.18}$$

$$b(m_1, m_2, n, h, j) = \frac{\bar{b}(m_1, m_2, n, h, j)}{(a(m_1, m_2, n, h) - j)!(j-1)!}, \tag{3.19}$$

$$\bar{b}(m_1, m_2, n, h, j) \tag{3.20}$$

$$
= \begin{cases} \dfrac{d^{v_1}}{ds^{v_1}}\left\{\dfrac{1}{(s+\mu_2)^{a(m_1,m_2,n,2)-u}}\left[(1-p)+\dfrac{p\,\mu_2}{s+\mu_2}\right]^u\right\}_{s=-\mu_1}, & \text{for } h=1; \\ \dfrac{d^{v_2}}{ds^{v_2}}\left\{\dfrac{1}{(s+\mu_1)^{a(m_1,m_2,n,1)-u}}\left[(1-p)+\dfrac{\mu_2-(1-p)\mu_1}{s+\mu_1}\right]^u\right\}_{s=-\mu_2}, & \\ & \text{for } h=2 \end{cases}
$$

$$
= \begin{cases} \displaystyle\sum_{\tilde{u}=0}^{u}\binom{\tilde{u}}{u}(1-p)^{\tilde{u}}(p\mu_2)^{u-\tilde{u}}(-\mu_1+\mu_2)^{-a(m_1,m_2,n,2)+\tilde{u}}, & \text{for } h=1, v_1=0; \\ \displaystyle\sum_{\tilde{u}=0}^{u}\binom{\tilde{u}}{u}(1-p)^{\tilde{u}}(p\mu_2)^{u-\tilde{u}}\prod_{\tilde{v}=0}^{v_1-1}[-a(m_1,m_2,n,2)+ & \\ \quad \tilde{u}-\tilde{v}](-\mu_1+\mu_2)^{-a(m_1,m_2,n,2)+\tilde{u}-v_1}, & \text{for } h=1, v_1\geq 1; \\ \displaystyle\sum_{\tilde{u}=0}^{u}\binom{\tilde{u}}{u}(1-p)^{\tilde{u}}[\mu_2-(1-p)\mu_1]^{u-\tilde{u}}(\mu_1-\mu_2)^{-a(m_1,m_2,n,1)+\tilde{u}}, & \\ & \text{for } h=2, v_2=0; \\ \displaystyle\sum_{\tilde{u}=0}^{u}\binom{\tilde{u}}{u}(1-p)^{\tilde{u}}[\mu_2-(1-p)\mu_1]^{u-\tilde{u}}\prod_{\tilde{v}=0}^{v_2-1}[-a(m_1,m_2,n,1)+ & \\ \quad \tilde{u}-\tilde{v}](\mu_1-\mu_2)^{-a(m_1,m_2,n,1)+\tilde{u}-v_2}, & \text{for } h=2, v_2\geq 1, \end{cases}
$$

where $u=n-m_1+1$, and $v_h=a(m_1,m_2,n,h)-j$ for $h=1,2$.

For a particular application, it may be that only the distribution of the counting process is required. In this situation, Eq. (3.5) and together with Eqs. (3.13) and (3.14) to obtain

$$
p_0^L(s)=\frac{s+\mu_2+\mu_1 p}{(s+\mu_1)(s+\mu_2)}; \tag{3.21}
$$

$$
p_n^L(s)=\frac{\mu_1^n(s+\mu_2+\mu_1 p)[(1-p)s+\mu_2]^n}{(s+\mu_1)^{n+1}(s+\mu_2)^{n+1}}, \qquad \text{for } n\geq 1. \tag{3.22}
$$

Again, if symbolic logic software is available, Eqs. (3.21) and (3.22) can be used directly; otherwise, the following provides the closed-form solutions.

$$
p_0(t)=\left(1-\frac{\mu_1 p}{\mu_1-\mu_2}\right)e^{-\mu_1 t}+\frac{\mu_1 p}{\mu_1-\mu_2}e^{-\mu_2 t}; \tag{3.23}
$$

$$p_n(t) = \sum_{h=1}^{2}\sum_{j=1}^{n+1} \frac{\mu_1^n \hat{b}(n,h,j)}{(n+1-j)!(j-1)!} t^{j-1} e^{-\mu_h t}, \qquad \text{for } n \geq 1, \tag{3.24}$$

where

$$\hat{b}(n,h,j) \tag{3.25}$$
$$= \begin{cases} \dfrac{d^{n+1-j}}{ds^{n+1-j}}\left\{\left(1+\dfrac{\mu_1 p}{s+\mu_2}\right)\left[(1-p)+\dfrac{p\,\mu_2}{s+\mu_2}\right]^n\right\}_{s=-\mu_1}, & \text{for } h=1; \\ \dfrac{d^{n+1-j}}{ds^{n+1-j}}\left\{\left[1+\dfrac{\mu_2-(1-p)\mu_1}{s+\mu_1}\right]\left[(1-p)+\dfrac{\mu_2-(1-p)\mu_1}{s+\mu_1}\right]^n\right\}_{s=-\mu_2}, & \text{for } h=2 \end{cases}$$

$$= \begin{cases} \sum_{\tilde{n}=0}^{n}\binom{\tilde{n}}{n}(1-p)^{\tilde{n}}(p\mu_2)^{n-\tilde{n}}[(-\mu_1+\mu_2)^{\tilde{n}-n}+\mu_1 p(-\mu_1+\mu_2)^{\tilde{n}-n-1}], & \\ & \text{for } h=1, j=n+1; \\ \sum_{\tilde{n}=0}^{n}\binom{\tilde{n}}{n}(1-p)^{\tilde{n}}(p\mu_2)^{n-\tilde{n}}\prod_{\tilde{v}=0}^{n-j}[(\tilde{n}-n-\tilde{v})(-\mu_1+\mu_2)^{\tilde{n}-2n-1+j}+ & \\ \quad \mu_1 p(\tilde{n}-n-1-\tilde{v})(-\mu_1+\mu_2)^{\tilde{n}-2n-2+j}], & \text{for } h=1, 1\leq j\leq n; \\ \sum_{\tilde{n}=0}^{n}\binom{\tilde{n}}{n}(1-p)^{\tilde{n}}(\mu_2-(1-p)\mu_1)^{n-\tilde{n}}[(\mu_1-\mu_2)^{\tilde{n}-n}+ & \\ \quad (\mu_2-(1-p)\mu_1)(\mu_1-\mu_2)^{\tilde{n}-n-1}], & \text{for } h=2, j=n+1; \\ \sum_{\tilde{n}=0}^{n}\binom{\tilde{n}}{n}(1-p)^{\tilde{n}}(\mu_2-(1-p)\mu_1)^{n-\tilde{n}}\prod_{\tilde{v}=0}^{n-j}[(\tilde{n}-n-\tilde{v})(\mu_1-\mu_2)^{\tilde{n}-2n-1+j}+ & \\ \quad (\mu_2-(1-p)\mu_1)(\tilde{n}-n-1-\tilde{v})(\mu_1-\mu_2)^{\tilde{n}-2n-2+j}], & \text{for } h=2, 1\leq j\leq n. \end{cases}$$

Case 2. A Coxian Model with $SCV < 1$

A simplified version of the Cox-r distribution is considered since it is general enough to model a distribution with any $SCV < 1.0$ and simple enough to obtain the inverse Laplace transform. Our distribution has $\alpha = (1, 0, \cdots, 0)$ and T with the form

$$T = \begin{pmatrix} -\mu & \mu p & & \mathbf{0} \\ & -\mu & \mu & \\ & & \ddots & \ddots \\ \mathbf{0} & & & -\mu \end{pmatrix}, \tag{3.26}$$

Eq. (3.4) yields the Laplace transform as:

$$P^L(0,s) = \begin{pmatrix} \frac{1}{s+\mu} & \frac{\mu p}{(s+\mu)^2} & \frac{\mu^2 p}{(s+\mu)^3} & \cdots & \frac{\mu^{r-1}p}{(s+\mu)^r} \\ & \frac{1}{s+\mu} & \frac{\mu}{(s+\mu)^2} & \cdots & \frac{\mu^{r-2}}{(s+\mu)^{r-1}} \\ & & \frac{1}{s+\mu} & \cdots & \frac{\mu^{r-3}}{(s+\mu)^{r-2}} \\ & & & \ddots & \vdots \\ \mathbf{0} & & & & \frac{1}{s+\mu} \end{pmatrix}; \tag{3.27}$$

$$P^L(n,s) = \tag{3.28}$$
$$\begin{pmatrix} \frac{\mu^n}{s+\mu} f(s)^n & \frac{\mu^{n+1}p}{(s+\mu)^2} f(s)^n & \frac{\mu^{n+2}p}{(s+\mu)^3} f(s)^n & \cdots & \frac{\mu^{n+r-1}p}{(s+\mu)^r} f(s)^n \\ \frac{\mu^{n+r-2}}{(s+\mu)^r} f(s)^{n-1} & \frac{\mu^{n+r-1}p}{(s+\mu)^{r+1}} f(s)^{n-1} & \frac{\mu^{n+r}p}{(s+\mu)^{r+2}} f(s)^{n-1} & \cdots & \frac{\mu^{n+2r-3}p}{(s+\mu)^{2r-1}} f(s)^{n-1} \\ \frac{\mu^{n+r-3}}{(s+\mu)^{r-1}} f(s)^{n-1} & \frac{\mu^{n+r-2}p}{(s+\mu)^r} f(s)^{n-1} & \frac{\mu^{n+r-1}p}{(s+\mu)^{r+1}} f(s)^{n-1} & \cdots & \frac{\mu^{n+2r-4}p}{(s+\mu)^{2r-2}} f(s)^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\mu^n}{(s+\mu)^2} f(s)^{n-1} & \frac{\mu^{n+1}p}{(s+\mu)^3} f(s)^{n-1} & \frac{\mu^{n+2}p}{(s+\mu)^4} f(s)^{n-1} & \cdots & \frac{\mu^{n+r-1}p}{(s+\mu)^{r+1}} f(s)^{n-1} \end{pmatrix},$$

for $n \geq 1$, where

$$f(s) = \frac{1-p}{s+\mu} + \frac{p\mu^{r-1}}{(s+\mu)^r}. \tag{3.29}$$

The components of this matrix are again a ratio of polynomials in s such that partial fraction expansion can be used to obtain the inverse Laplace transform yielding the conditional joint probability distributions in t. Therefore,

$$P(0,t) = \begin{pmatrix} e^{-\mu t} & \mu\, p\, t\, e^{-\mu t} & \frac{\mu^2 p}{2!} t^2 e^{-\mu t} & \cdots & \frac{\mu^{r-1} p}{(r-1)!} t^{r-1} e^{-\mu t} \\ & e^{-\mu t} & \mu\, t\, e^{-\mu t} & \cdots & \frac{\mu^{r-2}}{(r-2)!} t^{r-2} e^{-\mu t} \\ & & e^{-\mu t} & \cdots & \frac{\mu^{r-3}}{(r-3)!} t^{r-3} e^{-\mu t} \\ & & & \ddots & \vdots \\ \mathbf{0} & & & & e^{-\mu t} \end{pmatrix}; \tag{3.30}$$

$$P(n,t) = \tag{3.31}$$

$$\begin{pmatrix} \mu^n g_{1,n}(t) & \mu^{n+1} p g_{2,n}(t) & \mu^{n+2} p g_{3,n}(t) & \cdots & \mu^{n+r-1} p g_{r,n}(t) \\ \mu^{n+r-2} g_{r,n-1}(t) & \mu^{n+r-1} p g_{r+1,n-1}(t) & \mu^{n+r} p g_{r+2,n-1}(t) & \cdots & \mu^{n+2r-3} p g_{2r-1,n-1}(t) \\ \mu^{n+r-3} g_{r-1,n-1}(t) & \mu^{n+r-2} p g_{r,n-1}(t) & \mu^{n+r-1} p g_{r+1,n-1}(t) & \cdots & \mu^{n+2r-4} p g_{2r-2,n-1}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu^n g_{2,n-1}(t) & \mu^{n+1} p g_{3,n-1}(t) & \mu^{n+2} p g_{4,n-1}(t) & \cdots & \mu^{n+r-1} p g_{r+1,n-1}(t) \end{pmatrix},$$

for $n \geq 1$, where

$$g_{m,n}(t) = \sum_{j=0}^{n} c(m,n,j) t^{rn-rj+j+m-1} e^{-\mu t}, \tag{3.32}$$

and

$$c(m,n,j) = \binom{j}{n} \frac{(1-p)^j (p\mu^{r-1})^{n-j}}{(rn-rj+j+m-1)!}. \tag{3.33}$$

We also obtain the distribution for the counting process analogous to the preceding case:

$$p_0^L(s) = \frac{1}{s+\mu} + \frac{\mu\, p}{s(s+\mu)} - \frac{\mu^r p}{s(s+\mu)^r}; \tag{3.34}$$

$$p_n^L(s) = \Big[\frac{1}{s+\mu} + \frac{\mu\, p}{s(s+\mu)} - \frac{\mu^r p}{s(s+\mu)^r}\Big] \mu^n f(s)^n, \qquad \text{for } n \geq 1, \tag{3.35}$$

and

$$p_0(t) = (1-p)e^{-\mu t} - \mu^r p \sum_{j=1}^{r} \hat{b}(r,j) t^{j-1} e^{-\mu t}; \tag{3.36}$$

$$\begin{aligned} p_n(t) = \mu^n \sum_{i=0}^{n} \binom{i}{n} (1-p)^i (p\,\mu^{r-1})^{n-i} \Big\{ & \frac{1}{(rn-ri+i)!} t^{rn-ri+i} e^{-\mu t} + \\ & \mu\, p \Big[\sum_{j=1}^{rn-ri+i+1} \hat{b}(rn-ri+i+1, j) t^{j-1} e^{-\mu t} \Big] - \\ & \mu^r p \Big[\sum_{j=1}^{rn-ri+r+i} \hat{b}(rn-ri+r+i, j) t^{j-1} e^{-\mu t} \Big] \Big\}, \qquad \text{for } n \geq 1, \end{aligned} \tag{3.37}$$

where

$$\hat{b}(r,j) = \frac{1}{(j-1)!} \left(\frac{1}{s}\right)^{(r-j)}_{s=-\mu} = -\frac{\mu^{-r+j-1}}{(j-1)!}. \tag{3.38}$$

D. Conclusion

We proposed an approach to generate the distribution functions $P(n,t)$ and $p_n(t)$, for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$, for the PH renewal process. In particular, we provided closed-form expressions for the Cox-2 and simplified Cox-r models, which are applicable for a two-moment approximation of all distributions with $SCV > 0$. Both $p_n(t)$ and components of matrices $P(n,t)$ have up to $r(n+1)$ terms. Note that the parameters $\bar{b}(m_1, m_2, n, i, j)$ given by (3.20) and $\hat{b}(n,i,j)$ by (3.25) may become extremely large and, thus, subject to round-off error when n is large and the values of μ_1 and μ_2 are very close to each other.

CHAPTER IV

COMPUTATIONAL ANALYSIS FOR SM/PH/1/N QUEUE

The analysis for the SM/PH/1/N queue is carried out by using the embedded Markov approach and the matrix-analytic method. Computational algorithms are presented for both the system size probabilities and the waiting time distributions. A new approach applying the SM/PH/1/N model to tandem closed queueing networks will be attempted in Chapter V.

We shall use the Kronecker product of matrices throughout this chapter. As a reminder, assume that $A = (a_{ij})$ is an $m \times n$ matrix and B is an $r \times s$ matrix. Then the Kronecker product $A \otimes B$ is a matrix of dimension $mr \times ns$ and is given, using a block-partition form, as

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

In addition, $I_{(n)}$ denotes the $n \times n$ identity matrix and $\boldsymbol{e}_{(n)}$ denotes an n-dimensioned column vector of all ones; however, the subscripts will be omitted when the dimension is obvious from the context.

A. Problem Statement

For the SM/PH/1/N queue, let $T_0 = 0, T_1, T_2, \cdots$ be the instants of successive arrivals. Denote by $V_n \in \{1, 2, \cdots, \omega\}$ the state of the Markov renewal process of arrivals immediately after the n^{th} arrival. In other words, the stochastic process $(V, T) = \{V_n, T_n; n \in \mathcal{N}\}$ is a Markov renewal process with state space $E_\omega = \{1, 2, \cdots, \omega\}$, where ω is assumed finite and arrivals to the queueing system occur at every transition epoch of (V,T). The semi-Markov kernel associated with (V,T) is denoted by Q and is defined by

$$Q_{ij}(t) = \mathrm{P}\{V_{n+1} = j, T_{n+1} - T_n \leq t | V_n = i\} \quad \text{for } i,j \in E_\omega \text{ and } t \in \mathcal{R}^+. \tag{4.1}$$

The tilde will be used to denote limiting values for matrices so that $\tilde{Q}$ represents the probability transition matrix of the embedded Markov chain, $\{V_n; n \in \mathcal{N}\}$; that is,

$$\tilde{Q} = \lim_{t\to\infty} Q(t). \tag{4.2}$$

We assume that $\tilde{Q}$ is irreducible.

The service time has a PH distribution with representation $(\boldsymbol{\alpha}, T)$ of dimension r and mean service time given by $\eta = -\boldsymbol{\alpha}T^{-1}\boldsymbol{e}$. For an $(\boldsymbol{\alpha}, T)$ PH distribution, there exists a Markov process, $\{Z(t); t \geq 0\}$ that eventually dies whose infinitesimal generator is given by T, which implies that T is nonsingular. If the Z process is immediately "renewed" whenever it dies according to the probabilities given by $\boldsymbol{\alpha}$, a recurrent Markov process will be formed, which we denote by $\{J(t); t \geq 0\}$. The J process has state space $E_r = \{1, \cdots, r\}$ and its infinitesimal generator is given by $\overline{T} = T + \boldsymbol{t}^\circ \cdot \boldsymbol{\alpha}$, where $\boldsymbol{t}^\circ$ is a nonnegative column vector satisfying $T\boldsymbol{e} + \boldsymbol{t}^\circ = \mathbf{0}$. (Note that $\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha}$ is a $r \times r$ matrix.) The number of times that the process is renewed in the interval $(0,t]$ is denoted by $M(t)$, thus the M process is a PH Renewal Process which we studied in Chapter III. The time-dependent probabilities associated with this renewal process are defined, for $t \in \mathcal{R}^+$, $n \in \mathcal{N}$, $i,j \in E_r$, by

$$P_{ij}(n,t) = P\{M(t) = n, J(t) = j | M(0) = 0, J(0) = i\}. \tag{4.3}$$

If we sum over the values of n in (4.3), we have the time-dependent probabilities for the Markov process J yielding

$$\sum_{n=0}^{\infty} P(n,t) = e^{\overline{T}t} \quad \text{for } t \in \mathcal{R}^+. \tag{4.4}$$

The Laplace transform of Eq. (4.3) will be needed in our analysis, which is given by, for each $n \in \mathcal{N}$ and $s \in \mathcal{R}^+$,

$$\begin{aligned} P^L(n,s) &\equiv \int_0^\infty e^{-st} P(n,t)dt \\ &= (sI - T)^{-1}\Big((\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha})(sI - T)^{-1}\Big)^n. \end{aligned} \tag{4.5}$$

Denote by U_n the number of customers in the system immediately prior to the n^{th} arrival and J_n the phase of service immediately after the n^{th} arrival. The process $(U, J, V, T) = \{U_n, J_n, V_n, T_n; n \in \mathcal{N}\}$ is a Markov renewal process on the lexicographically ordered state space $\hat{E} = E_N \times E_r \times E_\omega$, where $E_N = \{0, 1, \cdots, N-1\}$, with semi-Markov kernel given by P_N. (We ask the reader to be cautious here, the subscripted matrix $P_N(t)$ for $t \in \mathcal{R}^+$ refers the time-dependent probabilities for the Markov renewal process representing the entire queueing system, and the matrix $P(n,t)$ for $t \in \mathcal{R}^+$ and $n \in \mathcal{N}$ refers the time-dependent probabilities for the PH renewal process representing the service system.) The semi-Markov kernel for the infinite capacity SM/PH/1 system is given in [37], so after appropriate modifications to their system, we have the following definition of the semi-Markov kernel,

$$P_N(t) = \begin{pmatrix} B_0(t) & A_0(t) & & & \mathbf{0} \\ B_1(t) & A_1(t) & A_0(t) & & \\ \vdots & \vdots & \vdots & \ddots & \\ B_{N-2}(t) & A_{N-2}(t) & \cdots & A_1(t) & A_0(t) \\ B_{N-1}(t) & A_{N-1}(t) & \cdots & A_2(t) & A_1(t) + A_0(t) \end{pmatrix} \quad \text{for } t \in \mathcal{R}^+ , \tag{4.6}$$

where each component is a $r\omega \times r\omega$ submatrix given by, for $k \in E_N$ and $t \in \mathcal{R}^+$,

$$A_k(t) = \int_0^t P(k,x) \otimes dQ(x)\,, \tag{4.7}$$

$$B_k(t) = \left(\int_0^t e^{\overline{T}x} \otimes dQ(x) - \sum_{n=0}^{k} A_n(t) \right) \left((\boldsymbol{e}_{(r)} \cdot \boldsymbol{\alpha}) \otimes I_{(\omega)} \right) . \tag{4.8}$$

The transition matrix of the embedded Markov chain $\{U_n, J_n, V_n; n \in \mathcal{N}\}$, is given by taking the limit of $P_N(t)$ as t goes to infinity yielding

$$\tilde{P}_N = \begin{pmatrix} \tilde{B}_0 & \tilde{A}_0 & & & \mathbf{0} \\ \tilde{B}_1 & \tilde{A}_1 & \tilde{A}_0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \tilde{B}_{N-2} & \tilde{A}_{N-2} & \cdots & \tilde{A}_1 & \tilde{A}_0 \\ \tilde{B}_{N-1} & \tilde{A}_{N-1} & \cdots & \tilde{A}_2 & \tilde{A}_1 + \tilde{A}_0 \end{pmatrix}, \tag{4.9}$$

where, for $k \in E_N$, $\tilde{A}_k = \lim_{t\to\infty} A_k(t)$ and $\tilde{B}_k = \lim_{t\to\infty} B_k(t)$. Its invariant probability vector satisfies

$$\boldsymbol{\pi}\tilde{P}_N = \boldsymbol{\pi} \qquad \text{and} \qquad \boldsymbol{\pi}\boldsymbol{e} = 1, \tag{4.10}$$

where $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_{N-1})$, and is partitioned into N vectors, each of dimension $r\omega$. Thus, for example, the scalar given by $\boldsymbol{\pi}_k\boldsymbol{e}$ equals the probability that an arriving customer (who may or may not enter the system) sees k customers in the queueing system upon arrival (unless $k = N-1$, in which case $\boldsymbol{\pi}_{N-1}\boldsymbol{e}$ is the probability that the arriving customer sees $N-1$ or N customers in the system).

Let $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N)$ denote the long-run state probabilities of the queueing system at an arbitrary point in time. In other words,

$$y_{kjv} = \lim_{t\to\infty} P\{U(t) = k, J(t) = j, V(t) = v\}, \tag{4.11}$$

where for a time $t \in \mathcal{R}^+$, $U(t)$ denotes the number of customers in the queueing system, $J(t)$ denotes the phase of the server process, and $V(t)$ denotes the state of the arrival process. In the above, if for some time t, $U(t) = 0$, then $J(t)$ is undefined so that $\boldsymbol{y}_0$ is a vector with dimension ω and $\boldsymbol{y}_k$, for $k = 1, \cdots, N$, are vectors with dimension $r\omega$.

To obtain expressions for $\boldsymbol{y}$, we need the invariant probability vector for the arrival process and its fundamental mean. That is, define $\boldsymbol{\nu}$ by

$$\boldsymbol{\nu}\tilde{Q} = \boldsymbol{\nu} \qquad \text{and} \qquad \boldsymbol{\nu}\boldsymbol{e} = 1, \tag{4.12}$$

and let λ be mean arrival rate, i.e., the inverse of the mean time between arrivals; namely,

$$\lambda = \frac{1}{\boldsymbol{\nu}\int_0^\infty t\, dQ(t)\boldsymbol{e}}\ , \tag{4.13}$$

where $\boldsymbol{\nu}$ is a row vector of dimension ω and $\boldsymbol{e}$ is a column vector of dimension ω. By making suitable adjustments to the results in [36] for the infinite capacity SM/PH/1 system, we have the following expressions for the long-run probabilities of the SM/PH/1/N queue

$$y_{0,v} = \lambda \sum_{i=0}^{N-1} \sum_{r=1}^{r} \sum_{\ell=1}^{\omega} \pi_{i,r,\ell} \int_0^\infty \sum_{n=i+1}^{\infty} (P(n,t)\boldsymbol{e})_r \left(\tilde{Q} - Q(t)\right)_{\ell\, v}\, dt; \tag{4.14}$$

$$y_{k,j,v} = \lambda \sum_{i=k-1}^{N-1} \sum_{r=1}^{r} \sum_{\ell=1}^{\omega} \pi_{i,r,\ell} \int_0^\infty P_{r\,j}(i-k+1,t) \left(\tilde{Q} - Q(t)\right)_{\ell\, v}\, dt, \tag{4.15}$$

for $k = 1, \cdots, N$, $j \in E_r$, $v \in E_\omega$, and where the subscripts refer to the associated matrix or vector component. Rewriting (4.14) and (4.15), respectively, in concise form and taking advantage of (4.4), we obtain

$$\begin{aligned}\boldsymbol{y}_0 &= \lambda \sum_{i=0}^{N-1} \boldsymbol{\pi}_i \left(\int_0^\infty \sum_{n=i+1}^{\infty} P(n,t) \otimes \left(\tilde{Q} - Q(t)\right)\, dt \right) \left(\boldsymbol{e}_{(r)} \otimes I_{(\omega)}\right) \\ &= \lambda \sum_{i=0}^{N-1} \boldsymbol{\pi}_i \left(\int_0^\infty \left(e^{\overline{T}t} - \sum_{n=0}^{i} P(n,t) \right) \otimes \left(\tilde{Q} - Q(t)\right)\, dt \right) \left(\boldsymbol{e}_{(r)} \otimes I_{(\omega)}\right) \end{aligned} \tag{4.16}$$

and

$$\boldsymbol{y}_k = \lambda \sum_{i=k-1}^{N-1} \boldsymbol{\pi}_i \int_0^\infty P(i-k+1,t) \otimes \left(\tilde{Q} - Q(t)\right) \, dt \ \text{ for } k = 1, \cdots, N. \tag{4.17}$$

The long-run probability that the system size is k at an arbitrary point in time is given by the vector product $\boldsymbol{y}_k \boldsymbol{e}$ for $k = 0, \cdots, N$.

We are interested in obtaining both the long-run time in the queue and the virtual time in the queue (i.e., the time a customer would wait if the customer arrived to the system at an arbitrary time). The distribution function for the actual waiting per customer and the virtual waiting time will be denoted by $W(\cdot)$ and $\hat{W}(\cdot)$, respectively. Extending [37, eq.(28)] to the SM/PH/1/N queue, we obtain the Laplace-Stieltjes transform of $W(\cdot)$, by

$$\begin{aligned} W^*(s) &= \sum_{i=0}^{N-1} \boldsymbol{\pi}_i (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) \left((sI - T)^{-1} (\boldsymbol{t}^\circ \cdot \boldsymbol{\alpha})\right)^i \boldsymbol{e}_{(r)} \\ &= \boldsymbol{\pi}_0 (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) \boldsymbol{e}_{(r)} + \sum_{i=1}^{N-1} \boldsymbol{\pi}_i (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) P^L(i-1,s) \boldsymbol{t}^\circ \\ &= \boldsymbol{\pi}_0 \boldsymbol{e} + \sum_{i=1}^{N-1} \boldsymbol{\pi}_i (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) (sI - T)^{-1} \boldsymbol{t}^\circ g^{i-1}(s), \end{aligned} \tag{4.18}$$

where $g(s) = \boldsymbol{\alpha}(sI - T)^{-1} \boldsymbol{t}^\circ$ for $s \in \mathcal{R}^+$. Upon differentiation and setting $s = 0$, we obtain the expected waiting time in the queue as

$$\begin{aligned} E[W] &= -\frac{d}{ds} W^*(s)|_{s=0} \\ &= -\sum_{i=1}^{N-1} \boldsymbol{\pi}_i (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) [-(sI - T)^{-2} \boldsymbol{t}^\circ g^{i-1}(s) \\ &\quad + (sI - T)^{-1} \boldsymbol{t}^\circ (i-1) g^{i-2}(s) (-\boldsymbol{\alpha}) (sI - T)^{-2} \boldsymbol{t}^\circ]_{s=0} \\ &= -\sum_{i=1}^{N-1} \boldsymbol{\pi}_i (I_{(r)} \otimes \boldsymbol{e}_{(\omega)}) T^{-1} \boldsymbol{e} + \eta \sum_{i=2}^{N-1} (i-1) \boldsymbol{\pi}_i \boldsymbol{e} \, . \end{aligned} \tag{4.19}$$

The final equality used the fact that $\boldsymbol{t}^\circ = -T\boldsymbol{e}$ and $\eta = -\boldsymbol{\alpha} T^{-1} \boldsymbol{e}$. Since the time in the system equals the time in the queue plus service time, we have

$$E[\bar{W}] = E(W) + \eta. \tag{4.20}$$

For the virtual waiting time, we proceed in an analogous manner to the derivation of Eqs.(4.18) and (4.19):

$$\hat{W}^*(s) = \boldsymbol{y}_0\boldsymbol{e} + \sum_{i=1}^{N} \boldsymbol{y}_i(I_{(r)} \otimes \boldsymbol{e}_{(\omega)})P^L(i-1, s)\boldsymbol{t}^\circ, \tag{4.21}$$

$$E[\hat{W}] = -\sum_{i=1}^{N} \boldsymbol{y}_i(I_{(r)} \otimes \boldsymbol{e}_{(\omega)})T^{-1}\boldsymbol{e} + \eta\sum_{i=2}^{N}(i-1)\boldsymbol{y}_i\boldsymbol{e}\ . \tag{4.22}$$

B. Computational Considerations

The computational effort presented by these equations can be quite extensive; however, we shall show that if the service distribution has a squared coefficient of variation (SCV) greater than or equal to 1/2 there may exist closed form expressions for which the computational effort is greatly reduced. It is common to use a Cox-2 distribution to approximate a server whose $SCV \geq 1/2$.

A function $f(t)$ for $t \in \mathcal{R}^+$ is said to have a generalized exponential form if there exist a real value a_0, an integer k and for each $i = 1, \cdots, k$ there exist real values a_i and u_i and a positive real value v_i such that it is possible to write the function as

$$f(t) = a_0 + \sum_{i=1}^{k} a_i\, t^{u_i} e^{-v_i\, t}, \tag{4.23}$$

Each component of the time-dependent probability matrix (Eq. 4.3) and each term of the matrix $e^{\overline{T}\,t}$ (Eq. 4.4) for the PH renewal process associated with a Cox-2 distribution has a generalized exponential form with its constant term being zero. It also turns out that many practical semi-Markov kernels are made up of components having a generalized exponential form. In particular, the output process from a

$\lambda(n)$/PH/1/N queueing system forms a semi-Markov process with each component of its kernel having this generalized exponential form. (The reason that we are interested in the output process is that the interest in the SM/PH/1/N arises from an interest in modeling a tandem queueing system.) The advantage of having the generalized exponential form is that the function form for the integrals given by Eqs. (4.7), (4.8), and (4.14) – (4.17) will also have a generalized exponential form and the constant term equals zero so that the limit as t approaches infinity is zero. For example, suppose that each component of Q and P_N have a generalized exponential form and we wish to evaluate the i, j component of $\tilde{A}_k$ for a fixed k. Further suppose that F is a function such that $F(dt) = (P(k,t) \otimes dQ(t))_{ij}$ for $t \in \mathcal{R}^+$, then F is of the generalized exponential form and we have $\tilde{A}_k = F(\infty) - F(0) = -F(0)$.

With the availability of relatively powerful symbolic logic software, it may not be necessary to explicitly give solutions to the above equations. In particular, if the T matrix for the service process is not too complex and the integer n is not large, then the symbolic logic software *Maple* would be capable of obtaining expressions for $P(n,t)$, $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$ by performing the inverse transform from Eq. (4.5). (Not too complex means that, for example when representing a Coxian distribution, T is upper triangular with the diagonal elements equal or unequal with no two elements almost equal.) If symbolic software is to be used, the following steps summarize the use of the previously developed equations when the service process is Coxian and the kernel for the arrival process has the generalized exponential form:

Initialization. Obtain $P(k,t)$, for $k = 0, \cdots, N-1$ and $t \in \mathcal{R}^+$, by taking the inverse Laplace transformation of the right-hand side of (4.5).

Step 1. Compute $\tilde{P}_{(N)}$ defined by (4.9), where $\tilde{A}_k$ and $\tilde{B}_k$, for $k = 0, \cdots, N-1$ are the limiting matrices as $t \to \infty$ of the matrices defined by (4.7) and (4.8). To evaluate the integral in (4.7), only the lower limit needs to be calculated since the integrals evaluated at the upper limit are zero. If the service process further conforms to a Cox-2 distribution, the calculation is also true for (4.8) where the matrix $e^{\overline{T}x}$ also has a generalized exponential form.

Step 2. Compute $\boldsymbol{\pi}$ defined by (4.10).

Step 3. Compute $\boldsymbol{y}_0$ using (4.16) and $\boldsymbol{y}_k$ using (4.17) for $k = 1, \cdots, N$. Again, for evaluating the integrals in (4.17), only the lower limit needs to be calculated since the integrals evaluated at the upper limit are zero, and this is also true for (4.16) for the Cox-2 service process.

Step 4. The waiting time distributions $W(\cdot)$ and $\hat{W}(\cdot)$ are obtained by taking the inverse Laplace-Stieltjes transform of the right-hand sides of (4.18) and (4.21), respectively.

Step 5. Compute mean waiting times $E(W)$ and $E(\hat{W})$ using (4.19) and (4.22), respectively.

The difficulty with using some symbolic logic software packages such as *Maple* is that numerical accuracy can quickly become problematic due to round-off errors, for example, when T of the Coxian distribution has different but almost equal diagonal elements. If symbolic logic software is not used, it may be convenient to obtain closed form solutions to the above equations. In other words, we obtain solutions that do not involve the demanding computation for the integrals and thus improve accuracy but require less time in computation. This is possible when a Cox-2 distribution is used for the service time, and the kernel of the arrival process has a closed form solution to its Laplace-Stieltjes transform (note that this accommodates not only the case of generalized exponential kernels but also other cases even if the kernel itself may not have a closed form expression). We shall also assume that the derivatives of the transform are known. In the following derivation, we need the factorial derivative and, for ease of notation, we shall modify the analogous definition (2.5) in Chapter II as

$$\overline{Q}^{*(n)}(s_0) = \frac{(-1)^n}{n!}\frac{d^n}{ds^n}Q^*(s)|_{s=s_0} \text{ for } n \in \mathcal{N}, \tag{4.24}$$

where $\overline{Q}^{*(0)} = Q^*$. (Also, remember that for fixed s and n, $\overline{Q}^{*(n)}(s)$ is a $r \times r$ matrix.)

The representation for the Cox-2 service time distribution is given by $\boldsymbol{\alpha} = (1, 0)$ and

$$T = \begin{pmatrix} -\mu_1 & \mu_1 p \\ 0 & -\mu_2 \end{pmatrix}, \tag{4.25}$$

where $\mu_1 \neq \mu_2$. By Chapter III the time dependent probabilities are given, for $t \in \mathcal{R}^+$, by

$$P(0,t) = \begin{pmatrix} e^{-\mu_1 t} & \frac{\mu_1 p}{\mu_1 - \mu_2}(-e^{-\mu_1 t} + e^{-\mu_2 t}) \\ 0 & e^{-\mu_2 t} \end{pmatrix}, \text{ and} \tag{4.26}$$

$$P(k,t) = \tag{4.27}$$

$$\sum_{h=1}^{2} \begin{pmatrix} \mu_1^k \sum_{j=1}^{a(1,1,k,h)} b(1,1,k,h,j) & \mu_1^{k+1} p \sum_{j=1}^{a(1,2,k,h)} b(1,2,k,h,j) \\ \mu_1^{k-1}\mu_2 \sum_{j=1}^{a(2,1,k,h)} b(2,1,k,h,j) & \mu_1^k \mu_2 p \sum_{j=1}^{a(2,2,k,h)} b(2,2,k,h,j) \end{pmatrix} t^{j-1} e^{-\mu_h t}$$

for $k = 1, \cdots, N-1$. The parameters $a(m_1, m_2, k, h)$ and $b(m_1, m_2, k, h, j)$ are determined by Eqs. (3.18) and (3.19), respectively.

To compute the probability vector $\boldsymbol{\pi}$ (Eq. 4.10), we use the expressions given in (4.26) and (4.27) to obtain expressions for $\tilde{A}_k$ and $\tilde{B}_k$. The matrices $\tilde{A}_k$ are given by

$$\tilde{A}_0 = \begin{pmatrix} Q^*(\mu_1) & \frac{\mu_1 p}{\mu_1 - \mu_2}\Big(-Q^*(\mu_1) + Q^*(\mu_2)\Big) \\ 0 & Q^*(\mu_2) \end{pmatrix}, \text{ and} \tag{4.28}$$

$$\tilde{A}_k = \tag{4.29}$$
$$\sum_{h=1}^{2} \begin{pmatrix} \mu_1^k \sum_{j=1}^{a(1,1,k,h)} b(1,1,k,h,j) & \mu_1^{k+1} p \sum_{j=1}^{a(1,2,k,h)} b(1,2,k,h,j) \\ \mu_1^{k-1}\mu_2 \sum_{j=1}^{a(2,1,k,h)} b(2,1,k,h,j) & \mu_1^k \mu_2 p \sum_{j=1}^{a(2,2,k,h)} b(2,2,k,h,j) \end{pmatrix} \otimes \overline{Q}^{*(j-1)}(\mu_h)$$

for $k = 1, \cdots, N-1$.

To obtain expressions for $\tilde{B}_k$, we must first obtain an expression for the integral $\int_0^\infty e^{\overline{T}x} \otimes dQ(x)$. Noting that

$$\overline{T} = \begin{pmatrix} -\mu_1 p & \mu_1 p \\ \mu_2 & -\mu_2 \end{pmatrix},$$

and

$$e^{\overline{T}x} = \frac{1}{\mu_1 p + \mu_2} \begin{pmatrix} \mu_2 + \mu_1 p\, e^{-(\mu_1 p + \mu_2)x} & \mu_1 p(1 - e^{-(\mu_1 p + \mu_2)x}) \\ \mu_2(1 - e^{-(\mu_1 p + \mu_2)x}) & \mu_1 p + \mu_2 e^{-(\mu_1 p + \mu_2)x} \end{pmatrix}$$

for $x \in \mathcal{R}^+$, we obtain,

$$\int_0^\infty e^{\overline{T}x} \otimes dQ(x) = \tag{4.30}$$
$$\frac{1}{\mu_1 p + \mu_2} \begin{pmatrix} \mu_2 Q^*(0) + \mu_1 p\, Q^*(\mu_1 p + \mu_2) & \mu_1 p\Big(Q^*(0) - Q^*(\mu_1 p + \mu_2)\Big) \\ \mu_2\Big(Q^*(0) - Q^*(\mu_1 p + \mu_2)\Big) & \mu_1 p\, Q^*(0) + \mu_2 Q^*(\mu_1 p + \mu_2) \end{pmatrix}.$$

Eqs. (4.28) – (4.30) can be used together to determine the $\tilde{B}_k$ matrices through the formula

$$\tilde{B}_k = \left(\int_0^\infty e^{\overline{T}x} \otimes dQ(x) - \sum_{n=0}^{k} \tilde{A}_n \right) \Big((\boldsymbol{e}_{(r)} \cdot \boldsymbol{\alpha}) \otimes I_{(\omega)} \Big) \quad \text{for } k = 0, \cdots, N-1, \tag{4.31}$$

where for the Cox-2 case

$$\Big((\boldsymbol{e}_{(r)} \cdot \boldsymbol{\alpha}) \otimes I_{(\omega)} \Big) = \begin{pmatrix} I_{(\omega)} & \mathbf{0} \\ I_{(\omega)} & \mathbf{0} \end{pmatrix}.$$

This completes the computations for $\tilde{P}_N$ and thus for $\boldsymbol{\pi}$.

The computation of the probability vector $\boldsymbol{y}$ using (4.16) and (4.17) will require a function of the form $(\tilde{Q} - Q^*(s))/s$ for $s \in \mathcal{R}^+$ and it factorial derivatives. For this purpose, we define the matrix functions G and $\overline{G}^{(n)}$, for $n \in \mathcal{N}$ and $s \in \mathcal{R}^+$, and note a nice property of G analogous to Eq. (2.28) as follows

$$\begin{aligned} G(s) &= \frac{\tilde{Q} - Q^*(s)}{s} \,, \\ \overline{G}^{(n)}(s_0) &= \frac{(-1)^n}{n!} \frac{d^n}{ds^n} G(s)|_{s=s_0} \,, \quad \text{and} \\ G(0) \;\equiv\; \lim_{s\to 0} G(s) &= -\lim_{s\to 0} \frac{d}{ds}\, Q^*(s) \;\equiv\; \overline{Q}^{*(1)}(0) \,. \end{aligned} \tag{4.32}$$

The last equality in (4.32) is seen by observing that $\tilde{Q} = Q^*(0)$. Thus in the matrices below, the modeler may want to replace $G(0)$ with $\overline{Q}^{*(1)}(0)$. The key integrals involved in Eqs. (4.16) and (4.17) can be shown to be as follows:

$$\int_0^\infty P(0,t) \otimes \Big(\tilde{Q} - Q(t)\Big)\, dt = \begin{pmatrix} G(\mu_1) & \dfrac{\mu_1 p}{\mu_1 - \mu_2}\Big(-G(\mu_1) + G(\mu_2)\Big) \\ 0 & G(\mu_2) \end{pmatrix}, \tag{4.33}$$

$$\int_0^\infty P(k,t) \otimes \Big(\tilde{Q} - Q(t)\Big)\, dt = \tag{4.34}$$
$$\sum_{h=1}^{2} \begin{pmatrix} \mu_1^k \displaystyle\sum_{j=1}^{a(1,1,k,h)} b(1,1,k,h,j) & \mu_1^{k+1} p \displaystyle\sum_{j=1}^{a(1,2,k,h)} b(1,2,k,h,j) \\ \mu_1^{k-1}\mu_2 \displaystyle\sum_{j=1}^{a(2,1,k,h)} b(2,1,k,h,j) & \mu_1^k \mu_2 p \displaystyle\sum_{j=1}^{a(2,2,k,h)} b(2,2,k,h,j) \end{pmatrix} \otimes \overline{G}^{(j-1)}(\mu_h) \,,$$

for $k = 1, \cdots, N-1$, and

$$\int_0^\infty e^{\overline{T}x} \otimes \Big(\tilde{Q} - Q(x)\Big)\, dx = \tag{4.35}$$
$$\frac{1}{\mu_1 p + \mu_2} \begin{pmatrix} \mu_2 G(0) + \mu_1 p G(\mu_1 p + \mu_2) & \mu_1 p\Big(G(0) - G(\mu_1 p + \mu_2)\Big) \\ \mu_2\Big(G(0) - G(\mu_1 p + \mu_2)\Big) & \mu_1 p G(0) + \mu_2 G(\mu_1 p + \mu_2) \end{pmatrix}.$$

To obtain an expression for the waiting time distributions, we will let $\mathcal{L}^{-1}$ denote the inverse Laplace transform; in other words, $\mathcal{L}^{-1}(W^L(\cdot)) = W(\cdot)$. Recall that the relationship between the Laplace transform and the Laplace-Stieltjes transform is given as $W^L(s) = W^*(s)/s$ for $s \in \mathcal{R}^+$. The distribution functions for the waiting time distribution and the virtual waiting time distribution are

$$W(t) = \boldsymbol{\pi}_0 \boldsymbol{e} + \sum_{k=1}^{N-1} \boldsymbol{\pi}_k \left(I_{(r)} \otimes \boldsymbol{e}_{(\omega)}\right) \mathcal{L}^{-1}\left(\frac{P^L(k-1,s)}{s}\right) \boldsymbol{t}^\circ \,, \text{ and} \tag{4.36}$$

$$\hat{W}(t) = \boldsymbol{y}_0 \boldsymbol{e} + \sum_{k=1}^{N-1} \boldsymbol{y}_k \left(I_{(r)} \otimes \boldsymbol{e}_{(\omega)}\right) \mathcal{L}^{-1}\left(\frac{P^L(k-1,s)}{s}\right) \boldsymbol{t}^\circ \,. \tag{4.37}$$

Noting that $\mathcal{L}^{-1}(P^L(k,s)/s) = \int_0^t P(k,x)dx$ for $t \in \mathcal{R}^+$. By substituting (4.26) we have,

$$\mathcal{L}^{-1}\left(\frac{P^L(0,s)}{s}\right) = \begin{pmatrix} \dfrac{1-e^{-\mu_1 t}}{\mu_1} & \dfrac{p(\mu_1 - \mu_2 + \mu_2 e^{-\mu_1 t} - \mu_1 e^{-\mu_2 t})}{\mu_2(\mu_1 - \mu_2)} \\ 0 & \dfrac{1-e^{-\mu_2 t}}{\mu_2} \end{pmatrix}. \tag{4.38}$$

For $k \geq 1$, substituting (4.27) yields

$$\mathcal{L}^{-1}\left(\frac{P^L(k,s)}{s}\right) = \tag{4.39}$$
$$\sum_{h=1}^{2} \begin{pmatrix} \mu_1^k \displaystyle\sum_{j=1}^{a(1,1,k,h)} b(1,1,k,h,j) & \mu_1^{k+1} p \displaystyle\sum_{j=1}^{a(1,2,k,h)} b(1,2,k,h,j) \\ \mu_1^{k-1}\mu_2 \displaystyle\sum_{j=1}^{a(2,1,k,h)} b(2,1,k,h,j) & \mu_1^k \mu_2 p \displaystyle\sum_{j=1}^{a(2,2,k,h)} b(2,2,k,h,j) \end{pmatrix} \int_0^t x^{j-1} e^{-\mu_h x} dx \,,$$

where

$$\int_0^t x^j e^{-\mu_h x} dx = \frac{\prod_{v=0}^{j-1}(j-v)}{\mu_h^{j+1}} - \frac{1}{\mu_h}\left(t^j + \sum_{u=1}^{j} \frac{\prod_{v=0}^{u-1}(j-v)}{\mu_h^u} t^{j-u}\right) e^{-\mu_h t}. \tag{4.40}$$

Thus, if the environmental process is such that its semi-Markov kernel's Laplace-Stieltjes transform can be obtained in closed form, the following procedure can be used to describe the SM/C_2/1/N queueing system where the two phases that describe the service process have unequal means.

Step 1. Compute the matrices $\tilde{A}_k$ for $k = 0, \cdots, N-1$ using (4.28) and (4.29), where Q^* is the Laplace-Stieltjes transform of the arrival process kernel and $\overline{Q}^{*(j)}$ is defined by (4.24). Compute $\int_0^\infty e^{\overline{T}x} \otimes dQ(x)$ from (4.30). Finally, compute $\tilde{B}_k$, for $k = 0, \cdots, N-1$ using (4.31).

Step 2. Build the matrix $\tilde{P}_N$ (4.9) and compute $\boldsymbol{\pi}$ defined by (4.10).

Step 3. Compute $\boldsymbol{y}_0$ using (4.16) and $\boldsymbol{y}_k$ using (4.17) for $k = 1, \cdots, N$, where the integrals in those equations are the matrices of the right-hand-side of Eqs. (4.33), (4.34), and (4.35).

Step 4. The waiting time distributions $W(\cdot)$ and $\hat{W}(\cdot)$ are obtained from Eqs. (4.36) and (4.37), respectively, where the inverse transform is given by (4.38) and (4.39).

Step 5. The mean waiting time and mean virtual waiting time are given by Eqs. (4.19) and (4.22), respectively.

The inverse of the T matrix (4.25) that is associated with the service distribution takes a different form when $\mu_1 = \mu_2 = \mu$. The following contains the analogous quantities to the above that can be used when the $SCV = 1/2$ which will force $\mu_1 = \mu_2 = \mu$:

$$P(0,t) = \begin{pmatrix} e^{-\mu t} & \mu\, t\, e^{-\mu t} \\ 0 & e^{-\mu t} \end{pmatrix}, \tag{4.41}$$

$$P(k,t) = \begin{pmatrix} \mu^k \sum_{j=0}^{k} c(1,k,j) t^{2k-j} & \mu^{k+1} p \sum_{j=0}^{k} c(2,k,j) t^{2k-j+1} \\ \mu^k \sum_{j=0}^{k-1} c(2,k-1,j) t^{2k-j+1} & \mu^{k+1} p \sum_{j=0}^{k-1} c(3,k-1,j) t^{2k-j+2} \end{pmatrix} e^{-\mu t}, \tag{4.42}$$

for $k = 1, \cdots, N-1$ and where the parameter $c(m,k,j)$ is given by (3.33).

Continuing as before,

$$\tilde{A}_0 = \begin{pmatrix} Q^*(\mu) & \mu\overline{Q}^{*(1)}(\mu) \\ 0 & Q^*(\mu) \end{pmatrix} \tag{4.43}$$

$$\tilde{A}_k = \tag{4.44}$$

$$\begin{pmatrix} \mu^k \sum_{j=0}^{k} c(1,k,j)\overline{Q}^{*(2k-j)}(\mu) & \mu^{k+1} p \sum_{j=0}^{k} c(2,k,j)\overline{Q}^{*(2k-j+1)}(\mu) \\ \mu^k \sum_{j=0}^{k-1} c(2,k-1,j)\overline{Q}^{*(2k-j+1)}(\mu) & \mu^{k+1} p \sum_{j=0}^{k-1} c(3,k-1,j)\overline{Q}^{*(2k-j+2)}(\mu) \end{pmatrix},$$

for $k = 1, \cdots, N-1$ and

$$\int_0^\infty e^{\overline{T}x} \otimes dQ(x) = \frac{1}{2} \begin{pmatrix} Q^*(0) + Q^*(2\mu) & Q^*(0) - Q^*(2\mu) \\ Q^*(0) - Q^*(2\mu) & Q^*(0) + Q^*(2\mu) \end{pmatrix} \tag{4.45}$$

$$\int_0^\infty P(0,x) \otimes [\tilde{Q} - Q(x)]dx = \begin{pmatrix} G(\mu) & \mu\overline{G}^{(1)}(\mu) \\ 0 & G(\mu) \end{pmatrix}, \tag{4.46}$$

$$\int_0^\infty P(k,t) \otimes [\tilde{Q} - Q(t)]dt = \tag{4.47}$$

$$\begin{pmatrix} \mu^k \sum_{j=0}^{k} c(1,k,j)\overline{G}^{(2k-j)}(\mu) & \mu^{k+1} p \sum_{j=0}^{k} c(2,k,j)\overline{G}^{(2k-j+1)}(\mu) \\ \mu^k \sum_{j=0}^{k-1} c(2,k-1,j)\overline{G}^{(2k-j+1)}(\mu) & \mu^{k+1} p \sum_{j=0}^{k-1} c(3,k-1,j)\overline{G}^{(2k-j+2)}(\mu) \end{pmatrix},$$

for $k = 1, \cdots, N-1$ and finally

$$\int_0^\infty e^{\overline{T}x} \otimes [\tilde{Q} - Q(x)]dx = \frac{1}{2} \begin{pmatrix} G(0) + G(2\mu) & G(0) - G(2\mu) \\ G(0) - G(2\mu) & G(0) + G(2\mu) \end{pmatrix}. \tag{4.48}$$

To compute the waiting-time distributions we again need the inverse Laplace transform which yields, for $\mu_1 = \mu_2 = \mu$,

$$\mathcal{L}^{-1}\left(\frac{P^L(0,s)}{s}\right) = \begin{pmatrix} \dfrac{1-e^{-\mu t}}{\mu} & \dfrac{p(1-e^{-\mu t} - \mu\, t\, e^{-\mu t})}{\mu} \\ 0 & \dfrac{1-e^{-\mu t}}{\mu} \end{pmatrix}, \tag{4.49}$$

and for $k = 1, \cdots, N-1$,

$$\mathcal{L}^{-1}\left(\frac{P^L(k,s)}{s}\right) = \tag{4.50}$$

$$\begin{pmatrix} \mu^k \sum_{j=0}^{k} c(1,k,j) \int_0^t x^{2k-j} e^{-\mu x} dx & \mu^{k+1} p \sum_{j=0}^{k} c(2,k,j) \int_0^t x^{2k-j+1} e^{-\mu x} dx \\ \mu^k \sum_{j=0}^{k-1} c(2,k-1,j) \int_0^t x^{2k-j+1} e^{-\mu x} dx & \mu^{k+1} p \sum_{j=0}^{k-1} c(3,k-1,j) \int_0^t x^{2k-j+2} e^{-\mu x} dx \end{pmatrix}.$$

If the environmental process is such that its semi-Markov kernel's Laplace-Stieltjes transform can be obtained in closed form, the following procedure can be used to describe the SM/C_2/1/N queueing system where the two phases that describe the service process have equal means.

Step 1. Compute the matrices $\tilde{A}_k$ for $k = 0, \cdots, N-1$ using (4.43) and (4.44), where Q^* is the Laplace-Stieltjes transform of the arrival process kernel and $\overline{Q}^{*(j)}$ is defined by (4.24). Compute $\int_0^\infty e^{\overline{T}x} \otimes dQ(x)$ from (4.45). Finally, compute $\tilde{B}_k$, for $k = 0, \cdots, N-1$ using (4.31).

Step 2. Build the matrix $\tilde{P}_N$ (4.9) and compute $\boldsymbol{\pi}$ defined by (4.10).

Step 3. Compute $\boldsymbol{y}_0$ using (4.16) and $\boldsymbol{y}_k$ using (4.17) for $k = 1, \cdots, N$, where the integrals in those equations are the matrices of the right-hand-side of Eqs. (4.46), (4.47), and (4.48). (Remember that the matrices $\overline{G}^{(j)}$ are defined by (4.32).)

Step 4. The waiting time distributions $W(\cdot)$ and $\hat{W}(\cdot)$ are obtained from Eqs. (4.36) and (4.37), respectively, where the inverse transform is given by (4.49) and (4.50).

Step 5. The mean waiting time and mean virtual waiting time are given by Eqs. (4.19) and (4.22), respectively.

C. Two Queues in Series

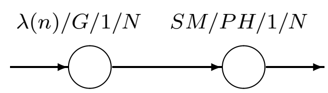

Fig. 2. Two Queues in Series

We implement the proposed method for a queueing system consisting of two queues in series (Figure 2). For the queue of the first stage, denoted by Q_1, the interarrival time distribution is exponential and load-dependent. That is, there exists an integer $N_1 \geq 2$ such that $\lambda(n_1) > 0$ for $0 \leq n_1 \leq N_1 - 1$, and $\lambda(n_1) = 0$ for all $n_1 \geq N_1$, where n_1 denotes the number of customers in Q_1. The service time distribution φ is arbitrary and its Laplace-Stieltjes transform φ^* is given in closed-form. Therefore, Q_1 is of the $\lambda(n)$/G/1/N type addressed in Chapter II. Let $T_0 = 0, T_1, T_2, \cdots$ be the instants of successive departures, and denote by X_i the number of customers left behind by the i^{th} departure. Then the stochastic process $(X, T) = \{X_i, T_i; i \in \mathcal{N}\}$ is a Markov renewal process with state space $\{0, 1, \cdots, N_1 - 1\}$. The process is assumed time-homogeneous, i.e., for $0 \leq k, j \leq N_1 - 1$ and $t \in \mathcal{R}^+$,

$$Q_{kj}(t) = P\{X_{i+1} = j, T_{i+1} - T_i \leq t | X_i = k\}, \tag{4.51}$$

which is independent of i. We obtain in Chapter II that, for $t \in \mathcal{R}^+$,

$$Q(t) = \begin{pmatrix} p_0(t) & p_1(t) & p_2(t) & \cdots & p_{N_1-1}(t) \\ q_{0,1}(t) & q_{1,1}(t) & q_{2,1}(t) & \cdots & q_{N_1-1,1}(t) \\ & q_{0,2}(t) & q_{1,2}(t) & \cdots & q_{N_1-2,2}(t) \\ & & \ddots & \vdots & \vdots \\ \mathbf{0} & & & q_{0,N_1-1}(t) & q_{1,N_1-1}(t) \end{pmatrix}, \tag{4.52}$$

where

$$q_{m,n}(t) = \int_0^t \varphi(dx)\psi_{m,n}(x) \text{ for } 1 \le n \le N_1 - 1 \text{ and } 1 \le m + n \le N_1, \tag{4.53}$$

$$p_m(t) = \int_0^t \lambda(0)e^{-\lambda(0)(t-x)} q_{m,1}(x)\, dx \text{ for } 0 \le m \le N_1 - 1. \tag{4.54}$$

$\psi_{m,n}(t)$ denotes the probability that there are exactly m arrivals in $(0,t]$, given the initial population n. The derivation and computation for $\psi_{m,n}(t)$ are contained in Chapter II. By (4.53) and substituting $\psi_{m,n}(t)$ by the right-hand side of Eq. (2.14), then taking the Laplace transformation yields

$$q^*_{m,n}(s) = \tag{4.55}$$
$$\begin{cases} \displaystyle\sum_{i=1}^{u(m+1,n)} \sum_{v=1}^{k_i(m+1,n)} c_{i,v}(m,n)\overline{\varphi}^{*(v-1)}(s+d_i), & \text{for } m \ge 1, 1 \le n \le N_1 - m; \\ \varphi^*(s+\lambda(n)), & \text{for } m = 0, 1 \le n \le N_1 - 1; \\ 0, & \text{otherwise,} \end{cases}$$

where $c_{i,v}(m,n)$ is the parameter given by (2.15). By (4.54) we have

$$p^*_m(s) = \frac{\lambda(0)q^*_{m,1}(s)}{s+\lambda(0)} \text{ for } 0 \le m \le N_1 - 1. \tag{4.56}$$

Note that Q is at the same time the semi-Markov kernel for the arrival process of the queue Q_2 of the second stage with finite buffer. If Q_2 reaches full capacity, the customer that completes service in Q_1 is released from the whole system, instead of remaining inside and blocking the system. Upon obtaining the kernel Q or its Laplace-Stieltjes transformation Q^*, and assuming Q_2 has the PH service distribution, we solve Q_2 as the SM/PH/1/N queue and obtain its steady-state system size probabilities and waiting time distributions by the method proposed in the preceding section.

D. Numerical Examples

We compare system size probabilities generated by the proposed method and simulation. In addition, we obtain waiting-time distribution functions and associated mean values. For all the simulation outcomes given in this section, we have 95% confidence interval with the half-width of the interval being less than 2% of the mean estimate. Note in the following tables $y_{0,1,\ell}$ is used to represent $y_{0,\ell}$ for ease of notation.

The SM/PH/1/N queue is degenerated as the GI/PH/1/N by letting $\tilde{Q} = 1$ and the kernel Q represents the interarrival time distribution. In Table V and the following, we solved again the example of the M/C_2/1/N queue shown in Chapter II by using the algorithm presented in this chapter.

$$\begin{aligned} W(t) = 1.00000 + (&-0.05436\,t^9 - 0.01268\,t^8 + 0.04729\,t^7 + 0.10217\,t^6 + 0.12619\,t^5 \\ &+0.10999\,t^4 + 0.06836\,t^3 + 0.02702\,t^2 + 0.00235\,t - 0.00407)e^{-7\,t} \\ &+(-0.00000\,t^9 - 0.00000\,t^8 - 0.00000\,t^7 - 0.00000\,t^6 - 0.00003\,t^5 \\ &-0.00083\,t^4 - 0.01175\,t^3 - 0.09745\,t^2 - 0.45312\,t - 0.96566)e^{-0.5\,t}, \end{aligned}$$

$$\begin{aligned} \hat{W}(t) = 1.00000 + (&-0.00542\,t^{10} - 0.00709\,t^9 - 0.00066\,t^8 + 0.01434\,t^7 + 0.03190\,t^6 \\ &+0.04222\,t^5 + 0.03907\,t^4 + 0.02505\,t^3 + 0.00875\,t^2 - 0.00236\,t - 0.00461)e^{-7\,t} \\ &+(-0.00000\,t^{10} - 0.00000\,t^9 - 0.00000\,t^8 - 0.00000\,t^7 - 0.00000\,t^6 - 0.00005\,t^5 \end{aligned}$$

Table V. Steady-state probabilities at arrivals and at an arbitrary point in time for $Q(t) = 1 - e^{-2t}$, $\mu_1 = 7$, $\mu_2 = 0.5$, $p = 1/3$, and $N = 11$

	Our Method			Simulation
i	$\pi_{i,1,1}$	$\pi_{i,2,1}$	$\pi(i)$	$\pi(i)$
0	0.03027	0	0.03027	0.03025
1	0.01179	0.01101	0.02280	0.02279
2	0.00803	0.01630	0.02432	0.02427
3	0.00821	0.02070	0.02891	0.02889
4	0.00965	0.02556	0.03521	0.03519
5	0.01173	0.03140	0.04312	0.04307
6	0.01436	0.03852	0.05287	0.05286
7	0.01760	0.04724	0.06484	0.06487
8	0.02158	0.05793	0.07951	0.07950
9	0.02646	0.07104	0.09751	0.09742
10	0.04173	0.47890	0.52063	0.52331
	$y_{i,1,1}$	$y_{i,2,1}$	$y(i)$	$y(i)$
0	0.03027	0	0.03027	0.03021
1	0.01179	0.01101	0.02280	0.02274
2	0.00803	0.01630	0.02432	0.02430
3	0.00821	0.02079	0.02891	0.02886
4	0.00965	0.02556	0.03521	0.03519
5	0.01173	0.03140	0.04312	0.04310
6	0.01436	0.03852	0.05287	0.05284
7	0.01760	0.04724	0.06484	0.06488
8	0.02158	0.05793	0.07951	0.07962
9	0.02646	0.07104	0.09751	0.09752
10	0.03245	0.08713	0.11958	0.11961
11	0.00927	0.39178	0.40105	0.40113

$$-0.00096\,t^4 - 0.01241\,t^3 - 0.09838\,t^2 - 0.45193\,t - 0.96512)e^{-0.5\,t},$$

$$E(W) = 7.43151, \qquad E(\hat{W}) = 7.75617.$$

We consider in the second example two queues in series. For the queue Q_1, the interarrival time distribution is exponential with load-dependent rates $\lambda(0) = 3$, $\lambda(1) = 2$, $\lambda(2) = 1$, and $\lambda(n) = 0$ for all $n \geq 3$. The service time distribution is given by $\varphi(t) = 1 - e^{-2\,t}$. By (4.52) we obtain

$$Q(t) =$$
$$\begin{pmatrix} \frac{3}{2}e^{-4\,t} - 2\,e^{-3\,t} + \frac{1}{2} & -3\,e^{-4\,t} + (-4\,t + \frac{8}{3})e^{-3\,t} + \frac{1}{3} & \frac{3}{2}e^{-4\,t} + (4\,t + \frac{4}{3})e^{-3\,t} - 3\,e^{-2\,t} + \frac{1}{6} \\ -\frac{1}{2}e^{-4\,t} + \frac{1}{2} & e^{-4\,t} - \frac{4}{3}e^{-3\,t} + \frac{1}{3} & -\frac{1}{2}e^{-4\,t} + \frac{4}{3}e^{-3\,t} - e^{-2\,t} + \frac{1}{6} \\ 0 & -\frac{2}{3}e^{-3\,t} + \frac{2}{3} & \frac{2}{3}e^{-3\,t} - e^{-2\,t} + \frac{1}{3} \end{pmatrix}.$$

The performance characteristics computed for Q_2 are shown in Table VI and the following:

$$W(t) = 1.00000 + (-0.65399\,t^3 - 22.05811\,t^2 - 275.29775\,t - 1283.16519)e^{-1.5\,t}$$
$$+(-2.85429\,t^3 + 44.31480\,t^2 - 367.27736\,t + 1282.16776)e^{-t},$$

$$\hat{W}(t) =$$
$$1.00000 + (0.17630\,t^4 + 9.79802\,t^3 + 223.12824\,t^2 + 2479.27850\,t + 11332.24652)e^{-1.5\,t}$$
$$+(-1.06550\,t^4 + 27.78475\,t^3 - 400.51150\,t^2 + 3185.85093\,t - 11333.24441)e^{-t},$$

$$E(W) = 5.25033, \qquad E(\hat{W}) = 6.15351.$$

Table VI. Steady-state probabilities at arrivals and at an arbitrary point in time for Q_2 with $\mu_1 = 1$, $\mu_2 = 1.5$, $p = 2/3$, and $N_2 = 5$

	Our Method							Simulation
i	$\pi_{i,1,1}$	$\pi_{i,1,2}$	$\pi_{i,1,3}$	$\pi_{i,2,1}$	$\pi_{i,2,2}$	$\pi_{i,2,3}$	$\pi(i)$	$\pi(i)$
0	0.00024	0.00078	0.00155	0	0	0	0.00257	0.00259
1	0.00125	0.00312	0.00340	0.00019	0.00054	0.00081	0.00932	0.00931
2	0.00528	0.01009	0.00983	0.00106	0.00233	0.00254	0.03113	0.03110
3	0.02085	0.03307	0.02582	0.00443	0.00800	0.00720	0.09938	0.09936
4	0.24945	0.23014	0.09816	0.11724	0.11193	0.05068	0.85760	0.85763
	$y_{i,1,1}$	$y_{i,1,2}$	$y_{i,1,3}$	$y_{i,2,1}$	$y_{i,2,2}$	$y_{i,2,3}$	$y(i)$	$y(i)$
0	0.00013	0.00049	0.00149	0	0	0	0.00211	0.00212
1	0.00067	0.00209	0.00373	0.00010	0.00036	0.00081	0.00775	0.00781
2	0.00295	0.00728	0.01140	0.00058	0.00161	0.00281	0.02663	0.02661
3	0.01216	0.02551	0.03314	0.00252	0.00589	0.00863	0.08784	0.08780
4	0.05126	0.08555	0.08540	0.01057	0.02051	0.02414	0.27743	0.27742
5	0.11511	0.14649	0.10812	0.06711	0.09019	0.07121	0.59824	0.59824

E. Conclusion

An explicit algorithm that generates numerical solutions for the steady-state system size probabilities and waiting time distribution functions of the SM/PH/1/N queue is developed. The computation is proportional to $(\omega^2 r^2 N^2)$. Moments of the waiting time can also be easily obtained. In addition, we demonstrate an efficient technique that implements the SM/PH/1/N model for the queue of the second stage of a sequential two-queue system without having to calculate performance characteristics of the first queue.

CHAPTER V

APPROXIMATE ANALYSIS FOR CLOSED QUEUEING NETWORKS

The queueing network under consideration is of single class, and its internal buffers are assumed sufficiently large that no blocking occurs at any time. There is only one server at each station. The service implements a first-come first-served (FCFS) discipline. The distribution of service times at each station is independent and can be any probability distribution as long as it has a closed-form Laplace transform and finite mean value. The routing probabilities are pre-specified and thus independent of the state of the network.

A. Norton's Theorem and Closed Network Approximations

Chandy et al. [10] demonstrated how to apply Norton's theorem in electrical circuit theory to queueing networks. According to their analysis, a closed network with K nodes and N customers can be substituted with a two-node cyclic system that consists of any node k, $k = 1, \cdots, K$, of the original network and an aggregated single node with load-dependent service rates $\mu_{(k)}(n)$, for $n = 1, \cdots, N$. To illustrate, fix k to be some given node and denote by $\tau_k(n)$, for each $n = 1, \cdots, N$, the throughput of that node when the original network is modified so that it contains n customers and the service time of node k is set to zero (called a short-circuiting of node k by [7]). (The network that contains all nodes with the same topology as the original network except that node k has been short-circuited and all other nodes have exponential servers is called the *complementary subnetwork associated with node k*.) Now consider the two-node closed network with N customers where the first node is the same as node k from the original network and the second node has a state-dependent exponential server with rate given by $\mu_{(k)}(n) = \tau_k(n)$ when the second node contains n customers. This two-node network is called the *aggregated cyclic network* associated with node k. For queueing networks that satisfy local balance [3], including networks with exponential servers and a FCFS discipline at all queues, the substitution is exact. That is, the system size and waiting time distributions of queue k in the original network are the same as in its complementary subnetwork.

As a direct implementation of Norton's theorem, Chandy et al. [11] further presented an approximate method for the analysis of closed queueing networks with general server distributions. Their method is an iterative procedure that first determines the system size and throughput at each node using the node's aggregated cyclic network which can be analyzed by the methods in [12, 20]. It should be noted that the difficulty in using the aggregated cyclic network is in determining values for $\tau_k(n)$ for $k = 1, 2, \cdots, K$ and $n = 1, 2, \cdots, N$. However, for exponential servers, a mean value analysis [42] or the convolution algorithm [8] can be easily applied, so the first step in the iterative procedure is to replace the general service distributions with exponential servers having the same mean service times. Because exponential service times are used instead of the actual service distribution, it is unlikely that the sum of all system sizes will equal N and the throughput at each node will satisfy the network topology, so the procedure of [11] gives a method of adjusting the mean service times for the next iteration based on the current computed system sizes and throughputs at all nodes.

The iterative method proposed by Marie [28] is an extension of the preceding method and has been shown to "behave very well both qualitatively and quantitatively and is the method of choice" [15]. In Marie's method, each node is analyzed as an isolated $\lambda(n)/\mathrm{C}_r/1/N$ model. The load-dependent arrival rate at a specific node, call it node k, is determined by forming the complementary subnetwork associated with node k and calculating the throughput rate, $\tau_k(n)$, at that (short-circuited) node when there are n jobs in the network, where n varies from 1 to N. The load-dependent arrival rate to node k, for $k = 1, \cdots, K$, is thus given as

$$\lambda_k(n) = \tau_k(N - n) \text{ for } n = 0, \cdots, N - 1 , \quad (5.1)$$

in other words $\lambda_k(n)$ is the mean arrival rate to node k when there are n jobs at node k.

The main difference between Marie's method and the method of Chandy et al. is that Marie uses load-dependent (exponential) service rates for the complementary subnetworks instead of a constant exponential rate. Let $\nu_k(n)$ be the load-dependent departure rates at node k for $1 \leq k \leq K$ and $1 \leq n \leq N$. $\nu_k(n)$ can be obtained by analyzing the $\lambda(n)/\mathrm{C}_r/1/N$ queue using the algorithm in [29] where particularly efficient algorithms have been developed for Cox-2 and Erlang-r service processes. The load-dependent (exponential) service rates for each queue k of the complementary subnetwork are updated equal to $\nu_k(n)$ for $n = 1, 2, \cdots, N$. The procedures to compute $\lambda_k(n)$ and $\nu_k(n)$ repeat iteratively until stopping conditions are fulfilled. See Appendix B for the steps of Marie's method.

It is known that, in two-moment approximations, a Coxian representation can be used to fit an arbitrary distribution that has a rational Laplace transform with SCV greater than zero. For $0 < SCV < 1$, the number of phases of the Cox-r model is chosen such that

$$\frac{1}{r} \leq SCV < \frac{1}{r-1}.$$

(In addition, Cox-2 can be used when SCV is greater than or equal to 1/2.) Obviously, when the SCV is close to zero, this approximation requires a large number of phases and therefore, the method of phases may become unattractive or impractical in queueing applications. Nojo and Watanabe [41] proposed an alternative model that requires only two exponential phases to fit a distribution with arbitrary SCV ($0 \leq SCV < \infty$) by allowing negative value of routing probabilities between the two phases. However, when applied to the $\lambda(n)/\mathrm{C}_r/1/N$ queue, the two-phase model with negative probabilities may generate negative values for the steady-state queue length probabilities.

B. Analysis and Algorithm

To avoid the difficulty that arises when the Coxian model causes Marie's method to be intractable (i.e., when the number of phases is large) or when it is important to model higher moments, we propose a new approach for the approximate analysis of closed queueing networks. Our approach follows that of Chandy et al. except that we analyze each node in isolation as a $\lambda(n)/\mathrm{G}/1/N$ model addressed in Chapter II.

Using the results in Chapter II, the steady-state probabilities for the $\lambda(n)/\mathrm{G}/1/N$ system can be obtained. In other words, we obtain $y_k(n)$, for $k = 1, \cdots, K$ and $n = 0, \cdots, N$, as the probability that n jobs are at node k, where the node is analyzed in isolation after the state-dependent arrival rates are determined. Based on these probabilities, we obtain both the net arrival rate to the node and the mean system size for the node as

$$\lambda_k = \sum_{n=0}^{N-1} \lambda_k(n) y_k(n) \tag{5.2}$$

$$L_k = \sum_{n=1}^{N} n y_k(n) \tag{5.3}$$

where k denotes the node and $k = 1, \cdots, K$. The values from Eqs. (5.2) and (5.3) will indicate how good the approximation is. If these values indicate that the approximation is not good, adjustments will be made to the exponential service rates in the complementary subnetworks and the procedures will be repeated.

The conditions for stopping the iteration are the same as that of Marie's method. In particular, a criterion will be established based on mean sizes; that is, we fix a value for ε so that the mean size stopping criterion is given as

$$\left| \frac{N - \sum_{k=1}^{K} L_k}{N} \right| < \varepsilon \; . \tag{5.4}$$

A second stopping criterion is established based on the net input rate to each node. For closed queueing networks, the vector, $\boldsymbol{x}$, of the relative input rates to each node can be found by solving the matrix equation

$$\boldsymbol{x} P = \boldsymbol{x} \tag{5.5}$$

where P is the routing matrix (in other words, $P(i,j)$ is the probability that a job will be routed to node j when it departs from node i). Eq. (5.5) is often called the balance equations and so the vectors $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ should only differ by a multiplicative constant; thus this becomes the second stopping criterion. To write this, define θ_k, for $k = 1, \cdots, K$, to be the ratio of the computed arrival rate to node k, to the solution to the balance equation; that is,

$$\theta_k = \frac{\lambda_k}{x_k} \quad \text{for } k = 1, \cdots, K \; . \tag{5.6}$$

If the values for $\boldsymbol{\lambda}$ (Eq. 5.2) are exact the ratios from (5.6) should be equal for all k, but if they are not, the algorithm should continue. Let θ denote the average value of the θ_k terms; namely,

$$\theta = \frac{1}{K} \sum_{k=1}^{K} \theta_k$$

so that the second stopping criterion is satisfied at node k when

$$\left| \frac{\theta_k - \theta}{\theta} \right| < \varepsilon \; . \tag{5.7}$$

The tolerance level ε does not necessarily have the same value for these two stopping conditions.

We use the same terminology introduced by Chandy et al. [11]. If the first condition (5.4) is not satisfied, the approximation is said to have an "insufficient-queue-length" error for the closed network if $\sum_{k=1}^{K} L_k < N$ and an "excessive-queue-length" error if $\sum_{k=1}^{K} L_k > N$. If the second condition (5.7) is not satisfied at node k, the approximation is said to have an "excessive-throughput" error at node k if $\theta_k > \theta$; otherwise, node k is said to have an "insufficient-throughput" error.

The algorithm that combines the $\lambda(n)/\mathrm{G}/1/N$ analysis with the method of Chandy et al. is as follows, with slight modifications in Steps 4, 5 and 6. For a closed network with exponential service time distributions, the proposed method is exact.

Step 1. Construct the substitute network that has the same topology as the original one but load-independent service rates μ_k for $k = 1, 2, \cdots, K$. For each node k, the initial values of μ_k are set all equal to the inverse of the mean value of the associated general service distribution.

Step 2. Determine load-dependent arrival rates $\lambda_k(n)$ for $k = 1, 2, \cdots, K$ and $n = 0, 1, \cdots, N-1$. The value of $\lambda_k(n)$ is set to $\tau_k(N-n)$ according to (5.1). Specifically, for each fixed k, $k = 1, 2, \cdots, K$, a complementary subnetwork is constructed by short-circuiting the node k (i.e., simply setting the service times of node k equal to zero) of the preceding substitute network. Then $\tau_k(N-n)$ is the load-dependent throughput rate at the short-circuiting node k when there are $(N-n)$ customers in the complementary subnetwork and can be analyzed by the methods for product-form closed networks ([8, 42]).

Step 3. Compute steady-state system length probabilities $y_k(n)$, for $k = 1, 2, \cdots, K$

and $n = 0, 1, \cdots, N$, by analyzing the $\lambda(n)$/G/1/N model using the method proposed in Chapter II.

Step 4. Check stopping conditions (5.4) and (5.7), where the tolerance level ε may be typically set to 1% or 2%. The value of θ_k in the second condition is given by (5.6), where λ_k is computed using (5.2). If both conditions are satisfied, then stop the iteration; otherwise go to Step 5.

Step 5. If there exists insufficient or excessive queue-length error, then go to Step 6. Otherwise, adjust the service rate μ_k of the equivalent exponential node by

$$\mu_{k(\text{new})} = \mu_{k(\text{old})} \frac{\theta_k}{\theta}, \qquad \text{for } k = 1, 2, \cdots, K, \tag{5.8}$$

then go to Step 2 starting the next iteration.

Step 6. For the case of excessive-queue-length error, identify all nodes with insufficient-throughput error; for the case of insufficient-queue-length error, identify all nodes with excessive-throughput error. If there exist such nodes, for each node h identified in this manner its service rate is adjusted by

$$\mu_{h(\text{new})} = \mu_{h(\text{old})} \frac{\theta_h}{\theta}, \tag{5.9}$$

then go to Step 2. If no such node is identified, go to Step 7.

Step 7. Adjust the service rates for all nodes by

$$\mu_{k(\text{new})} = \frac{\mu_{k(\text{old})}}{2} \Big(\frac{N}{\sum_{k=1}^{K} L_k} + 1 \Big), \qquad \text{for } k = 1, 2, \cdots, K, \tag{5.10}$$

then go to Step 2.

It is pointed out by Bondi and Whitt [6] that the accuracy of Marie's method is achieved by accounting for the Coxian representation when the system length probabilities are computed at each queue in isolation, and by fitting the service times with equivalent load-dependent service times used in analyzing the complementary subnetwork. From the positive aspect, the $\lambda(n)$/G/1/N model analyzed in our proposed method uses the exact expression of the service distribution rather than the two-moment Coxian approximation used in Marie's method. From another aspect, when constructing the complementary subnetwork using the proposed method, we substitute the general service times at the queues by flow-equivalent exponential times with independent rates, instead of the load-dependent rates used in Marie's method. Therefore, the likelihood for better approximate results obtained by the proposed method primarily relies on whether or not its advantage of exactly using the service time distributions outweighs its disadvantage of failing to apply load-dependent service times to the complementary subnetwork.

C. Tandem Closed Networks

Recall in Chapter IV we provided a computational technique to analyze a sequential two-queue system. The queue of the first stage is of the $\lambda(n)$/G/1/N type, and the queue of the second stage can be analyzed in isolation as a SM/PH/1/N queue as long as the semi-Markov departure process of the first queue is described. In this section, we implement the preceding two-queue technique to the approximate analysis of tandem closed networks (Figure 3). The service times at each node of the network are assumed to conform to a phase-type distribution and thus, we use the notation $\lambda(n)$/PH/1/N to stand for the $\lambda(n)$/G/1/N queue.

For each node k analyzed by the SM/PH/1/N model, steady-state joint probabilities of the system length and service phase can be computed using the method in Chapter IV. Then the load-dependent departure rates $\nu_k(n)$ are obtained straight-

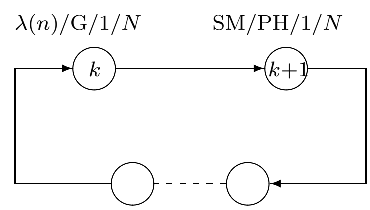

Fig. 3. The Two-Queue Technique

forwardly. However, if we substitute the arbitrary service distribution at each node k of the original network by an exponential distribution with load-dependent rates $\nu_k(n)$ and proceed along similar lines to Marie's method, stopping conditions (5.4, 5.7) are generally not fulfilled within a finite number of iterations. We suppose the failing of convergence is caused by the overflow taking place in the sequential system of the $\lambda(n)$/PH/1/N and SM/PH/1/N queues. The customer leaves the whole system when observing that the second queue is full, which violates the assumptions for the closed network analyzed by Marie's method. Therefore, we employ the approach proposed in Section V.B with necessary modifications to accommodate the two-queue technique.

The nodes of a tandem closed network are indexed sequentially from 1 to K. Before the steps of iteration, we determine the phase-type or Coxian representation for the service distribution of each node k, $k = 1, 2, \cdots, K$ by two-moment approximation. These Coxian representations are required when analyzing the $\lambda(n)$/PH/1/N and SM/PH/1/N system in Step 3. In Step 3, starting from some arbitrary node, say, node 1 of the network, we obtain a queueing system that consists of node 1 and node 2 in series and subjects to load-dependent arrivals with rates $\lambda_1(n)$ at node 1, $n = 0, 1, \cdots, N-1$. The steady-state probabilities for node 2, denoted by $y_2(n)$ for $n = 0, 1, \cdots, N$, are produced by means of the two-queue technique in Chapter IV.

The computation is repeated for each pair of sequential nodes, i.e., node 2 and node 3, $\cdots$, node $(K-1)$ and node K, node K and node 1, and system size probabilities $y_3(n), \cdots, y_K(n), y_1(n)$ are obtained for each pair. Other steps of the second approach in Section V.B shall be followed without changes.

As we pointed out in Section IV.D, for the SM/PH/1/N queue the computational work increases significantly as the network population N, the dimension of the SM kernel, or the number of phases for the PH distribution increase. Therefore, this method is only practical for tandem closed networks of small size and with small phase number at each node for the PH distribution of service times.

D. Numerical Examples

In the following tables, we use "T" to denote tandem closed networks, and use "G" to denote the networks that have the same topology as that of Figure 4. (The tandem closed network consists sequently of the nodes 1, 2, 3 and 4 of Figure 4.) The *SCV* values for all the nodes of the "general" and tandem networks range between 0 and 1/2 for Table VII, are equal to 1/2 for Table VIII, range between 1/2 and 1 for Table IX, and between 1 and 10 for Tables X and XI. The mean values of service times are respectively 1/16, 1/20, 2/25, and 3/50 for the nodes 1, 2, 3 and 4 for all the tables.

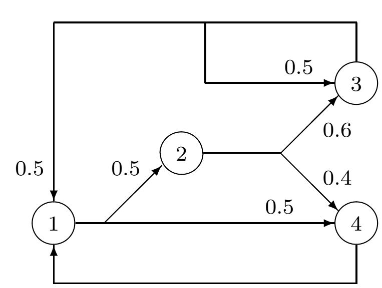

Fig. 4. A Closed Queueing Network with General Topology

Let "New_1" and "New_2" denote the two methods proposed in Sections V.B and V.C, respectively. In the Tables VII through XI, we compare the results generated by the maximum entropy (ME) method [24, 47], Marie's method and "New_1" against simulation. In the Table XII, we demonstrate the comparisons for using "New_2" to approximate the tandem networks given in Tables VII through XI. The results in Table XII are shown only for the case of $N = 3$ since the computation using "New_2" becomes very laborious when the network population gets larger. The notation n_k stands for the mean number of customers at node k, and u_k for the steady-state utilization at queue k, $k = 1, 2, 3$ and 4. For all simulation outcomes, we have 95% confidence intervals with half-width less than 2% of the mean estimate. We set the tolerance level ε equal to 1.5% for our approximations. The relative deviations shown in each table are computed by

$$\delta(\%) = \frac{\text{approximation} - \text{simulation}}{\text{simulation}} \times 100.$$

The specified Gamma distributions are used in simulation and the "New_1" method for all the tables. Associated Coxian representations are used to approximate the general service distributions at all nodes by Marie's and "New_2" methods. In addition, for the situations shown in Tables VII and XII with SCV less than 1/2, when using Marie's method and "New_2" we set the SCV values equal to 1/2 for the service distributions at all nodes so that a Cox-2 representation can be applied.

E. Conclusion

In general, all these methods are fairly accurate for the approximate values of utilization at the nodes, and the maximum entropy method is as an alternative to Marie's method. Our numerical examples show that the approximation accuracy for sys-

Table VII. Mean numbers in the system and utilizations for node 1 having service times with constant value 1/16, and nodes k for $k = 2$, 3 and 4 having Gamma(β_k, α_k)-distributed service times with $\beta_2 = 1/250$, $\alpha_2 = 25/2$, $\beta_3 = 6/625$, $\alpha_3 = 25/3$, $\beta_4 = 6/375$, $\alpha_4 = 15/4$

	$N = 3$				$N = 10$				$N = 18$			
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
T	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	0.716	-1.94	1.31	-0.15	1.012	3.14	96.58	83.03	1.012	3.31	142.08	114.24
n_2	0.526	-2.36	2.39	1.27	0.644	-12.43	83.36	70.05	0.644	-12.46	105.14	81.20
n_3	1.048	5.87	-0.92	-1.34	7.290	3.20	-31.23	-28.94	15.288	1.58	-20.99	-18.98
n_4	0.710	-4.99	-0.79	-1.86	1.055	-17.52	69.28	71.72	1.056	-18.36	103.18	102.89
u_1	0.648	1.93	-19.52	-17.05	0.781	0.00	-3.37	-1.23	0.781	-0.02	-0.20	1.14
u_2	0.518	1.93	-19.56	-17.98	0.625	0.02	-3.39	-1.87	0.625	-0.01	-0.19	0.39
u_3	0.829	1.94	-19.46	-17.68	1.000	0.00	-3.36	-2.42	1.000	0.00	-0.17	-0.09
u_4	0.622	1.95	-19.52	-18.09	0.750	0.01	-3.38	-3.23	0.750	-0.02	-0.20	0.74
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
G	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	1.127	6.05	5.27	2.25	6.546	18.75	-9.05	-8.27	14.342	14.19	-4.96	-5.09
n_2	0.322	1.99	6.87	5.56	0.406	-6.92	41.47	30.95	0.409	8.29	43.31	34.65
n_3	0.885	-8.65	-10.20	-9.30	2.070	-49.87	-0.55	-4.22	2.261	-73.17	-0.96	-4.89
n_4	0.666	0.29	-0.81	1.52	0.978	-17.05	49.85	47.71	0.988	-41.85	56.31	54.83
u_1	0.794	2.44	-8.66	-6.64	0.993	0.67	-0.87	-0.26	1.000	0.02	-0.01	0.02
u_2	0.317	2.46	-8.85	-7.69	0.397	0.68	-0.33	-0.28	0.400	0.02	-0.01	-0.31
u_3	0.609	2.45	-8.73	-5.92	0.763	0.63	-0.30	0.84	0.768	-0.05	-0.08	0.64
u_4	0.533	2.46	-8.78	-6.79	0.667	0.69	-0.30	0.33	0.672	0.03	-0.00	0.24

tem sizes tends to decrease as the network population increases. Marie's method is usually regarded as the method of choice. However, we would like to demonstrate specific situations where Marie's method may not produce accurate approximations. We show in Tables VII and VIII that the effect of a bottleneck is significant for the tandem closed network with a large network population as well as small values of SCV ($SCV \leq 1/2$) for the queueing nodes. In such situations, both Marie's method and the proposed method do not produce satisfying approximations, and the maximum entropy method is shown to be a better choice. (Even though in Table VII the Cox-2 representation was used in Marie's method, where the SCV values are not exactly as specified, it is shown in Tables VIII through XI that the results of Marie's method may not be significantly better than those of the proposed method.) As we pointed out in Section I.D, the maximum entropy method does not involve aggregation, which

Table VIII. Mean numbers in the system and utilizations for nodes k, $k = 1, 2, 3$ and 4, having Gamma(β_k, 2)-distributed service times with $\beta_1 = 1/32$, $\beta_2 = 1/40$, $\beta_3 = 1/25$, $\beta_4 = 3/100$

	$N=3$				$N=10$				$N=18$			
		$\delta(\%)$				$\delta(\%)$				$\delta(\%)$		
T	Simu.	ME	Marie	New	Simu.	ME	Marie	New	Simu.	ME	Marie	New
n_1	0.725	-0.14	0.41	-0.41	1.831	0.76	8.85	13.71	2.080	2.45	17.84	37.02
n_2	0.537	-1.12	0.74	-0.74	1.030	-0.68	14.85	16.80	1.084	-0.37	22.23	33.12
n_3	1.055	1.23	-1.14	-1.71	5.543	1.05	-9.16	-12.11	13.065	0.30	-7.55	-12.55
n_4	0.681	-0.59	1.03	0.15	1.596	-4.07	12.16	16.42	1.770	-4.80	21.19	38.02
u_1	0.549	1.09	-4.74	-7.29	0.769	-0.13	-1.69	-2.34	0.781	0.00	-0.13	2.30
u_2	0.439	1.14	-4.56	-7.52	0.614	0.00	-1.47	-2.44	0.624	0.16	0.00	0.96
u_3	0.703	1.00	-4.69	-6.97	0.983	0.00	-1.63	-3.05	0.999	0.00	-0.10	-0.40
u_4	0.527	1.14	-4.74	-7.40	0.738	-0.14	-1.76	-2.30	0.750	0.00	-0.13	1.47
		$\delta(\%)$				$\delta(\%)$				$\delta(\%)$		
G	Simu.	ME	Marie	New	Simu.	ME	Marie	New	Simu.	ME	Marie	New
n_1	1.174	2.21	1.53	0.77	5.928	6.56	1.15	-3.49	13.381	8.13	1.87	-4.77
n_2	0.335	0.30	3.28	0.90	0.517	-5.42	10.06	10.25	0.532	-9.77	10.15	11.47
n_3	0.833	-3.48	-4.08	-4.32	2.232	-15.05	-10.39	-3.27	2.674	-32.54	-16.23	-2.95
n_4	0.658	0.30	0.91	0.30	1.326	-2.11	8.30	13.20	1.413	-11.75	9.27	16.91
u_1	0.727	1.38	0.00	-2.48	0.982	0.71	0.41	-0.81	0.999	0.10	0.10	-0.10
u_2	0.291	1.37	0.00	-3.44	0.392	0.77	0.51	-0.51	0.399	0.25	0.25	-0.25
u_3	0.559	1.25	-0.18	-2.68	0.752	0.93	0.66	0.27	0.767	0.13	0.13	-0.39
u_4	0.489	1.23	0.00	-2.86	0.659	0.76	0.61	-0.15	0.671	0.15	0.15	0.15

is a principal technique employed in Marie's method. In order to use the techniques of aggregation and product-form approximations, it is assumed that the decomposition for the closed network is such that the performance of each subnetwork depends only on the state of this subnetwork (i.e., independent of the rest of the network). Since the bottleneck is a form of state-dependent behavior, it is one of the main reasons resulting in the inaccuracy of Marie's method and the proposed method. See [4] for detailed discussions regarding the assumptions required for a feasible decomposition of the closed network.

Almost all the methods in the literature for general queueing networks are of two-moment approximations. The proposed method applies exact forms of service time distributions to queueing network analysis. Our experiments for closed networks with various topologies and parameters show that the proposed method is expected

Table IX. Mean numbers in the system and utilizations for nodes k, $k = 1$, 3 and 4, having Gamma(β_k, α_k)-distributed service times with $\beta_1 = 1/32$, $\alpha_1 = 2$, $\beta_3 = 6/125$, $\alpha_3 = 5/3$, $\beta_4 = 6/125$, $\alpha_4 = 5/4$, and node 2 having exponential service times with the mean value 1/20

	$N = 3$				$N = 10$				$N = 18$			
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
T	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	0.712	0.17	0.01	0.18	1.889	0.43	3.90	6.14	2.333	2.07	7.45	16.50
n_2	0.554	-1.15	0.38	-0.88	1.214	-0.14	9.17	8.25	1.364	0.63	14.46	17.70
n_3	1.046	0.49	-1.35	-1.32	5.122	0.09	-5.92	-7.37	12.153	-0.33	-5.45	-8.66
n_4	0.689	0.01	1.04	0.28	1.775	-0.63	8.31	8.75	2.151	-0.76	13.56	21.49
u_1	0.523	0.54	-2.66	-3.65	0.754	-0.17	-1.14	-1.70	0.780	-0.09	-0.22	0.05
u_2	0.419	0.50	-2.75	-5.16	0.603	-0.18	-1.18	-2.75	0.624	-0.11	-0.25	0.12
u_3	0.670	0.49	-2.67	-3.66	0.965	-0.19	-1.34	-2.16	0.998	-0.06	-0.20	-0.51
u_4	0.500	1.04	-2.18	-4.16	0.720	0.36	-0.63	-1.86	0.745	0.44	0.30	-0.41
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
G	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	1.168	1.34	0.25	-0.33	5.738	0.43	0.23	-2.45	13.064	7.13	1.48	-2.46
n_2	0.346	-0.23	2.08	0.46	0.571	-0.74	8.89	8.44	0.596	-11.35	8.78	13.18
n_3	0.822	-2.70	-3.45	-3.04	2.243	-4.69	-7.50	-2.47	2.744	-24.49	-13.70	-3.33
n_4	0.664	1.10	1.61	1.15	1.448	5.86	8.19	10.46	1.595	-11.95	8.21	17.05
u_1	0.712	1.01	0.07	-1.40	0.975	0.13	0.42	-0.38	0.998	0.14	0.11	-0.03
u_2	0.285	1.02	-0.07	-2.77	0.390	0.13	0.64	-0.82	0.399	0.10	0.08	-0.63
u_3	0.547	0.97	0.00	-1.46	0.749	0.16	0.75	0.12	0.766	0.18	0.16	-0.42
u_4	0.476	1.60	0.57	-1.72	0.651	0.71	1.21	-0.06	0.667	0.69	0.66	-0.45

to perform more accurately than Marie's method in situations where the network population is large, most of the nodes have SCV greater than 1.0, and the network topology is complex (e.g., in Tables X and XI for the "general" closed networks with $N = 10$ and 18). For these situations, we believe that the influence of the third and higher moments is important for closed queueing networks.

Future research into an efficient way for estimating the load-dependent departure rates $\nu(n)$ of the $\lambda(n)$/G/1/N queue is promising. (In the method of Marie [29], the algorithm of computing $\nu(n)$ is derived by involving the fictitious phases of Coxian representation.) Upon producing $\nu(n)$ by some efficient algorithm, then following exact procedures of Marie's method in which load-dependent service rates at each queue k of the substitute network are updated by $\nu_k(n)$ in the iterative procedure, we would expect to achieve convergence of iteration and to yield an approach with more

Table X. Mean numbers in the system and utilizations for nodes k, $k = 1, 2, 3$ and 4, having Gamma (β_k, α_k)-distributed service times with $\beta_1 = 1/2$, $\alpha_1 = 1/8$, $\beta_2 = 1/2$, $\alpha_2 = 1/10$, $\beta_3 = 6/25$, $\alpha_3 = 1/3$, $\beta_4 = 3/10$, $\alpha_4 = 1/5$

	$N = 3$				$N = 10$				$N = 18$			
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
T	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	0.760	-1.72	-1.34	4.22	2.490	0.55	-0.72	9.05	4.371	1.43	0.59	11.04
n_2	0.596	-1.39	-2.57	3.19	1.893	0.79	-5.25	2.28	3.203	2.65	-6.83	6.46
n_3	0.965	0.36	1.35	-3.68	3.560	-2.92	1.33	-6.47	7.042	-4.27	0.90	-12.54
n_4	0.678	2.64	1.84	0.65	2.057	3.66	3.29	0.64	3.385	4.54	3.84	0.68
u_1	0.365	-7.29	3.56	9.23	0.523	-3.52	3.00	12.21	0.608	-1.63	4.05	8.74
u_2	0.291	-6.97	3.92	9.34	0.418	-3.40	3.23	11.80	0.485	-1.37	4.33	7.65
u_3	0.466	-6.87	3.97	10.72	0.668	-3.32	3.25	13.62	0.775	-1.36	4.38	7.72
u_4	0.348	-6.69	4.19	10.16	0.501	-3.41	3.23	12.37	0.581	-1.27	4.48	9.54
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
G	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	1.191	-3.49	-3.48	5.26	4.651	-1.01	-6.11	9.34	9.360	0.49	-5.29	4.75
n_2	0.398	-0.20	-3.80	-3.29	1.089	4.76	-6.10	-8.15	1.636	7.11	-9.33	0.96
n_3	0.739	3.55	6.31	-1.16	2.162	-1.28	16.62	-3.85	3.517	-4.06	20.69	-4.53
n_4	0.673	2.41	1.47	-2.14	2.098	1.10	-0.41	-10.72	3.487	-0.55	-2.27	-12.76
u_1	0.545	-7.04	-2.18	2.48	0.774	-3.93	-3.85	3.05	0.874	-1.42	-2.15	0.18
u_2	0.217	-6.72	-1.84	1.98	0.308	-3.54	-3.44	2.95	0.349	-1.27	-1.99	0.85
u_3	0.417	-6.64	-1.80	3.17	0.593	-3.68	-3.58	2.51	0.669	-1.16	-1.86	0.58
u_4	0.364	-6.33	-1.46	2.01	0.519	-3.66	-3.57	1.60	0.585	-1.08	-1.82	-1.24

Table XI. Mean numbers in the system and utilizations for nodes k, $k = 1, 2, 3$ and 4, having Gamma (β_k, α_k)-distributed service times with $\beta_1 = 5/8$, $\alpha_1 = 1/10$, $\beta_2 = 1/2$, $\alpha_2 = 1/10$, $\beta_3 = 2/5$, $\alpha_3 = 1/5$, $\beta_4 = 3/10$, $\alpha_4 = 1/5$

	$N = 3$				$N = 10$				$N = 18$			
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
T	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	0.761	-1.75	-1.46	3.44	2.517	0.09	-0.94	8.31	4.438	1.14	1.22	9.38
n_2	0.585	-0.31	-1.98	0.94	1.842	1.14	-4.65	-3.09	3.129	2.16	-7.04	-2.79
n_3	0.977	-0.40	1.36	1.59	3.577	-2.62	1.74	0.51	6.968	-3.56	2.53	-8.19
n_4	0.676	2.80	1.38	-0.21	2.063	3.42	2.29	-4.20	3.465	3.74	-0.31	-3.45
u_1	0.353	-7.75	4.36	8.46	0.501	-4.35	2.56	11.80	0.583	-2.25	3.88	10.65
u_2	0.282	-7.62	4.50	9.32	0.401	-4.37	2.52	11.38	0.466	-2.19	3.97	9.57
u_3	0.451	-7.43	4.70	11.69	0.641	-4.29	2.56	13.65	0.744	-1.96	4.21	5.75
u_4	0.338	-7.48	4.64	12.36	0.480	-4.21	2.75	14.29	0.558	-1.99	4.21	7.69
			$\delta(\%)$				$\delta(\%)$				$\delta(\%)$	
G	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1	Simu.	ME	Marie	New_1
n_1	1.190	-4.25	-3.98	6.15	4.604	-2.65	-7.30	9.44	9.141	-1.14	-6.47	4.78
n_2	0.391	0.31	-3.30	-3.12	1.077	2.81	-6.91	-9.83	1.625	5.40	-11.24	-1.80
n_3	0.751	5.19	6.91	1.96	2.250	6.28	18.31	2.52	3.754	5.47	24.45	-0.68
n_4	0.669	1.57	1.26	-3.40	2.069	1.97	-0.06	-12.58	3.479	-5.43	-4.12	-14.74
u_1	0.531	-8.05	-1.87	3.03	0.751	-5.65	-5.14	2.68	0.852	-2.99	-3.40	-0.90
u_2	0.212	-7.80	-1.61	3.40	0.300	-5.60	-5.06	4.13	0.341	-2.94	-3.32	0.85
u_3	0.406	-7.71	-1.53	1.95	0.574	-5.19	-4.64	1.24	0.653	-2.79	-3.14	-4.06
u_4	0.356	-7.79	-1.63	2.81	0.504	-5.44	-4.91	2.72	0.572	-2.95	-3.32	-2.76

Table XII. Mean numbers in the system and utilizations for nodes k, $k = 1, 2, 3,$ and 4, of tandem closed networks with $N = 3$

	Table VII		Table VIII		Table IX		Table X		Table XI	
	Simu.	$\delta(\%)$ New$_2$	Simu.	$\delta(\%)$ New$_2$	Simu.	$\delta(\%)$ New$_2$	Simu.	$\delta(\%)$ New$_2$	Simu.	$\delta(\%)$ New$_2$
n_1	0.716	3.03	0.725	0.83	0.711	3.04	0.760	0.85	0.761	1.12
n_2	0.526	3.73	0.537	0.74	0.554	4.29	0.596	2.89	0.585	3.88
n_3	1.048	-3.30	1.055	-4.83	1.046	-4.48	0.965	-3.33	0.977	-3.63
n_4	0.710	-4.90	0.681	-1.47	0.689	1.41	0.678	5.35	0.676	3.68
u_1	0.648	-23.63	0.549	-10.56	0.523	-6.03	0.365	22.18	0.353	25.69
u_2	0.518	-23.58	0.439	-10.48	0.419	-7.12	0.291	22.59	0.282	25.87
u_3	0.829	-23.93	0.703	-9.53	0.670	-5.69	0.466	23.81	0.451	27.12
u_4	0.622	-23.83	0.527	-9.68	0.500	-5.72	0.348	23.07	0.338	26.54

accuracy than Marie's method. In addition, the proposed method is promising for networks that contain multi-server queues by making an extension of the $\lambda(n)/G/1/N$ queue to the associated model with multiple servers. In general, the multi-server queues are not computationally tractable for exact solutions. Approximate techniques for the M/G/c queues with both infinite and finite buffers can be found, for example, in [46].

Comparing Table XII with corresponding results in Tables VII through XI, we observe that the method proposed in Section V.C for tandem closed networks is less accurate than Marie's method and the method of Section V.B. We believe this is primarily due to the overflows occuring in the sequential two-queue system.

CHAPTER VI

SUMMARY OF CONTRIBUTIONS

A number of problems in queueing theory and stochastic processes are addressed in this dissertation. These problems arise in an attempt to improve approximations for closed queueing networks.

For the $\lambda(n)$/G/1/N queue, the method of Gupta and Rao is possibly the most popular in the literature that provides computational solutions of system size probabilities. However, there are two main difficulties in implementing their method. First, whenever a new $\lambda(n)$/G/1/N system is analyzed with a different combination of identical and distinct values of $\lambda(n)$, $n = 0, 1, \cdots, N-1$, the basic equations of their method need to be rewritten and the procedure to solve the equations repeated. Second, when the values of $\lambda(n)$ are distinct but close to each other, their method may fail to produce accurate or reasonable results. Our algorithm developed in Chapter II significantly eliminates these two difficulties and, thus, produces more stable and accurate results for the $\lambda(n)$/G/1/N queue.

The computation of the time-dependent probability matrix $P(n,t)$ for $n \in \mathcal{N}$ and $t \in \mathcal{R}^+$ associated with a renewal process is essential in queueing and stochastic analysis. However, there usually does not exist a concise explicit solution for $P(n,t)$, as pointed out by Neuts in [40, pp. 389] where he derived formulas for the moments of $U(t)$ for the counting process of PH renewals. An approach was also outlined by Neuts in [36, pp. 68] to obtain the approximate value of $P(n,t)$ for every point of t, $t \in \mathcal{R}^+$, which requires extensive calculations. In Chapter III, we presented a procedure to generate the exact solution of $P(n,t)$ for PH renewal processes. In particular, we developed closed-form solutions for $P(n,t)$ when the Cox-2 and simplified Cox-r

distributions are used to fit inter-renewal times of arbitrary mean and variance.

Extensive discussion for the SM/PH/1/N queue has not appeared in the literature so far. Based on Neuts and Chakravarthy's analysis for the SM/PH/1 queue [37], which has infinite waiting space, and our preceding work for obtaining exact expressions for the $P(n, t)$ terms, an explicit procedure has been proposed in Chapter IV for the system length probabilities and waiting time distributions of the SM/PH/1/N system. In addition, closed-form solutions are given for the case that the Laplace-Stieltjes transform of the semi-Markov kernel can be obtained in closed form and the service times conform to a Cox-2 distribution.

Among the numerous methods for the approximate analysis of general closed queueing networks, Marie's method is regarded as the best choice in accuracy and computational efficiency, where each queue of the network is analyzed in isolation as a $\lambda(n)/\mathrm{C}_r/1/N$ model. By implementing our preceding algorithm for the $\lambda(n)/\mathrm{G}/1/N$ queue, in Chapter V we proposed a procedure to approximate the state probabilities for the closed network with an arbitrary topology. The proposed procedure is unique in that, to the best of our knowledge, almost all existing methods employ two-moment approximations rather than exact forms of the service distributions of the queues of the network. For tandem closed networks, we further developed a new approach using the queueing model consisting of the $\lambda(n)/\mathrm{PH}/1/N$ and SM/PH/1/N in series.

REFERENCES

[1] I. F. Akyildiz and A. Sieber, “Approximate analysis of load dependent general queueing networks”, *IEEE Transactions on Software Engineering*, vol. 14, pp. 1537–1545, 1988.

[2] T. Altiok, “On the phase-type approximations of general distributions”, *IIE Transactions*, vol. 17, pp. 110–116, 1985.

[3] F. Baskett, K. M. Chandy, R. R. Muntz, and F. Palacios-Gomez, “Open, closed and mixed networks of queues with different classes of customers”, *Journal of the ACM*, vol. 22, pp. 248–260, 1975.

[4] B. Baynat and Y. Dallery, “A unified view of product-form approximation techniques for general closed queueing networks”, *Performance Evaluation*, vol. 18, pp. 205–224, 1993.

[5] B. Baynat and Y. Dallery, “A product-form approximation method for general closed queueing networks with several classes of customers”, *Performance Evaluation*, vol. 24, pp. 165–188, 1996.

[6] A. B. Bondi and W. Whitt, “The influence of service-time variability in a closed network of queues”, *Performance Evaluation*, vol. 6, pp. 219–234, 1986.

[7] A. Brandwajn, “Equivalence and decomposition in queueing systems: A unified approach”, *Performance Evaluation*, vol. 5, pp. 175–186, 1985.

[8] J. P. Buzen, “Computational algorithms for closed queueing networks with exponential servers”, *Communications of the ACM*, vol. 16, pp. 527–531, 1973.

[9] S. R. Chakravarthy and A. S. Alfa, Eds., *Matrix-Analytic Mehtods in Stochastic Models*. New York, NY: Marcel Dekker, 1997.

[10] K. M. Chandy, U. Herzog, and L. Woo, “Parametric analysis of queueing networks”, *IBM Journal of Research and Development*, vol. 19, pp. 36–42, 1975.

[11] K. M. Chandy, U. Herzog, and L. Woo, “Approximation analysis of general queueing networks”, *IBM Journal of Research and Development*, vol. 19, pp. 43–49, 1975.

[12] P. J. Courtois and J. Georges, “On a single-server finite queuing model with state-dependent arrival and service processes”, *Operations Research*, vol. 19, pp. 424–435, 1971.

[13] D. R. Cox, "A use of complex probabilities in the theory of stochastic processes", in *Proceedings of Cambridge Philosophical Society*, vol. 51, pp. 313–319, 1955.

[14] E. Çinlar, *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[15] Y. Dallery and X. Cao, "Operational analysis of stochastic closed queueing networks", *Performance Evaluation*, vol. 14, pp. 43–61, 1992.

[16] J. G. Dai and J. M. Harrison, "The QNET method for two-moment analysis of closed manufacturing systems", *The Annals of Applied Probability*, vol. 3, pp. 968–1012, 1993.

[17] U. C. Gupta and T. S. S. Srinivasa Rao, "On the M/G/1 machine interference model with spares", *European Journal of Operational Research*, vol. 89, pp. 164–171, 1996.

[18] U. C. Gupta and T. S. S. Srinivasa Rao, "Computing steady state probabilities in $\lambda(n)/G/1/K$ queue", *Performance Evaluation*, vol. 24, pp. 265–275, 1996.

[19] J. M. Harrison, R. J. Williams, and H. Chen, "Brownian models of closed queueing networks with homogeneous customer populations", *Stochastics and Stochastic Reports*, vol. 29, pp. 37–74, 1990.

[20] U. Herzog, L. Woo, and K. M. Chandy, "Solution of queuing problems by a recursive technique", *IBM Journal of Research and Development*, vol. 19, pp. 295–300, 1975.

[21] W. J. Hopp and M. L. Spearman, *Factory Physics*. New York, NY: McGraw-Hill, 2nd edition, 2001.

[22] H. Kobayashi, "Application of the diffusion approximation to queueing networks I: Equilibrium queue distributions", *Journal of the ACM*, vol. 21, pp. 316–328, 1974.

[23] H. Kobayashi, "Application of the diffusion approximation to queueing networks II: Nonequilibrium distributions and applications to computer modeling", *Journal of the ACM*, vol. 21, pp. 459–469, 1974.

[24] D. D. Kouvatsos, "Maximum entropy methods for general queueing networks", in *Modeling Techniques and Tools for Performance Analysis*. D. Potier, Ed., New York, NY: Elsevier Science, 1985, pp. 589–608.

[25] C. Langaris, "On a two-stage birth and death queueing process", *Operations Research*, vol. 37, pp. 488–497, 1989.

[26] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia, PA: American Statistical Association and Society for Industrial and Applied Mathematics, 1999.

[27] G. Latouche and P. Taylor, Eds., in *Matrix-Analytic Methods: Theory and Applications: Proceedings of the Fourth International Conference*, Adelaide, Australia, July 2002.

[28] R. A. Marie, "An approximate analytical method for general queueing networks", *IEEE Transaction on Software Engineering*, vol. 5, pp. 530–538, 1979.

[29] R. A. Marie, "Calculating equilibrium probabilities for $\lambda(n)/\mathrm{C}_k/1/N$ queue", *Performance Evaluation Review*, vol. 9, pp. 117–125, 1980.

[30] R. A. Marie and J. M. Pellaumail, "Steady-state probabilities for a queue with a general service distribution and state-dependent arrivals", *IEEE Transactions on Software Engineering*, vol.9, pp. 109–113, 1983.

[31] S. Nahmias, *Production and Operations Analysis*. Chicago, IL: IRWIN, 3rd edition, 1997.

[32] D. Neuse and K.M. Chandy, "HAM: The heuristic aggregation method for solving general closed queueing network models of computer systems", *Performance Evaluation Review*, vol. 11, pp. 195–212, 1982.

[33] M. F. Neuts, "Two queues in series with a finite, intermediate waitingroom", *Journal of Applied Probability*, vol. 5, pp. 123–142, 1968.

[34] M. F. Neuts, "Two servers in series, studied in terms of a Markov renewal branching process", *Advances in Applied Probability*, vol. 2, pp. 110–149, 1970.

[35] M. F. Neuts, "Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector", *Advances in Applied Probability*, vol. 10, pp. 185–212, 1978.

[36] M. F. Neuts, *Matrix-geometric Solutions in Stochastic Models: an Algorithmic Approach*. Baltimore, MD: Johns Hopkins University Press, 1981.

[37] M. F. Neuts and S. Chakravarthy, "A single server queue with platooned arrivals and phase type services", *European Journal of Operational Research*, vol. 8, pp. 379–389, 1981.

[38] M. F. Neuts, "Matrix-analytic methods in queueing theory", *European Journal of Operational Research*, vol. 15, pp. 2–12, 1984.

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

[40] M. F. Neuts, *Algorithmic Probability: a Collection of Problems*. London, England: Chapman & Hall, 1995.

[41] S. Nojo and H. Watanabe, "A new stage method getting arbitrary coefficient of variation by two stages", *The Transactions of the IEICE*, vol. E70, pp. 33–36, 1987.

[42] M. Reiser and S. S. Lavenberg, "Mean value analysis of closed multichain queueing networks", *Journal of the ACM*, vol. 27, pp. 313–322, 1980.

[43] E. Schwerer and J. A. Van Mieghem, "Brownian models of closed queueing networks: Explicit solutions for balanced three-station systems", *The Annals of Applied Probability*, vol. 4, pp. 448–477, 1994.

[44] W. J. Stewart and R. A. Marie, "A numerical solution for the $\lambda(n)/\mathrm{C}_k/r/N$", *European Journal of Operational Research*, vol. 5, pp. 56–68, 1980.

[45] W. T. Thomson, *Laplace Transformation*. Englewood Cliffs, NJ: Prentice-Hall, 1960.

[46] H. C. Tijms, *A First Course in Stochastic Models*. New York, NY: John Wiley & Sons, 2003.

[47] R. J. Walstra, "Nonexponential networks of queues: A maximum entropy analysis", *Performance Evaluation Review*, vol. 13, pp. 27–37, 1985.

[48] W. Whitt, "Open and closed models for networks of queues", *AT&T Bell Laboratories Technical Journal*, vol. 63, pp. 1911–1979, 1984.

[49] C. J. Willits and D. C. Dietz, "Rapid, efficient analysis of the $\lambda(n)/\mathrm{C}_k/r/N$ queue, with application to decomposition of closed queuing networks", *Computers & Operations Research*, vol. 25, pp. 543–556, 1998.

APPENDIX A

GUPTA AND RAO'S METHOD FOR $\lambda(n)$/G/1/N QUEUE

The method of Gupta and Rao [17, 18] is possibly the most popular in the literature to analyze and produce numerical solutions for the $\lambda(n)$/G/1/N Queue. It is assumed there exists an integer N such that $\lambda(n) > 0$ for all $0 \leq n < N$ and $\lambda(n) = 0$ for all $n \geq N$. The service time distribution φ is arbitrary as long as its Laplace transform φ^* can be obtained in closed form and the mean service time is of finite value. Let $V(t)$ and $U(t)$ denote the number of customers at present and the remaining service time for the customer in service at time t, respectively, and define

$$p_j(t) = \mathrm{P}\{V(t) = j\} \text{ for } j = 0, 1, \cdots, N, \text{ and} \tag{A.1}$$

$$p_j(u,t)\,du = \mathrm{P}\{V(t) = j, u < U(t) \leq u + du\} \text{ for } u \geq 0, j = 1, 2, \cdots, N. \tag{A.2}$$

If the $\lambda(n)$ values are either all identical or all distinct from each other, the basic equations are given in Gupta and Rao [18] as follows:

$$\frac{d}{dt}p_0(t) = -\lambda(0)p_0(t) + p_1(0,t); \tag{A.3}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_1(u,t) = -\lambda(1)p_1(u,t) + \lambda(0)p_0(t)\varphi(u) + p_2(0,t)\varphi(u); \tag{A.4}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_j(u,t) = -\lambda(j)p_j(u,t) + \lambda(j-1)p_{j-1}(t) + p_{j+1}(0,t)\varphi(u), \tag{A.5}$$

for $j = 2, 3, \cdots, N-1$; and

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_N(u,t) = -\lambda(N-1)p_{N-1}(u,t). \tag{A.6}$$

The steady-state system size probability at an arbitrary point of time is then given by $y_j = \lim_{t\to\infty} p_j(t)$ for $j = 0, 1, \cdots, N$. The following steps of a recursive algorithm are derived after solving the preceding basic equations:

Step 1. Compute $\nu(1)$ by

$$\nu(1) = \lambda(0)\frac{[1-\varphi^*(\lambda(1))]}{\varphi^*(\lambda(1))}p_0, \tag{A.7}$$

where p_0 is set equal to 1.

Step 2. Compute p_1 by

$$p_1 = \frac{\nu(1)}{\lambda(1)}. \tag{A.8}$$

Step 3. If the $\lambda(n)$ values are all distinct, then for each $j = 2, 3, \cdots, N-1$ do the Steps 3.1 through 3.3. Otherwise if $\lambda(n) = \lambda$ for all $n = 0, 1, \cdots, N-1$, then do Steps 3.4 through 3.6.

3.1. Compute $\overline{p}_1(\lambda(j))$ by

$$\overline{p}_1(\lambda(j)) = \frac{1}{\lambda(1)-\lambda(j)}\Big[\lambda(0)\,p_0\,(\varphi^*(\lambda(j))-1)+\nu(1)\,\varphi^*(\lambda(j))\Big]. \tag{A.9}$$

3.2. If $j = 2$ then go to Step 3.3. Otherwise, for each $i = 2, \cdots, j-1$, compute $\overline{p}_i(\lambda(j))$ by

$$\overline{p}_i(\lambda(j)) = \frac{1}{\lambda(i)-\lambda(j)}\Big[\lambda(i-1)\,\overline{p}_{i-1}(\lambda(j))+\nu(i)\,\varphi^*(\lambda(j))-\nu_(i-1)\Big]. \tag{A.10}$$

3.3. Compute $\nu(j)$ using

$$\nu(j) = \frac{1}{\varphi^*(\lambda(j))}\Big[\nu(j-1)-\lambda(j-1)\,\overline{p}_{j-1}(\lambda(j))\Big]. \tag{A.11}$$

Go to Step 4.

3.4. For each $i = 0, 1, \cdots, N-3$, compute $\bar{p}_1(i,\lambda)$ using

$$\bar{p}_1(i,\lambda) = -\frac{1}{i+1}(\lambda\,p_0+\nu(1))\,\varphi^{*(i+1)}(\lambda). \tag{A.12}$$

3.5. For each $j = 2, \cdots, N-2$ and $i = 0, 1, \cdots, N-2-j$, compute $\nu(j)$ and

$\bar{p}_j(i,\lambda)$ using Eq. (A.11) and

$$\bar{p}_j(i,\lambda) = -\frac{1}{i+1}\Big[\lambda\,\bar{p}_{j-1}(i+1,\lambda) + \nu(j)\,\varphi^{*(i+1)}(\lambda)\Big]. \tag{A.13}$$

3.6. Set $\bar{p}_j(\lambda)$ equal to $\bar{p}_j(0,\lambda)$ for all $j = 1, 2, \cdots, N-2$, then compute $\nu(N-1)$ using Eq. (A.11).

Step 4. For all $j = 2, 3, \cdots, N-1$, compute p_j using

$$p_j = \frac{1}{\lambda(j)}\Big[\lambda(j-1)\,p_{j-1} + \nu(j) - \nu(j-1)\Big]. \tag{A.14}$$

Step 5. Compute p_N using

$$p_N = -\lambda(N-1)\,\hat{p}_{N-1}, \tag{A.15}$$

where $\hat{p}_{N-1}$ is determined by

$$\hat{p}_1 = \frac{1}{\lambda(1)}\Big[-\lambda(0)\,p_0\,b - \nu(1)\,b + p_1\Big], \tag{A.16}$$

$$\hat{p}_j = \frac{1}{\lambda(j)}\Big[\lambda(j-1)\,\hat{p}_{j-1} - \nu(j)\,b + p_j\Big], \qquad \text{for } N-1 \geq j \geq 2, \tag{A.17}$$

and b is the mean service time.

Step 6. For all $j = 0, 1, \cdots, N-1$, the steady-state probabilities immediately after a departure, denoted by π_j, are computed by

$$\pi_j = \frac{\nu_j}{\sum_{i=0}^{N-1} \nu_i}, \tag{A.18}$$

where $\nu(0)$ is equal to $\lambda(0)\,p_0$.

Step 7. For all $j = 0, 1, \cdots, N$, the steady-state probabilities y_j are computed by

$$y_j = \frac{p_j}{\sum_{i=0}^{N} p_i}. \tag{A.19}$$

The above procedure can be modified to accommodate the situation where some of the $\lambda(n)$ values are identical and some are distinct. For example, in [17], Gupta and Rao calculated the system size probabilities of a M/G/1/N machine interference problem with Y spares, which is essentially a $\lambda(n)$/G/1/$(N+Y)$ model with $\lambda(n) = N\lambda$ for $0 \leq n \leq Y$ and $\lambda(n) = (N+Y-n)\lambda$ for $Y+1 \leq n \leq N+Y-1$. The basic equations are rewritten by,

$$\frac{d}{dt}p_0(t) = -N\,\lambda\,p_0(t) + p_1(0,t); \tag{A.20}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_1(u,t) = -N\,\lambda\,p_1(u,t) + N\,\lambda\,p_0(t)\varphi(u) + p_2(0,t)\varphi(u); \tag{A.21}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_j(u,t) = -N\,\lambda\,p_j(u,t) + N\,\lambda\,p_{j-1}(t) + p_{j+1}(0,t)\varphi(u), \tag{A.22}$$

for $j = 2, 3, \cdots, Y$;

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_j(u,t) = -(N+Y-j)\,\lambda\,p_j(u,t) + (N+Y-n+1)\,\lambda\,p_{j-1}(t) + p_{j+1}(0,t)\varphi(u), \tag{A.23}$$

for $j = Y+1, Y+2, \cdots, N+Y-1$; and

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) p_{N+Y-1}(u,t) = -\lambda\,p_{N+Y-1}(u,t). \tag{A.24}$$

The analogous procedure is therefore derived for the new system with a different combination of identical and distinct values of $\lambda(n)$. See Gupta and Rao [17] for detailed discussion.

APPENDIX B

MARIE'S METHOD FOR GENERAL CLOSED QUEUEING NETWORKS

Marie's method [28, 29] has been empirically shown to be the best choice for approximate analysis of non-product-form closed queueing networks, and was extended for networks of multi-server queues by Stewart and Marie [44] and Willits [49], respectively. The principle of Marie's method is to analyze each queue of the network (consisting of K queueing nodes and having N customers inside) in isolation as a model with load-dependent arrival rates $\lambda_k(n)$ for $k = 1, 2, \cdots, K$ and $n = 0, 1, \cdots, N-1$. To determine the unknown values $\lambda_k(n)$, the non-product-form network is approximated by a product-form network by substituting exponential service times with load-dependent rates $\mu_k(n)$, $n = 1, 2, \cdots, N$, for the original queues that have general service times.

Step 1. Construct the substitute network which has the same topology as the original one but load-dependent service rates $\mu_k(n)$ for $k = 1, 2, \cdots, K$ and $n = 1, 2, \cdots, N$. For each queue k, the initial values of $\mu_k(n)$ are set all equal to the inverse of the mean value of the associated general service distribution.

Step 2. Determine load-dependent arrival rates $\lambda_k(n)$ for $k = 1, 2, \cdots, K$ and $n = 0, 1, \cdots, N-1$. The value of $\lambda_k(n)$ is set to $\tau_k(N-n)$ according to (5.1). Specifically, for each fixed k, $k = 1, 2, \cdots, K$, a complementary subnetwork is constructed by short-circuiting the queue k (i.e., simply setting the service times of queue k equal to zero) of the preceding substitute network. Then $\tau_k(N-n)$ is the load-dependent throughput rate at the short-circuiting queue k when there are $(N-n)$ customers in the complementary subnetwork, and can be analyzed by the methods for product-form closed networks ([8, 42]).

Step 3. Analyze each queue k for $k = 1, 2, \cdots, K$ as an individual $\lambda(n)/\mathrm{C}_r/1/N$ queue. The computation for load-dependent departure rates $\nu_k(n)$ is given in [29], which is particularly efficient for the Cox-2 and Erlang-r service distributions. Then the steady-state system size probabilities $y_k(n)$ are determined by

$$y_k(n) = y_k(0) \sum_{i=0}^{n-1} \frac{\lambda_k(i)}{\nu_k(i+1)} \qquad \text{for } n = 1, 2, \cdots, N, \text{ and} \tag{B.1}$$

$$\sum_{n=0}^{N} y_k(n) = 1. \tag{B.2}$$

Step 4. Check stopping conditions (5.4) and (5.7). If both conditions are satisfied, then stop the iteration; otherwise go to Step 5.

Step 5. The load-dependent service rates $\mu_k(n)$ are updated equal to $\nu_k(n)$, for $k = 1, 2, \cdots, K$ and $n = 1, 2, \cdots, N$, for the substitute network. Then return to Step 2.

VITA

Feng Sun received his B.S. degree in Electrical Engineering in 1992 and M.S. degree in Aerocraft Design in 1995, both from Beijing University of Aeronautics and Astronautics (BUAA) in China, and Ph.D. degree in Industrial Engineering in 2006 from Texas A&M University at College Station. He worked as a full-time lecturer at BUAA Department of Automatic Control from 1995 to 1998. His research interests include queueing systems, applied probability, and stochastic processes. He can be reached at fsun70@hotmail.com.