Doctoral Dissertation

Studies on Queueing Models Interacting with Underlying Processes

Yoshiaki Inoue

January 2016

Graduate School of Engineering,

Osaka University

Preface

Congestion of data traffic is one of the most significant issues to be addressed in designing communication systems. Queueing theory is a branch of applied mathematics, which provides a set of mathematical tools for evaluating the impact of congestion on the performance of systems. In queueing theory, analytical methods for abstract mathematical models, called queueing models, are studied. Along with the development of communication technologies, a large variety of queueing models have been analyzed.

In particular, queueing models with underlying processes are fundamental models for the current communication networks, where data traffic has the following two characteristics: (i) traffic patterns are bursty rather than completely random and (ii) traffic consists of multiple data streams with different properties such as text, audio, video, and control packets. This class of queueing models is well studied, and efficient computational algorithms for performance measures are available in the literature.

Presently, adaptive resource allocation mechanisms are being developed in each field of communication technology. For example, dynamic spectrum allocation in cognitive radio, reconfigurable wavelength division multiplexing (WDM), and dynamic route control based on the software-defined network (SDN) technology are regarded as adaptive resource allocation mechanisms. To make efficient use of bandwidths, these technologies dynamically change the allocation of bandwidths and restrict the volume of incoming traffic through admission control. Mathematically, communication systems with adaptive resource allocation mechanisms can be formulated as queueing models *interacting with* underlying processes. Interaction between queues and underlying processes makes the behavior of systems more complicated, so that known results for queueing models with underlying processes cannot be applied directly.

The main contribution of this dissertation is the development of analytical and computational methods for two kinds of queueing models interacting with underlying processes, which are regarded as fundamental models for communication systems with adaptive resource allocation mechanisms. In the first model, the state of the underlying process is switched when the system becomes empty. In the second model, on the other hand, the state of the underlying process is assumed to change *continually* according to the workload in system. Although these two models do not cover the whole class of queueing models interacting with underlying processes, they formulate two fundamental behaviors of adaptive resource allocation mechanisms. The first model corresponds to allocation mechanisms that bandwidth is added when congestion continues for a while, and the additional bandwidth is released when the congestion gets relieved. This is reasonable because the addition and the removal of bandwidth are usually performed with a delay, so that they cannot be executed very frequently. On the other hand, the second model corresponds to allocation mechanisms that are performed in real-time, such as an adaptive admission control. Note that the first model describes relatively long-term changes in the state of the underlying process, while the second model describes short-term changes in it. In this dissertation, we analyze these two kinds of queueing models separately so that their mathematical structures to be well understood.

Chapter 1 provides a background of this study, and introduces the two kinds of queueing models mentioned above. Chapter 2 presents a basic approach to analyzing the first model, through an extensive analysis of a closely related model referred to as the multi-class M/G/1 queue with working vacations. Chapter 3 analyzes the queueing model with disasters and multiple Markovian arrival streams, where the approach of Chapter 2 is extended into a more general case that arrival streams are governed by an underlying Markov chain. Chapter 4 generalizes the results in Chapters 2 and 3 by considering a continuous-time Markov process with the skip-free to the left property and reducible generators for busy periods, which provides a unified way to analyze the first model.

Chapters 5, 6, and 7 are devoted to a queueing model with impatient customers, which is formulated as a stochastic process equivalent to the second model. This model has been studied for a long time, and a basic result for the stationary workload was already obtained in 1961 by Kovalenko. While this result for the stationary workload is often used as a starting point of the analysis of another performance measures of interest such as the stationary queue length and busy periods, we could not find further results for the stationary workload in the literature. Chapters 5 revisits the formula obtained by Kovalenko and provides a new perspective on it. This leads to a unified understanding of special cases of this model, for which analytical results were reported independently in several research papers. Chapter 6 analyzes the stationary loss probability, which is the fundamental quantity of interest in the second model. Based on the results in Chapter 5, various properties of the loss probability such as theoretical lower and upper bounds and stochastic ordering relations are derived. In Chapter 7, a computational algorithm for the loss probability is developed. This computational algorithm has a remarkable feature that it also outputs an upper bound of its numerical error. Finally, we conclude this dissertation in Chapter 8.

This dissertation summarizes my studies on queueing theory in Doctor's Course of Department of Information and Communications Technology, Graduate School of Engineering, Osaka University. Contents of every chapter in this dissertation except Chapters 1 and 8 are based on some publications shown in the publication list as follows:

Chapter 2: Publications A-1 and C-1,

Chapter 3: Publications A-2, B-1, C-3, and C-4,

Chapter 4: Publications A-4 and C-7,

Chapter 5: Publications A-3 and C-8,

Chapter 6: Publications A-3, A-5, and C-11,

Chapter 7: Publications C-10 and C-11.

Yoshiaki Inoue

Osaka University
January 2016

Acknowledgements

I would like to express my deepest appreciation to Professor Tetsuya Takine of Osaka University for his supervision and continuing encouragement through the course of my research in Osaka University. I really enjoyed attending his lecture on queueing theory in my undergraduate course, which became the entrance to my research on this field. It was a great pleasure to work with him, and I have learned a lot of things from him. His profound insights always stimulated me, and his professional, diligent, and rigorous style of work has had a great influence on me. Without his tremendous support, none of this work would have been done.

I would like to express my gratitude to my committee members, Professor Atsuko Miyaji and Associate Professor Takahiro Matsuda of Osaka University for their invaluable comments on this dissertation. Associate Professor Takahiro Matsuda helped me a lot in my course of study in Osaka University, and I am very grateful to him for his persistent support.

I also take pleasure in thanking Professor Noboru Babaguchi, Professor Kyou Inoue, Professor Kenichi Kitayama, Professor Kazunori Komatani, Professor Seiichi Sampei, and Professor Takashi Washio. Their valuable discussions and favorable comments greatly enhanced the quality of this dissertation.

My sincere thanks are due to Associate Professor Masahiro Sasabe of Nara Institute of Science and Technology, Associate Professor Kouji Hirata of Kansai University, Lecturer Takanori Kudo of Setsunan University, Assistant Professor Tomotaka Kimura of Tokyo University of Science, and all other colleagues of Takine Laboratory for their kind help in many ways. I also would like to express my sincere gratitude to Mrs. Yumi Shimoyashiki for her enormous support for my research activities.

I would like to express my sincere gratitude to Mr. Kenji Higashi, the CEO of Da Vinci Co., Ltd. I was taught by him a lot of things about business administration and management, as well as various technical perspectives of heat flows, which largely broadened my outlook. Also I would like to express my appreciation to Associate Professor Atsue Ishii of Osaka University and all members of Ishii Laboratory. Discussions with them about nursing, health-care systems, operations

research, and so on greatly motivated me.

I am very grateful to Professor Onno Boxma of Eindhoven University of Technology and all other colleagues of his research group. My half-year stay in Eindhoven was a great experience, and it stimulated me in many ways. It was my real pleasure to spend time with them.

Last, but not least, I heartily thank my parents for their understanding, constant support, and sincere encouragement.

Abbreviations and Conventions

Throughout this dissertation, we use the following abbreviations:

FCFS	: First-Come First-Served,
i.i.d.	: independent and identically distributed,
LCFS-PR:	Last-Come First-Served Preemptive-Resume,
LST	: Laplace-Stieltjes transform,
MAP	: Markovian arrival process,
PDF	: probability distribution function,
p.d.f.	: probability density function.

In addition, we follow the following conventions of mathematical notation unless otherwise mentioned:

- Vectors are denoted by bold-type lower-case letters.
- Matrices are denoted by bold-type upper-case letters.
- Inequalities between matrices or vectors imply that they hold elementwise.
- Empty sum terms are defined as zero.
- Empty product terms are defined as one.
- 0^0 is defined as one.

Contents

List of Figures

1 Introduction

1.1 Queueing theory

For the last several decades, communication technologies have been highly developed, and they are still under rapid development. The most notable example is the emergence and the growth of the Internet, which is a huge communication network connecting billions of people world-wide. Today, the Internet has become a vital part of our everyday life and it is used for a variety of services, such as the World Wide Web, E-mail, E-commerce, video and photo sharing, and social networks. Technically, the Internet is *a network of networks*, which interconnects a large number of networks operated by different service providers. In those networks, various types of physical media are used for communication links, e.g., copper wire, optical fiber, and radio spectrum.

Although properties and performances of communication links highly depend on the nature of the underlying physical media, there is a common important restriction on them: each communication link has a *limited* bandwidth (or capacity). Because of this fact, data packets arriving at a communication link have to wait in a buffer to be transmitted when the instantaneous volume of the input traffic exceeds the bandwidth. This causes congestion of data traffic, which has a significant impact on the performance of communication systems. Therefore, regardless of the physical media used, it is important in designing communication systems to understand the congestion phenomena occurring in them.

Queueing theory provides a set of mathematical tools for evaluating the impact of the congestion on the performance of a communication system. Based on the theory of probability and stochastic processes, queueing theory analyzes various properties of abstract mathematical models called *queueing models*. In this dissertation, we focus our attention on single-server queueing models (Figure 1.1), where the system consists of a server and a waiting room with infinite capacity. Customers arrive at the system, wait for their turns, and leave the system on their service completion. Single-server queueing models are used as models of single communication links, and they are building blocks of networked systems. Note here that customers

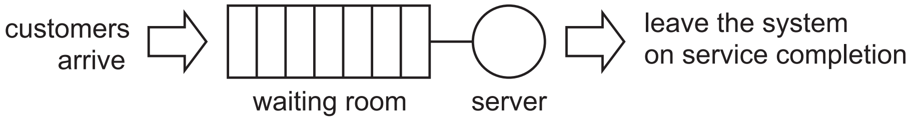

Figure 1.1: Single-server queueing model.

in this case represent incoming packets.

Using Kendall's notation, single-server queueing models are usually denoted by A/B/1, where A and B represent the arrival process of customers and the service time distribution, respectively, and the last symbol "1" means that there is only one server. Various symbols are used for A and B to specify the arrival process and the service time distribution, e.g., "D" denotes the deterministic distribution, "M" denotes the exponential distribution, "G" denotes the general non-negative probability distribution, and so on. To determine the behavior of a queueing system, we also need to specify the *service discipline*, i.e., the order of services. The most basic service discipline is first-come first-served (FCFS). Other service disciplines are also often considered, e.g., last-come first-served (LCFS), services in random order, and processor-sharing.

One of the most basic models of customer arrival processes is the Poisson arrival process, where arrivals of customers occur completely randomly in time. A Poisson process is characterized only by mean arrival rate λ ($\lambda > 0$), and its inter-arrival times are independent and identically distributed (i.i.d.) according to an exponential distribution with mean $1/\lambda$. Single-server queueing models with Poisson arrivals and generally distributed service times, denoted by M/G/1, have been widely studied and various results for M/G/1 queues can be found in the literature. For classical results of single-server queueing models including M/G/1 queues, readers are referred to [Coh82, Kle75].

While the occurrences of Poisson arrivals are assumed to be completely random, data traffic of current communication networks are usually bursty. Markovian arrival processes (MAPs) are an extension of the Poisson arrival process [LMN90], which can be used to model bursty arrivals. MAPs form a rich class of arrival processes, and they include Markov modulated Poisson processes (MMPPs) and phase-type renewal processes. In particular, MAPs have an important property that they are dense in the set of all stationary point processes [AK93]. In this sense, MAPs are one of the most general arrival processes.

In addition to the burstiness of arrivals, data traffic of current communication networks has another important property: it is *a superposition* of several packet streams with different properties such as text, audio, video, and control packets. Such a superposition of different arrival streams can be modeled using marked MAPs [He96], which consists of several arrival streams of MAP. As we will see below, the superposition of different arrival streams makes the behavior of queueing models essentially more complicated.

The single-server queue with MAP arrivals (MAP/G/1) and marked MAP arrivals (multi-class MAP/G/1) are formulated as queueing models with underlying processes, for which efficient computational algorithms for performance measures are known in the literature. In the next section, we briefly review these results as preliminaries to this dissertation.

1.2 Queueing models with underlying processes

1.2.1 MAP/G/1 queue [LMN90, Neu89]

Throughout this chapter, the service discipline is always assumed to be FCFS unless otherwise mentioned. We start with the definition of the MAP. Consider an irreducible continuous-time Markov chain with finite state space $\mathcal{M} = \{1, 2, \dots, M\}$. We refer to this Markov chain as the *underlying Markov chain*. The behavior of a MAP is governed by this Markov chain as follows. The underlying Markov chain stays in state i ($i \in \mathcal{M}$) for an exponential interval of time with mean $1/\sigma_i$ ($\sigma_i > 0$), and when the sojourn time in state i is elapsed,

- with probability $q_{i,j}$, an arrival of a customer occurs, and the underlying Markov chain changes its state to j ($j \in \mathcal{M}$),
- with probability $p_{i,j}$, no customer arrivals occur, and the underlying Markov chain changes its state to j ($j \in \mathcal{M}$, $j \neq i$).

It is assumed that

$$\sum_{j \in \mathcal{M}} (p_{i,j} + q_{i,j}) = 1, \qquad i \in \mathcal{M},$$

where $p_{i,i}$ is defined as zero. The MAP is usually represented using $M \times M$ matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ whose (i,j)-th ($i, j \in \mathcal{M}$) elements are given by

$$[\boldsymbol{C}]_{i,j} = \begin{cases} \sigma_i p_{i,j}, & i \neq j, \\ -\sigma_i, & i = j, \end{cases} \qquad [\boldsymbol{D}]_{i,j} = \sigma_i q_{i,j}.$$

Note that $\boldsymbol{C}$ and $\boldsymbol{D}$ denote a defective infinitesimal generator and a transition rate matrix, respectively, and they satisfy

$$(\boldsymbol{C} + \boldsymbol{D})\boldsymbol{e} = \boldsymbol{0}, \tag{1.1}$$

where $\boldsymbol{e}$ denotes an $M \times 1$ vector whose elements are all equal to one. It is readily verified that $\boldsymbol{C}+\boldsymbol{D}$ represents the irreducible infinitesimal generator of the underlying Markov chain. Therefore, $\boldsymbol{C}+\boldsymbol{D}$ has its invariant probability vector $\boldsymbol{\pi}$ uniquely determined by

$$\boldsymbol{\pi}(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{0}, \qquad \boldsymbol{\pi}\boldsymbol{e} = 1. \tag{1.2}$$

Let $\boldsymbol{N}(n,t)$ ($n = 0,1,\ldots$, $t \geq 0$) denote an $M \times M$ matrix whose (i,j)-th ($i,j \in \mathscr{M}$) element represents the probability that n customers arrive in time period $(0,t]$ and the state of the underlying Markov chain is equal to j at time t, given that the state of the underlying Markov chain is equal to i at time 0. We define $\boldsymbol{N}^*(z,t)$ ($|z| \leq 1$, $t \geq 0$) as

$$\boldsymbol{N}^*(z,t) = \sum_{n=0}^{\infty} \boldsymbol{N}(n,t) z^n.$$

Note that for $\Delta t \geq 0$, $\boldsymbol{N}^*(z,t)$ satisfies

$$\boldsymbol{N}^*(z,t+\Delta t) = \boldsymbol{N}^*(z,t)[\boldsymbol{I} + \boldsymbol{C}\Delta t + z\boldsymbol{D}\Delta t + \boldsymbol{o}(\Delta t)],$$

where $\boldsymbol{I}$ denotes a unit matrix. It is readily verified that this equation implies

$$\frac{\partial \boldsymbol{N}^*(z,t)}{\partial t} = \boldsymbol{N}^*(z,t)(\boldsymbol{C} + z\boldsymbol{D}).$$

Therefore, noting $\boldsymbol{N}^*(z,0) = \boldsymbol{I}$, we obtain

$$\boldsymbol{N}^*(z,t) = \exp[(\boldsymbol{C} + z\boldsymbol{D})t].$$

The MAP/G/1 queue is a single-server queueing model, where customers arrive according to a MAP, and service times are i.i.d. according to a general distribution. Let $H(x)$ ($x \geq 0$) denote the probability distribution function (PDF) of service times. Also let $\mathrm{E}[H]$ denote the mean service time. Because the mean arrival rate is given by $\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}$, the traffic intensity ρ of this model is given by

$$\rho = \boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e} \cdot \mathrm{E}[H].$$

The stability of the MAP/G/1 queue is ensured if ρ satisfies

$$\rho < 1. \tag{1.3}$$

Under the stability condition (1.3), performance measures of the system can be obtained through an analysis of an embedded queue length process obtained by observing the system only at departure time instants. Let $\boldsymbol{x}(n)$ ($n = 0,1,\ldots$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the stationary joint probability that the number of customers in the system is equal to n and the state of the underlying Markov chain is equal to j just after a customer departure. We can verify that $\boldsymbol{x}(n)$ is given by the stationary distribution of a positive-recurrent discrete-time Markov chain with transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.4}$$

where $\boldsymbol{A}_n$ (resp. $\boldsymbol{B}_n$) denotes an $M \times M$ matrix whose (i,j)-th $(i,j \in \mathscr{M})$ element represents the probability that n $(n = 0,1,\ldots)$ customers arrive and the state of the underlying Markov chain changes from i to j in the service time of a customer (resp. in time period from an instant that the system becomes empty, to the next departure of a customer). It is easy to see that

$$\sum_{n=0}^{\infty} \boldsymbol{A}_n z^n = \int_0^{\infty} \exp[(\boldsymbol{C} + z\boldsymbol{D})x]dH(x). \tag{1.5}$$

From this equation, the coefficient matrices $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots$ can be obtained with an algorithmic approach [TMSH94]. Furthermore, $\boldsymbol{B}_n$ is given by

$$\boldsymbol{B}_n = \int_0^{\infty} \exp[\boldsymbol{C}x]\boldsymbol{D}dx \cdot \boldsymbol{A}_n = (-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{A}_n, \qquad n = 0,1,\ldots.$$

Markov chains with the transition structure (1.4) are said to have *the skip-free to the left property*, and such Markov chains are called M/G/1-type Markov chains. For this class of Markov chains, an efficient computational algorithm for the stationary distributions can be found in the literature [Neu89], so that we can compute the stationary distribution $\boldsymbol{x}(n)$ $(n = 0,1,\ldots)$ of the queue length at departure instants.

Other performance measures in the MAP/G/1 queue are given in terms of $\boldsymbol{x}(n)$. Let $\boldsymbol{y}(0)$ denote a $1 \times M$ vector whose j-th $(j \in \mathscr{M})$ element represents the stationary probability that the system is empty and the state of the underlying Markov chain is equal to j. Further let $\boldsymbol{y}(n,t)$ $(t \geq 0,\ n = 1,2,\ldots)$ denote a $1 \times M$ vector whose j-th $(j \in \mathscr{M})$ element represents the joint probability that the number of customers in the stationary system is equal to n, the remaining service time of the customer being served is not greater than t, and the state of the underlying Markov chain is equal to j. We define a joint transform $\boldsymbol{y}^*(z,s)$ $(|z| \leq 1,\ \mathrm{Re}(s) > 0)$ as

$$\boldsymbol{y}^*(z,s) = \boldsymbol{y}(0) + \sum_{n=1}^{\infty} z^n \int_{t=0}^{\infty} \exp[-st]d\boldsymbol{y}(n,t).$$

It can be verified that $\boldsymbol{y}^*(z,s)$ is given by

$$\boldsymbol{y}^*(z,s) = (1-\rho) \cdot \frac{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}}{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{e}} + \rho\Big(\boldsymbol{x}^*(z) - \boldsymbol{x}(0) + z\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{D}\Big)\tilde{\boldsymbol{A}}^*(z,s), \tag{1.6}$$

where

$$\tilde{\boldsymbol{A}}^*(z,s) = \int_0^{\infty} \frac{xdH(x)}{\mathrm{E}[H]} \int_0^x \frac{dt}{x} \cdot \exp[-s(x-t)]\exp[(\boldsymbol{C} + z\boldsymbol{D})t].$$

Let $\boldsymbol{y}^*(z)$ $(|z| \leq 1)$ denote the probability generating function of $\boldsymbol{y}(n)$.

$$\boldsymbol{y}^*(z) = \sum_{n=0}^{\infty} \boldsymbol{y}(n)z^n = \lim_{s \to 0+} \boldsymbol{y}^*(z,s).$$

Also let $\tilde{\boldsymbol{A}}(z)$ ($|z| \geq 0$) denote an $M \times M$ matrix given by

$$\begin{aligned}\tilde{\boldsymbol{A}}(z) = \lim_{s \to 0+} \tilde{\boldsymbol{A}}^*(z,s) &= \int_0^\infty \frac{dH(x)}{\mathrm{E}[H]} \int_0^x \exp[(\boldsymbol{C} + z\boldsymbol{D})t]dt \\ &= \int_0^\infty \exp[(\boldsymbol{C} + z\boldsymbol{D})t]dt \int_t^\infty \frac{dH(x)}{\mathrm{E}[H]} \\ &= \int_0^\infty \exp[(\boldsymbol{C} + z\boldsymbol{D})t]\tilde{h}(t)dt,\end{aligned}$$

where $\tilde{h}(x)$ ($x \geq 0$) denotes the probability density function (p.d.f.) of the equilibrium distribution of service times. Note here that $\tilde{\boldsymbol{A}}(z)$ takes the form

$$\tilde{\boldsymbol{A}}(z) = \sum_{n=0}^{\infty} \tilde{\boldsymbol{A}}_n z^n,$$

where the coefficient matrices $\tilde{\boldsymbol{A}}_0, \tilde{\boldsymbol{A}}_1, \ldots$ can be calculated in the same way as $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots$ (cf. (1.5)). It then follows from (1.6) that

$$\boldsymbol{y}^*(z) = (1-\rho) \cdot \frac{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}}{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{e}} + \rho\Big(z\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{D} + \boldsymbol{x}^*(z) - \boldsymbol{x}(0)\Big)\tilde{\boldsymbol{A}}^*(z),$$

from which we obtain

$$\begin{aligned}\boldsymbol{y}(0) &= (1-\rho) \cdot \frac{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}}{\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{e}}, \\ \boldsymbol{y}(n) &= \rho\Big(\boldsymbol{x}(0)(-\boldsymbol{C})^{-1}\boldsymbol{D}\tilde{\boldsymbol{A}}_{n-1} + \sum_{i=1}^{n} \boldsymbol{x}(i)\tilde{\boldsymbol{A}}_{n-i}\Big), \qquad n = 1,2,\ldots.\end{aligned}$$

$\boldsymbol{y}(n)$ is thus given in terms of $\boldsymbol{x}(n)$.

The actual waiting time and the sojourn time are also analyzed based on (1.6). Let $\boldsymbol{u}(x)$ ($x \geq 0$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the joint probability that the workload in system is not greater than x and the underlying Markov chain is in state j in steady state. We define $\boldsymbol{u}^*(s)$ (Re(s) > 0) as the Laplace-Stieltjes transform (LST) of $\boldsymbol{u}(x)$.

$$\boldsymbol{u}^*(s) = \int_0^\infty \exp[-sx]d\boldsymbol{u}(x).$$

We then have from (1.6),

$$\boldsymbol{u}^*(s) = \boldsymbol{y}(0) + \frac{\boldsymbol{y}^*(h^*(s),s) - \boldsymbol{y}(0)}{h^*(s)},$$

where $h^*(s)$ (Re(s) > 0) denotes the LST of the service time distribution. After some calculations based on this equation, it can be verified that $\boldsymbol{u}^*(s)$ (Re(s) > 0) satisfies

$$\boldsymbol{u}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + h^*(s)\boldsymbol{D}] = \boldsymbol{y}(0)s. \tag{1.7}$$

We define $w^*(s)$ and $\overline{w}^*(s)$ (Re$(s) > 0$) as the LSTs of the actual waiting time and the sojourn time distributions, respectively. It is easy to verify that $w^*(s)$ and $\overline{w}^*(s)$ are given in terms of $\boldsymbol{u}^*(s)$ by

$$w^*(s) = \frac{\boldsymbol{u}^*(s)\boldsymbol{De}}{\boldsymbol{\pi De}}, \qquad \overline{w}^*(s) = \frac{\boldsymbol{u}^*(s)\boldsymbol{De}}{\boldsymbol{\pi De}} \cdot h^*(s). \tag{1.8}$$

With a straightforward calculations based on (1.7) and (1.8), we can obtain a computational algorithm for the moments of these distributions [TH94].

1.2.2 Multi-class MAP/G/1 queue [He96, TH94, Tak01, Tak05]

We next consider an extension of the single-class MAP/G/1 queue described in Section 1.2.1 to a multi-class queueing model. We start with an extension of the MAP, called the marked MAP, where K ($K = 1,2,\ldots$) classes of customers exist. Let $\mathscr{K} = \{1,2,\ldots,K\}$ denote the set of customer classes. The arrival process of class k customers is assumed to follow a MAP $(\boldsymbol{C},\boldsymbol{D}_k)$ with the same state space and a common generator $\boldsymbol{C}$. Also, service times of class k ($k \in \mathscr{K}$) customers are assumed to be i.i.d. with PDF $H_k(x)$ ($x \ge 0$).

If all customer classes have the same service time distribution $H(x)$, i.e., $H_k(x) = H(x)$ ($k \in \mathscr{K}$), this model reduces to a single-class MAP/G/1 queue with

$$\boldsymbol{D} = \sum_{k\in\mathscr{K}} \boldsymbol{D}_k.$$

However, if service times depend on the customer class, this relation no longer holds because service times and the state of the underlying Markov chain are dependent in general, so that service times are not i.i.d. To deal with the multi-class MAP/G/1 queue with different service time distributions among classes, it is useful to introduce a more general model described as follows.

Similarly to the case of the MAP, consider an irreducible underlying Markov chain with finite state space $\mathscr{M} = \{1,2,\ldots,M\}$. There are K ($K = 1,2,\ldots$) classes of customers, and the set of customer classes are denoted by $\mathscr{K} = \{1,2,\ldots,K\}$. The underlying Markov chain stays in state i ($i \in \mathscr{M}$) for an exponential interval of time with mean $1/\sigma_i$ ($\sigma_i > 0$), and when the sojourn time in state i is elapsed,

- with probability $p_{i,j}$, it changes its state to j ($j \in \mathscr{M}$, $j \neq i$) without customer arrivals,
- with probability $q_{k,i,j}$, it changes its state to j ($j \in \mathscr{M}$) and a class k ($k \in \mathscr{K}$) customer arrives.

It is assumed that

$$\sum_{j\in\mathscr{M}} \left(p_{i,j} + \sum_{k\in\mathscr{K}} q_{k,i,j} \right) = 1, \qquad i \in \mathscr{M},$$

where $p_{i,i}$ is defined as zero. Furthermore, service times of class k customers who arrive with state transitions from i to j are assumed to be i.i.d. according to a general distribution with PDF $H_{k,i,j}(x)$ ($x \geq 0$, $i,j \in \mathcal{M}$, $k \in \mathcal{K}$).

We then define $\boldsymbol{C}$ and $\boldsymbol{D}_k(x)$ ($x \geq 0$, $k \in \mathcal{K}$) as $M \times M$ matrices whose (i,j)-th ($i,j \in \mathcal{M}$) elements are given by

$$[\boldsymbol{C}]_{i,j} = \begin{cases} \sigma_i p_{i,j}, & i \neq j, \\ -\sigma_i, & i = j, \end{cases} \qquad [\boldsymbol{D}_k(x)]_{i,j} = \sigma_i q_{k,i,j} H_{k,i,j}(x).$$

We further define $\boldsymbol{D}(x)$ ($x \geq 0$) as

$$\boldsymbol{D}(x) = \sum_{k \in \mathcal{K}} \boldsymbol{D}_k(x).$$

Throughout this dissertation, we refer to a single-server queue with this general arrival stream as the multi-class MAP/G/1 queue. We can readily verify that this model includes the MAP/G/1 and the marked MAP/G/1 queues as special cases.

We define $\boldsymbol{D}^*(s)$ ($\mathrm{Re}(s) > 0$) and $\boldsymbol{D}$ as

$$\boldsymbol{D}^*(s) = \int_0^\infty \exp[-sx] d\boldsymbol{D}(x), \qquad \boldsymbol{D} = \lim_{x \to \infty} \boldsymbol{D}(x) = \lim_{s \to 0+} \boldsymbol{D}^*(s).$$

Because the underlying Markov chain is assumed to be irreducible, $\boldsymbol{C} + \boldsymbol{D}$ satisfies (1.1), and it has the unique invariant probability vector $\boldsymbol{\pi}$ determined by (1.2). The traffic intensity ρ of this model is given by

$$\rho = \boldsymbol{\pi} \int_0^\infty x d\boldsymbol{D}(x) \boldsymbol{e},$$

and the system is stable if $\rho < 1$. Below we assume that the system is stable.

Unlike the single-class MAP/G/1 queue, the queue length process in the multi-class MAP/G/1 queue is not easy to analyze directly, because it is necessary to keep track of the classes of all waiting customers to construct the embedded Markov chain at customer departures, which results in an explosion of the state space. On the other hand, the workload process in the multi-class MAP/G/1 queue is easy to analyze. From a balance equation for the stationary workload distribution $\boldsymbol{u}(x)$, we can verify that its LST $\boldsymbol{u}^*(s)$ satisfies (cf. (1.7))

$$\boldsymbol{u}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = (1-\rho)\boldsymbol{\kappa}, \tag{1.9}$$

where $\boldsymbol{\kappa}$ denotes a $1 \times M$ vector whose j-th element ($j \in \mathcal{M}$) represents the conditional probability that the underlying Markov chain is in state j given that the system is empty. An effective computational algorithm for $\boldsymbol{\kappa}$ is reported in the literature [TH94].

Let $w_k^*(s)$ and $\overline{w}_k^*(s)$ ($k \in \mathscr{K}$, $\mathrm{Re}(s) > 0$) denote the LSTs of the actual waiting time and the sojourn time distributions of class k customers, respectively. Similarly to (1.8), they are given in terms of $\boldsymbol{u}^*(s)$ by

$$w_k^*(s) = \frac{\boldsymbol{u}^*(s)\boldsymbol{D}_k\boldsymbol{e}}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}}, \qquad \overline{w}_k^*(s) = \frac{\boldsymbol{u}^*(s)\boldsymbol{D}_k^*(s)\boldsymbol{e}}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}},$$

where

$$\boldsymbol{D}_k^*(s) = \int_0^\infty \exp[-sx] d\boldsymbol{D}_k(x), \qquad \boldsymbol{D}_k = \lim_{x\to\infty} \boldsymbol{D}_k(x) = \lim_{s\to 0+} \boldsymbol{D}_k^*(s).$$

Furthermore, a computational algorithm for the joint queue length distribution of respective classes can be derived based on (1.9), as shown in [Tak01].

In [Tak05], the multi-class MAP/G/1 queue is further extended to a model where the service speed of the server varies depending on the state of the underlying Markov chain. [Tak05] shows that the analysis of this model is reduced to that of a multi-class MAP/G/1 queue with constant service rate, through the following observation: by extending the time axis of a sample path of the original model so that the service rate is constant, we obtain a sample path equivalent to that of an ordinary multi-class MAP/G/1 queue.

1.3 Motivation of this study

Presently, adaptive resource allocation mechanisms are being developed in each field of communication technology. For example, dynamic spectrum allocation in cognitive radio, reconfigurable wavelength division multiplexing (WDM), and dynamic route control based on the software-defined network (SDN) technology are regarded as adaptive resource allocation mechanisms. To make efficient use of bandwidths, these technologies dynamically change the allocation of bandwidths and restrict the volume of incoming traffic through admission control.

Mathematically, systems with adaptive resource allocation mechanisms can be formulated as queueing models *interacting with* underlying processes, where states of the queues such as the queue lengths and the workloads have an effect on state transitions of the underlying processes. Because the ordinary analytic methods reviewed in the previous section assume that the dynamics of the underlying process is invariant with respect to the state of the queue, they cannot be applied directly to these systems. Note that there are some recent studies on computational algorithms of the stationary distribution of Markov chains with *level-dependent* transition structures [BS12, Mas15, PMKT10] that can formulate a wide class of queueing models interacting with underlying processes. However, they deal with only Markov *chains*, i.e., the case that the state space is countable, which corresponds to the queue length process in queueing models. Therefore, it is difficult to apply them to: (i) multi-class queueing models, where we need to analyze the workload process, and (ii) queueing models where underlying processes interact with the workload processes (or wating times) not the queue length processes.

In this study, we consider two types of queueing models interacting with underlying processes, which are regarded as fundamental models for communication systems with adaptive resource allocation mechanisms. In the first model, we assume that the state of the queue affect the underlying state only when the system becomes empty. We assume that the underlying process is a continuous-time Markov chain with finite state-space similarly to the ordinary MAP/G/1 queue, and its state is switched at the instant that the system becomes empty. This corresponds to a system where bandwidth is added when congestion continues for a while, and the additional bandwidth is released when the congestion gets relieved.

On the other hand, the second model assumes that the state of the underlying process changes *continually* according to the workload in system. Specifically, when the workload in system is equal to x ($x \geq 0$), the arrival rate of customers is reduced from λ to $\lambda\overline{G}(x)$, where $\overline{G}(x)$ denotes a non-increasing function taking value in $[0,1]$. This model corresponds to a system with real-time traffic control such as an adaptive admission control. Note that the first model describes relatively long-term changes in the state of the underlying process, while the second model describes short-term changes in it.

In this dissertation, we develop analytical methods for these two fundamental models through an analysis of closely related queueing models: queueing models with working vacations and queueing models with impatient customers. As we will see, queueing models with working vacations (resp. impatient customers) has a similar mathematical structure to the first (resp. second) fundamental model introduced above. In the rest of this section, we review these models more specifically, discussing the relation between those and our two fundamental models.

1.3.1 Queueing models with working vacations

In queueing models with working vacations, the server takes a period called working vacation when the system becomes empty. During working vacation periods, arriving customers are served at processing rate σ ($\sigma > 0$), which may differ from the normal processing rate of one. This model is an extension of ordinary vacations models, where the server does not serve customers during vacation periods, i.e., $\sigma = 0$. In what follows, time intervals during which customers are served at processing rate one are referred to as *normal service periods*.

The queueing model with working vacations was first introduced in [SF02], as a model of an access router in a reconfigurable WDM optical access network. While each access router has its own wavelength, there are some additional wavelengths that are shared among several access routers, and those additional wavelengths are assigned to those access routers cyclically. A working vacation period then corresponds to the situation that the access router has no additional wavelengths, and the following normal service periods corresponds to the situation that the access router utilizes the additional wavelengths as well.

[SF02] studies an M/M/1 queue with exponential working vacations, where service times are assumed to be i.i.d. according to an exponential distribution. Furthermore, in [KCC03, LTZL09, WT06], the model of [SF02] is generalized to the M/G/1 queue. These models can be viewed as special cases of our first model of the two fundamental models introduced above, regarding that they have two underlying states {WV, NP}, where WV and NP denote working vacation period and normal service period, respectively. Therefore, to make analytical methods applicable to the general case of our first model, it is necessary to extend the state space of the underlying process to N-states.

1.3.2 Queueing models with impatient customers

In queueing models with impatient customers, each customer has his/her own maximum allowable waiting time, which is referred to as the impatience time. If elapsed waiting times of customers reach their impatience times, they leave the system immediately without receiving their services. Our fundamental model of the second kind is equivalent to the M/G/1 queue with impatient customers and generally distributed impatience times, which is usually denoted by M/G/1+G. The equivalence between these models can be verified as follows. Note first that we define the workload in the M/G/1+G queue as the sum of the remaining service times of all customers in the system *who receive their services eventually*. Therefore, customers who leave the system without receiving their services do not contribute to the workload in system. The workload process in the M/G/1+G queue, where impatience times have a complementary PDF $\overline{G}(x)$, is thus equivalent to that in our second model, because the arrival rate λ is essentially reduced to $\lambda\overline{G}(x)$ when the workload in system is equal to x.

Queueing models with impatient customers have been studied for a long time, and a large number of research papers have been published. The numbers of customers in the M/M/s+D, M/M/1+M, and M/G/1+M queues are studied in [Bar57], [AG62], and [Rao67], respectively. [Kov61] shows that the p.d.f. $v(x)$ of the workload distribution satisfies a Volterra integral equation of the second kind. [Dal65] studies the actual waiting time in the GI/G/1+G queue and derives the actual waiting time distributions in the M/G/1+D and M/G/1+M queues and the workload distribution in the M/G/1+M queue. [Coh69] studies the workload in the GI/G/1+D queue and it shows the result for the M/G/1+D queue as a special case. In [Coh69], the workload in the M/G/1+M queue is also analyzed. [Sta79] studies the GI/G/1+G queue, where the stationary distributions of the actual waiting time and the number of customers in the system are related to the stationary workload distribution. Furthermore, in [Sta79], some special cases are also discussed, which include the stationary workload distribution in the M/M/1+G queue. The GI/G/1+G queue is also studied in [BH81] and [BBH84], and the p.d.f. $v(x)$ of the stationary workload in the M/G/1+G queue is derived as a special case, reproducing the result in [Kov61]. In addition, a formula for the M/G/1+Er queue is obtained. In recent papers, some special cases of the M/G/1+G queue are discussed, where $v(x)$ has a simpler expression than the general case, e.g., the M/G/1+D queue [BKL01], the M/PH/1+D queue [KK13], the M/G/1+PH queue [BB13], and the M/G/1 queue with discrete impatience times [Bae13]. Moreover, there exist further works which analyze the busy period [BPS11, BPSZ10, BB13, KBL01, LK08, PA95, PSZ00] and the joint distribution of the residual/original waiting times of all waiting customers [BB99].

The fundamental quantity of interest in queues with impatient customers is the stationary loss probability P_{loss}, i.e., the probability that a randomly chosen customer leaves the system without receiving his/her service. P_{loss} is also the most important quantity in our second model, where P_{loss} represents the ratio of rejected arrivals. To the best of our knowledge, however, any further results for P_{loss} in the M/G/1+G queue beyond those of [BBH84, Kov61] are not found in the literature. As we will see, under the formulation of [BBH84, Kov61], P_{loss} is given in terms of the sequence of recursively determined functions, so that it is not easy to evaluate the impacts of the parameters of the model on P_{loss}.

1.4 Overview of the dissertation

The rest of this dissertation is organized into 7 chapters. Chapters 2, 3, and 4 are related to our first model, while Chapters 5, 6, and 7 are related to our second model.

In Chapter 2, we analyze the multi-class FCFS M/G/1 queue with working vacations. Past studies on the single-class M/G/1 queues with working vacations [KCC03, LTZL09, WT06] take an approach that the queue length process is analyzed first, and other performance measures are derived using the queue length distribution. Note that queue length processes in multi-class FCFS queues are not easy to analyze directly as mentioned in Section 1.2.2, so that the approach of [KCC03, LTZL09, WT06] cannot be applied to our model considered in Chapter 2. Furthermore, to make the analysis of the queue length simple, those studies assume the preemptive-repeat with resampling when working vacations end, i.e., the server always restarts the ongoing service at the beginning of normal service periods, where the new service time is resampled according to the service time distribution. In Chapter 2, we present an analytical method for the multi-class M/G/1 queue with working vacations, where the server continues the ongoing service at the beginning of a normal service period in a preemptive-resume manner. We further generalize the conventional model with working vacations assuming that the arrival rate in the working vacation and normal service periods may be different.

Our analysis takes an approach different from those in the past studies. We first analyze the workload process, and other performance measures are derived using the workload distribution, similarly to the analysis of the ordinary multi-class MAP/G/1 queue without working vacations.

As we will see in Chapter 2, analyzing queues with working vacations are essentially the same as analyzing the corresponding queueing models with disasters. We thus focus our attention on queues with disasters in Chapter 3, and extend the approach based on the workload distribution shown in Chapter 2 into the multi-class MAP/G/1 queue with disasters. The model discussed in Chapter 3 is considered as a generalization of the single-class M/G/1 queue with disasters studied in [JS96], where both the customer arrival process and the disaster occurrence process are assumed to follow Poisson processes. [DN99, Shi04] generalize the model of [JS96] such that customers arrive according to a batch MAP (BMAP) and disasters occur according to a MAP. While [JS96] analyzes the workload distribution directly, [DN99, Shi04] takes an approach that the queue length distribution is analyzed first. Note again that the multi-class model considered in Chapter 3 cannot be analyzed with the approach of [DN99, Shi04] based on the queue length distribution.

In Chapter 4, we generalize the results in Chapters 2 and 3 by considering a continuous-time bivariate Markov process $\{(U(t),S(t)); t \geq 0\}$, where $U(t)$ and $S(t)$ are referred to as the level and the phase, respectively, at time t. $U(t)$ $(t \geq 0)$ takes values in $[0,\infty)$ and $S(t)$ $(t \geq 0)$ takes values in a finite set $\mathscr{M} = \{1,2,\ldots,M\}$. $U(t)$ and $S(t)$ correspond to the workload in system and the state of the underlying process in a queueing model. $U(t)$ $(t \geq 0)$ is assumed to be skip-free to the left, and therefore we call it the M/G/1-type Markov *process*. The M/G/1-type Markov process was first introduced in [Tak96] as a generalization of the workload process in the MAP/G/1 queue and its stationary distribution was analyzed under a strong assumption that the conditional infinitesimal generator of the underlying Markov chain $S(t)$ given $U(t) > 0$ is irreducible. In Chapter 4, we extend known results for the stationary distribution to the case that the conditional infinitesimal generator of the underlying Markov chain given $U(t) > 0$ is reducible. This extension provides a unified way to analyze our fundamental model of the first kind.

In Chapter 5, we revisit the formula for the p.d.f. $v(x)$ of the workload distribution in the M/G/1+G queue obtained in [Kov61], and provide a new perspective on it. More specifically, we consider a last-come first-served preemptive-resume (LCFS-PR) M/G/1 queue with workload dependent loss, whose workload process is identical to that of the M/G/1+G queue. Through an analysis of the model, we show that $v(x)$ can be interpreted as the p.d.f. of a random sum of dependent random variables. As we will see, this new perspective leads to a unified understanding of special cases of the M/G/1+G queue.

Chapter 6 analyzes the stationary loss probability P_{loss} in the M/G/1+G queue. Based on the results in Chapter 5, various properties of the loss probability such as theoretical lower and upper bounds and stochastic ordering relations are derived.

In particular, as a consequence of the stochastic ordering relations, we show a theoretically interesting result that the loss probability P_{loss} in the M/D/1+D queue is smallest among all M/G/1+G queues with the same and finite arrival rate, mean service time, and mean impatience time.

In Chapter 7, a computational algorithm for P_{loss} in the M/G/1+PH queue is developed, where impatience times are assumed to follow a phase-type distribution. The phase-type distribution is a probability distribution defined as the length of absorbing time of a continuous-time absorbing Markov chain, which has suitable properties for numerical computation. The set of phase-type distributions is known to be dense in the class of all non-negative probability distributions [Neu89], so that it is one of the most general probability distribution function. To the best of our knowledge, the M/G/1+PH queue has been studied only in [BB13]. Unfortunately, formulas obtained in [BB13] are not suitable for numerical computation because if we followed it, we would have to deal with exponentially growing number of terms. We develop a computational algorithm for P_{loss} based on the uniformization technique [Tij94, Page 154] and the results in Chapters 5 and 6. With this algorithm, we can effectively compute P_{loss} without being involved with the exponentially growing number of terms. Moreover, this computational algorithm has a remarkable feature that it also outputs an upper bound of its numerical error.

Finally, we conclude this dissertation in Chapter 8.

2 Multi-Class M/G/1 Queue with Working Vacations

2.1 Introduction

In this chapter, we consider a stationary multi-class FCFS M/G/1 queue with exponential working vacations. When the system becomes empty, the server takes a working vacation, during which customers are served at processing rate σ ($\sigma > 0$). If the system is empty at the end of the working vacation, the server takes another working vacation. On the other hand, if a customer is being served at the end of the working vacation, the server switches its processing rate to one and continues to serve customers in a *preemptive-resume* manner, until the system becomes empty. We assume that lengths of working vacations are i.i.d. according to an exponential distribution with parameter γ ($\gamma > 0$). Let V denote a generic random variable for lengths of working vacations.

$$\Pr(V \le x) = 1 - \exp[-\gamma x], \qquad x \ge 0.$$

There are K classes of customers, labeled one to K. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ denote the set of customer classes. During working vacation periods (resp. normal service periods), class k ($k \in \mathcal{K}$) customers arrive according to a Poisson process with rate $\lambda_{\mathrm{WV},k}$ (resp. $\lambda_{\mathrm{NP},k}$). Let λ_{WV} and λ_{NP} denote the total arrival rates during working vacation periods and during normal service periods, respectively.

$$\lambda_{\mathrm{WV}} = \sum_{k \in \mathcal{K}} \lambda_{\mathrm{WV},k}, \qquad \lambda_{\mathrm{NP}} = \sum_{k \in \mathcal{K}} \lambda_{\mathrm{NP},k},$$

where we assume $\lambda_{\mathrm{WV}} > 0$ to avoid trivialities. The amounts of service requirements of class k ($k \in \mathcal{K}$) customers who arrive in working vacation periods (resp. normal service periods) are assumed to be i.i.d. according to a general distribution with PDF $H_{\mathrm{WV},k}(x)$ (resp. $H_{\mathrm{NP},k}(x)$). For each k ($k \in \mathcal{K}$), let $H_{\mathrm{WV},k}$ and $H_{\mathrm{NP},k}$ denote generic random variables with PDFs $H_{\mathrm{WV},k}(x)$ and $H_{\mathrm{NP},k}(x)$, respectively.

$$\Pr(H_{\mathrm{WV},k} \le x) = H_{\mathrm{WV},k}(x), \quad x \ge 0, \qquad \Pr(H_{\mathrm{NP},k} \le x) = H_{\mathrm{NP},k}(x), \quad x \ge 0.$$

Let H_{WV} (resp. H_{NP}) denote a random variable representing the amount of the service requirement brought by a customer randomly chosen among those arriving in working vacation periods (resp. normal service periods). We define $h^*_{\mathrm{WV}}(s)$ and $h^*_{\mathrm{NP}}(s)$ ($\mathrm{Re}(s) > 0$) as the LSTs of H_{WV} and H_{NP}, respectively.

$$h^*_{\mathrm{WV}}(s) = \sum_{k \in \mathcal{K}} \frac{\lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}} \cdot h^*_{\mathrm{WV},k}(s), \qquad h^*_{\mathrm{NP}}(s) = \sum_{k \in \mathcal{K}} \frac{\lambda_{\mathrm{NP},k}}{\lambda_{\mathrm{NP}}} \cdot h^*_{\mathrm{NP},k}(s),$$

where $h^*_{\mathrm{WV},k}(s)$ and $h^*_{\mathrm{NP},k}(s)$ ($\mathrm{Re}(s) > 0$) denote the LSTs of $H_{\mathrm{WV},k}$ and $H_{\mathrm{NP},k}$, respectively.

$$h^*_{\mathrm{WV},k}(s) = \int_0^\infty \exp[-sx] dH_{\mathrm{WV},k}(x), \qquad h^*_{\mathrm{NP},k}(s) = \int_0^\infty \exp[-sx] dH_{\mathrm{NP},k}(x).$$

We define $\rho_{\mathrm{WV},k}$ ($k \in \mathcal{K}$), $\rho_{\mathrm{NP},k}$ ($k \in \mathcal{K}$), ρ_{WV}, and ρ_{NP} as

$$\begin{aligned} \rho_{\mathrm{WV},k} &= \lambda_{\mathrm{WV},k} \mathrm{E}[H_{\mathrm{WV},k}], \qquad \rho_{\mathrm{NP},k} = \lambda_{\mathrm{NP},k} \mathrm{E}[H_{\mathrm{NP},k}], \\ \rho_{\mathrm{WV}} &= \sum_{k \in \mathcal{K}} \rho_{\mathrm{WV},k}, \qquad \rho_{\mathrm{NP}} = \sum_{k \in \mathcal{K}} \rho_{\mathrm{NP},k}. \end{aligned}$$

In what follows, we assume $\rho_{\mathrm{NP}} < 1$. The service discipline is assumed to be FCFS, unless otherwise mentioned, and services are nonpreemptive.

Remark 2.1. *When $\gamma > 0$, the system is stable if and only if $\rho_{\mathrm{WV}} < \infty$ and $\rho_{\mathrm{NP}} < 1$. To see this, consider the length $\hat{\Phi}$ of an interval between successive starts of working vacations. Note that the system is stable if and only if $\mathrm{E}[\hat{\Phi}] < \infty$. By definition, $\hat{\Phi}$ can be divided into two parts, one of which is the length of a working vacation period $\hat{\Phi}_{\mathrm{WV}}$ with mean $1/\gamma$ and the other is the length of the following normal service period $\hat{\Phi}_{\mathrm{NP}}$. Let $U^{\mathrm{E}}_{\mathrm{WV}}$ denote the total workload in system at the end of the working vacation period. If $\rho_{\mathrm{WV}} < \infty$ and $\rho_{\mathrm{NP}} < 1$, the stability of the system is ensured because $\mathrm{E}[\hat{\Phi}_{\mathrm{NP}}] = \mathrm{E}[U^{\mathrm{E}}_{\mathrm{WV}}]/(1 - \rho_{\mathrm{NP}})$ and*

$$\mathrm{E}[\hat{\Phi}] = \frac{1}{\gamma} + \frac{\mathrm{E}[U^{\mathrm{E}}_{\mathrm{WV}}]}{1 - \rho_{\mathrm{NP}}} \le \frac{1}{\gamma} + \frac{\rho_{\mathrm{WV}}/\gamma}{1 - \rho_{\mathrm{NP}}} < \infty,$$

where the first inequality comes from the fact that in every sample path, $U^{\mathrm{E}}_{\mathrm{WV}}$ is bounded above by the total workload brought in the working vacation period.

Conversely, if the system is stable, $\mathrm{E}[U^{\mathrm{E}}_{\mathrm{WV}}] < \infty$ holds, and therefore $\rho_{\mathrm{WV}} < \infty$. Furthermore, in an ordinary $M/G/1$ queue, the first passage time to the idle state with finite initial workload is finite if and only if the traffic intensity is less than one. Therefore, we have $\rho_{\mathrm{WV}} < \infty$ and $\rho_{\mathrm{NP}} < 1$ if the system is stable.

Remark 2.2. *If we ignore customer classes, the above model is reduced to a single-class $M/G/1$ with exponential working vacations characterized by arrival rates λ_{WV} and λ_{NP}, amounts of service requirements H_{WV} and H_{NP}, processing rate σ during working vacation periods, and exponential lengths of working vacation periods with mean $1/\gamma$.*

We first analyze the stationary workload in system and obtain its LST. Using this result, we derive the joint LST of the attained waiting time [Sen89] and the remaining service requirement in terms of the LST of the workload. Because the server has two different processing rates, the analysis of the attained waiting time distribution in our model is not as simple as in queues without working vacations [BT03, Tak01]. This also makes the joint LST of the attained waiting time and the remaining service requirement complicated. We classify the attained waiting time into several cases, so that the formula for the joint LST of the attained waiting time and the remaining service requirement is given in a comprehensible form.

Note that all *waiting* customers in the FCFS system arrived in the attained waiting time [BT03, Tak01]. Based on this observation, we obtain the joint transform of the queue lengths and the the workloads in system in respective classes. We also derive the LSTs of the stationary distributions of waiting time and sojourn time and the joint transform of the length of a randomly chosen busy cycle and the number of customers served in the cycle.

Owing to the independent and stationary increment of Poisson arrival processes, the stationary system behavior conditioned that the server is on working vacation is equivalent to that in the corresponding queue with disasters. Therefore, as by-products, we also obtain various formulas for the multi-class FCFS M/G/1 queue with Poisson disasters, which are generalized into the multi-class FCFS MAP/G/1 queue with disasters in Chapter 3.

The rest of this chapter is organized as follows. In Section 2.2, the stationary workload in system is analyzed. In Section 2.3, the actual waiting time and sojourn time distributions are analyzed. In Section 2.4, we study the joint distribution of the numbers of customers and the workloads in system in respective classes. In Section 2.5, we analyze the busy cycle. Finally, we conclude this chapter in Section 2.6.

2.2 Total workload in system

In this section, we discuss the total workload in system in steady state. Let U denote the total workload in system. We define U_{WV} (resp. U_{NP}) as the conditional workload in system given the server being on working vacation (resp. being in a normal service period). Let $u^*(s)$, $u^*_{\mathrm{WV}}(s)$, and $u^*_{\mathrm{NP}}(s)$ denote the LSTs of U, U_{WV}, and U_{NP}, respectively. We then have

$$u^*(s) = P_{\mathrm{WV}} \cdot u^*_{\mathrm{WV}}(s) + P_{\mathrm{NP}} \cdot u^*_{\mathrm{NP}}(s), \tag{2.1}$$

where P_{WV} (resp. P_{NP}) denotes the time-average probability of the server being on working vacation (resp. being in a normal service period).

Let $U^{\mathrm{E}}_{\mathrm{WV}}$ denote the total workload in system at the end of a working vacation. We denote the LST of $U^{\mathrm{E}}_{\mathrm{WV}}$ by $u^*_{\mathrm{WV,E}}(s)$. Consider a censored workload process by removing all normal service periods. In the resulting process, the ends of working vacations occur according to a Poisson process with rate γ. Therefore, owing to PASTA [Wol82], we have

$$u^*_{\mathrm{WV,E}}(s) = u^*_{\mathrm{WV}}(s), \qquad \mathrm{E}[U^{\mathrm{E}}_{\mathrm{WV}}] = \mathrm{E}[U_{\mathrm{WV}}]. \tag{2.2}$$

We then have the following two lemmas, whose proofs are given in Appendices 2.A and 2.B, respectively.

Lemma 2.1. *$u^*_{\mathrm{NP}}(s)$ is given by*

$$u^*_{\mathrm{NP}}(s) = \frac{1-u^*_{\mathrm{WV}}(s)}{s\mathrm{E}[U_{\mathrm{WV}}]} \cdot u^*_{\mathrm{M/G/1}}(s), \tag{2.3}$$

where $u^*_{\mathrm{M/G/1}}(s)$ denotes the LST of the workload in system in an ordinary $M/G/1$ queue and it is given by

$$u^*_{\mathrm{M/G/1}}(s) = \frac{(1-\rho_{\mathrm{NP}})s}{s-\lambda_{\mathrm{NP}}+\lambda_{\mathrm{NP}}h^*_{\mathrm{NP}}(s)}. \tag{2.4}$$

Lemma 2.2. *P_{WV} and P_{NP} are given by*

$$P_{\mathrm{WV}} = \frac{1-\rho_{\mathrm{NP}}}{1-\rho_{\mathrm{NP}}+\gamma\mathrm{E}[U_{\mathrm{WV}}]}, \qquad P_{\mathrm{NP}} = \frac{\gamma\mathrm{E}[U_{\mathrm{WV}}]}{1-\rho_{\mathrm{NP}}+\gamma\mathrm{E}[U_{\mathrm{WV}}]}, \tag{2.5}$$

respectively.

With Lemma 2.1, $u^*(s)$ is given in terms of $u^*_{\mathrm{WV}}(s)$ and $\mathrm{E}[U_{\mathrm{WV}}]$.

$$u^*(s) = P_{\mathrm{WV}} \cdot u^*_{\mathrm{WV}}(s) + P_{\mathrm{NP}} \cdot \frac{1-u^*_{\mathrm{WV}}(s)}{s\mathrm{E}[U_{\mathrm{WV}}]} \cdot u^*_{\mathrm{M/G/1}}(s), \tag{2.6}$$

where P_{WV} and P_{NP} are given in (2.5).

We now characterize $u^*_{\mathrm{WV}}(s)$. Note that the conditional workload U_{WV} given the server being on working vacation is equivalent to that in the corresponding M/G/1 queue with Poisson disasters [JS96, YKC02]. Therefore we can readily obtain $u^*_{\mathrm{WV}}(s)$ using the results in [JS96, YKC02]. Note that a similar observation with respect to the queue length is made in [KCC03] for a single-class M/G/1 queue with exponential working vacations.

Lemma 2.3. *$u^*_{\mathrm{WV}}(s)$ and $\mathrm{E}[U_{\mathrm{WV}}]$ are given by*

$$u^*_{\mathrm{WV}}(s) = \frac{(1-\nu)s-\gamma/\sigma}{s-\lambda_{\mathrm{WV}}/\sigma+(\lambda_{\mathrm{WV}}/\sigma)h^*_{\mathrm{WV}}(s)-\gamma/\sigma}, \qquad \mathrm{E}[U_{\mathrm{WV}}] = \frac{\rho_{\mathrm{WV}}-\sigma\nu}{\gamma}, \tag{2.7}$$

respectively, where ν denotes the conditional steady state probability that the server is busy given that it is on working vacation. Note that ν is given by

$$\nu = \frac{(1-r)\lambda_{\mathrm{WV}}}{(1-r)\lambda_{\mathrm{WV}}+\gamma}, \tag{2.8}$$

where r $(r>0)$ denotes the unique real root of the following equation.

$$z = h^*_{\mathrm{WV}}\big(\gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma)z\big), \qquad |z|<1. \tag{2.9}$$

The proof of Lemma 2.3 is given in Appendix 2.C.

Remark 2.3 (Remark 2.2 in [YKC02]). *The solution r of (2.9) represents the probability that a randomly chosen busy period starting in a working vacation ends within the working vacation. To see this, consider an $M/G/1$ queue with arrival rate λ_{WV}, the LST $h^*_{\mathrm{WV}}(s)$ of service requirements of customers, and the processing rate σ. The LST $\phi^*(s)$ of the lengths of busy periods is then given by $\phi^*(s) = h^*_{\mathrm{WV}}(s/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma)\phi^*(s))$. Comparing this with (2.9), we have $r = \phi^*(\gamma) > 0$.*

Rearranging terms on the right side of $u^*_{\mathrm{WV}}(s)$ in (2.7) yields

$$u^*_{\mathrm{WV}}(s) = \frac{1-\nu}{1-\nu\tilde{f}^*_{\mathrm{WV}}(s)}, \tag{2.10}$$

where $\tilde{f}^*_{\mathrm{WV}}(s)$ is given by

$$\tilde{f}^*_{\mathrm{WV}}(s) = \frac{h^*_{\mathrm{WV}}(s)-r}{(\sigma\nu/\lambda_{\mathrm{WV}})\{\gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma)r - s\}}. \tag{2.11}$$

Remark 2.4. *Theorem 2 in [JS96] shows that $\tilde{f}^*_{\mathrm{WV}}(s)$ represents the LST of the remaining service requirement $\tilde{F}_{\mathrm{WV}}$ of a randomly chosen customer present in working vacation periods when customers are served under the LCFS-PR basis. Note that (2.7) and (2.10) imply*

$$\mathrm{E}[\tilde{F}_{\mathrm{WV}}] = \frac{1-\nu}{\nu}\cdot \mathrm{E}[U_{\mathrm{WV}}] = \frac{1-\nu}{\nu}\cdot\frac{\rho_{\mathrm{WV}}-\sigma\nu}{\gamma}. \tag{2.12}$$

Theorem 2.1. *$u^*(s)$ is given by*

$$u^*(s) = u^*_{\mathrm{WV}}(s)\cdot\left(P_{\mathrm{WV}} + P_{\mathrm{NP}}\cdot\frac{1-\tilde{f}^*_{\mathrm{WV}}(s)}{s\mathrm{E}[\tilde{F}_{\mathrm{WV}}]}\cdot u^*_{\mathrm{M/G/1}}(s)\right), \tag{2.13}$$

where $u^*_{\mathrm{M/G/1}}(s)$, $u^*_{\mathrm{WV}}(s)$, $\tilde{f}^*_{\mathrm{WV}}(s)$, and $\mathrm{E}[\tilde{F}_{\mathrm{WV}}]$ are given by (2.4), (2.7), (2.11), and (2.12), respectively, and P_{WV} and P_{NP} are given by

$$P_{\mathrm{WV}} = \frac{1-\rho_{\mathrm{NP}}}{1-\rho_{\mathrm{NP}}+\rho_{\mathrm{WV}}-\sigma\nu}, \qquad P_{\mathrm{NP}} = \frac{\rho_{\mathrm{WV}}-\sigma\nu}{1-\rho_{\mathrm{NP}}+\rho_{\mathrm{WV}}-\sigma\nu}, \tag{2.14}$$

respectively.

Proof. It follows from (2.10) and (2.12) that

$$\frac{1-u^*_{\mathrm{WV}}(s)}{s\mathrm{E}[U_{\mathrm{WV}}]} = \frac{1-\nu}{1-\nu\tilde{f}^*_{\mathrm{WV}}(s)} \cdot \frac{1-\tilde{f}^*_{\mathrm{WV}}(s)}{s\mathrm{E}[\tilde{F}_{\mathrm{WV}}]} = u^*_{\mathrm{WV}}(s) \cdot \frac{1-\tilde{f}^*_{\mathrm{WV}}(s)}{s\mathrm{E}[\tilde{F}_{\mathrm{WV}}]}. \tag{2.15}$$

Substituting (2.15) into (2.6) yields (2.13). Further (2.14) follows from (2.5) and (2.7). □

Remark 2.5. *Theorem 2.1 shows that U is stochastically decomposed into two independent non-negative random variables, i.e., $U = U_{\mathrm{WV}} + U_{\mathrm{I}}$, where the LST of non-negative random variable U_{I} is given by*

$$u^*_{\mathrm{I}}(s) = P_{\mathrm{WV}} + P_{\mathrm{NP}} \cdot \frac{1-\tilde{f}^*_{\mathrm{WV}}(s)}{s\mathrm{E}[\tilde{F}_{\mathrm{WV}}]} \cdot u^*_{\mathrm{M/G/1}}(s).$$

2.3 Waiting time and sojourn time

In this section, we consider the actual waiting time and sojourn time distributions of class k ($k \in \mathcal{K}$) customers in steady state, assuming the FCFS service discipline. Let W_k ($k \in \mathcal{K}$) denote the waiting time of a randomly chosen class k customer. For each k ($k \in \mathcal{K}$), we define $W_{\mathrm{WV},k}$ (resp. $W_{\mathrm{NP},k}$) as the waiting time of a randomly chosen class k customer arriving in a working vacation period (resp. a normal service period). Let $w^*_k(s)$, $w^*_{\mathrm{WV},k}(s)$, and $w^*_{\mathrm{NP},k}(s)$ ($k \in \mathcal{K}$) denote the LSTs of W_k, $W_{\mathrm{WV},k}$, and $W_{\mathrm{NP},k}$, respectively. Similarly, let $\overline{W}_k$ ($k \in \mathcal{K}$) denote the stationary sojourn time of class k customers. For each k ($k \in \mathcal{K}$), we define $\overline{W}_{\mathrm{WV},k}$ (resp. $\overline{W}_{\mathrm{NP},k}$) as the sojourn time of a randomly chosen class k customer arriving in a working vacation period (resp. a normal service period). Let $\overline{w}^*_k(s)$, $\overline{w}^*_{\mathrm{WV},k}(s)$, and $\overline{w}^*_{\mathrm{NP},k}(s)$ ($k \in \mathcal{K}$) denote the LSTs of $\overline{W}_k$, $\overline{W}_{\mathrm{WV},k}$, and $\overline{W}_{\mathrm{NP},k}$, respectively.

For each k ($k \in \mathcal{K}$), we define $P^{\mathrm{A}}_{\mathrm{WV},k}$ (resp. $P^{\mathrm{A}}_{\mathrm{NP},k}$) as the probability that a randomly chosen class k customer finds the server being on working vacation (resp. being in a normal service period) upon arrival. By definition, $w^*_k(s)$ and $\overline{w}^*_k(s)$ ($k \in \mathcal{K}$) are given by

$$w^*_k(s) = P^{\mathrm{A}}_{\mathrm{WV},k} \cdot w^*_{\mathrm{WV},k}(s) + P^{\mathrm{A}}_{\mathrm{NP},k} \cdot w^*_{\mathrm{NP},k}(s), \tag{2.16}$$

$$\overline{w}^*_k(s) = P^{\mathrm{A}}_{\mathrm{WV},k} \cdot \overline{w}^*_{\mathrm{WV},k}(s) + P^{\mathrm{A}}_{\mathrm{NP},k} \cdot \overline{w}^*_{\mathrm{NP},k}(s), \tag{2.17}$$

respectively. Because class k customers arrive according to a Poisson process with rate $\lambda_{\mathrm{WV},k}$ during working vacation periods and rate $\lambda_{\mathrm{NP},k}$ during normal service periods, $P^{\mathrm{A}}_{\mathrm{WV},k}$ and $P^{\mathrm{A}}_{\mathrm{NP},k}$ satisfy

$$\frac{P^{\mathrm{A}}_{\mathrm{WV},k}}{P^{\mathrm{A}}_{\mathrm{NP},k}} = \frac{\lambda_{\mathrm{WV},k} P_{\mathrm{WV}}}{\lambda_{\mathrm{NP},k} P_{\mathrm{NP}}}.$$

Therefore, using $P^{\mathrm{A}}_{\mathrm{WV},k} + P^{\mathrm{A}}_{\mathrm{NP},k} = 1$, we obtain

$$P^{\mathrm{A}}_{\mathrm{WV},k} = \frac{\lambda_{\mathrm{WV},k} P_{\mathrm{WV}}}{\lambda_{\mathrm{WV},k} P_{\mathrm{WV}} + \lambda_{\mathrm{NP},k} P_{\mathrm{NP}}} = \frac{\lambda_{\mathrm{WV},k}(1-\rho_{\mathrm{NP}})}{\lambda_{\mathrm{WV},k}(1-\rho_{\mathrm{NP}}) + \lambda_{\mathrm{NP},k}(\rho_{\mathrm{WV}} - \sigma\nu)}, \tag{2.18}$$

$$P^{\mathrm{A}}_{\mathrm{NP},k} = \frac{\lambda_{\mathrm{NP},k} P_{\mathrm{NP}}}{\lambda_{\mathrm{WV},k} P_{\mathrm{WV}} + \lambda_{\mathrm{NP},k} P_{\mathrm{NP}}} = \frac{\lambda_{\mathrm{NP},k}(\rho_{\mathrm{WV}} - \sigma\nu)}{\lambda_{\mathrm{WV},k}(1-\rho_{\mathrm{NP}}) + \lambda_{\mathrm{NP},k}(\rho_{\mathrm{WV}} - \sigma\nu)}. \tag{2.19}$$

Both W_k and $\overline{W}_k$ ($k \in \mathscr{K}$) are considered as the processing time of a certain amount of workload. More specifically, W_k ($k \in \mathscr{K}$) corresponds to the stationary processing time of the workload in system seen by an arriving customer of class k. On the other hand, $\overline{W}_k$ ($k \in \mathscr{K}$) corresponds to the stationary processing time of the sum of workload in system seen by an arriving customer of class k and his/her service requirement. To treat W_k and $\overline{W}_k$ in a unified way, we define $\chi_{\mathrm{WV}}(U_X)$ (resp. $\chi_{\mathrm{NP}}(U_X)$) as the processing time of the amount U_X of workload conditioned that the server is on working vacation (resp. in a normal service period) when its processing starts, where U_X is assumed to be a non-negative random variable whose distribution function and LST are given by $U_X(x)$ and $u^*_X(s)$, respectively. Because the processing rate in $\chi_{\mathrm{WV}}(U_X)$ may change from σ to one, we divide $\chi_{\mathrm{WV}}(U_X)$ into two parts, $\chi^{(\sigma)}_{\mathrm{WV}}(U_X)$ and $\chi^{(1)}_{\mathrm{WV}}(U_X)$, where $\chi^{(\sigma)}_{\mathrm{WV}}(U_X)$ (resp. $\chi^{(1)}_{\mathrm{WV}}(U_X)$) is defined as the length of a subinterval in $\chi_{\mathrm{WV}}(U_X)$, during which the processing rate is equal to σ (resp. one). By definition, $\chi_{\mathrm{WV}}(U_X) = \chi^{(\sigma)}_{\mathrm{WV}}(U_X) + \chi^{(1)}_{\mathrm{WV}}(U_X)$, where $\chi^{(\sigma)}_{\mathrm{WV}}(U_X) > 0$ for $U_X > 0$, and $\chi^{(1)}_{\mathrm{WV}}(U_X) \geq 0$. We then define $\chi^{**}_{\mathrm{WV}}(\omega, s \mid U_X)$ and $\chi^*_{\mathrm{NP}}(s \mid U_X)$ as

$$\begin{aligned} \chi^{**}_{\mathrm{WV}}(\omega, s \mid U_X) &= \mathrm{E}\left[\exp[-\omega\chi^{(\sigma)}_{\mathrm{WV}}(U_X)]\exp[-s\chi^{(1)}_{\mathrm{WV}}(U_X)]\right], \\ \chi^*_{\mathrm{NP}}(s \mid U_X) &= \mathrm{E}\left[\exp[-s\chi_{\mathrm{NP}}(U_X)]\right], \end{aligned}$$

respectively.

Lemma 2.4. *$\chi^{**}_{\mathrm{WV}}(\omega, s \mid U_X)$ and $\chi^*_{\mathrm{NP}}(s \mid U_X)$ are given by*

$$\chi^{**}_{\mathrm{WV}}(\omega, s \mid U_X) = u^*_X\left(\frac{\omega+\gamma}{\sigma}\right) + \frac{u^*_X(s) - u^*_X\left(\dfrac{\omega+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega+\gamma)/\sigma - s\}}, \qquad \chi^*_{\mathrm{NP}}(s \mid U_X) = u^*_X(s),$$

respectively.

Proof. We first consider $\chi^*_{\mathrm{NP}}(s \mid U_X)$. When the processing of U_X starts in a normal service period, the processing rate is fixed to one throughout its processing. We then have $\chi_{\mathrm{NP}}(U_X) = U_X/1$, from which $\chi^*_{\mathrm{NP}}(s \mid U_X) = u^*_X(s)$ follows. On the other hand, when the processing of U_X starts in a working vacation period, we have

$$(\chi^{(\sigma)}_{\mathrm{WV}}(U_X), \chi^{(1)}_{\mathrm{WV}}(U_X)) = \begin{cases} (\dfrac{U_X}{\sigma}, 0), & \tilde{V}_{\mathrm{S}} > \dfrac{U_X}{\sigma}, \\ (\tilde{V}_{\mathrm{S}}, U_X - \sigma\tilde{V}_{\mathrm{S}}), & \tilde{V}_{\mathrm{S}} \leq \dfrac{U_X}{\sigma}, \end{cases}$$

where $\tilde{V}_{\mathrm{S}}$ denotes the remaining length of the working vacation when the processing starts. Owing to the memoryless property of the exponential distribution, $\tilde{V}_{\mathrm{S}}$ is exponentially distributed with parameter γ. We then have

$$\begin{aligned}
&\chi_{\mathrm{WV}}^{**}(\omega,s \mid U_X) \\
&\quad = \int_0^\infty dU_X(x) \Bigg[\exp[-\gamma(x/\sigma)] \exp[-\omega(x/\sigma)] \\
&\qquad + \{1-\exp[-\gamma(x/\sigma)]\} \int_0^{x/\sigma} \frac{\gamma \exp[-\gamma\tau]}{1-\exp[-\gamma(x/\sigma)]} \cdot \exp[-\omega\tau] \exp[-s(x-\sigma\tau)] d\tau \Bigg],
\end{aligned} \tag{2.20}$$

from which the expression of $\chi_{\mathrm{WV}}^{**}(\omega,s \mid U_X)$ follows. □

Using Lemma 2.4, $\mathrm{E}[\chi_{\mathrm{WV}}^{(\sigma)}(U_X)]$, $\mathrm{E}[\chi_{\mathrm{WV}}^{(1)}(U_X)]$, and $\mathrm{E}[\chi_{\mathrm{NP}}(U_X)]$ are obtained to be

$$\mathrm{E}\left[\chi_{\mathrm{WV}}^{(\sigma)}(U_X)\right] = (-1) \cdot \lim_{\omega \to 0+} \frac{d}{d\omega} [\chi_{\mathrm{WV}}^{**}(\omega,0 \mid U_X)] = \frac{1-u_X^*(\gamma/\sigma)}{\gamma}, \tag{2.21}$$

$$\mathrm{E}\left[\chi_{\mathrm{WV}}^{(1)}(U_X)\right] = (-1) \cdot \lim_{s \to 0+} \frac{d}{ds} [\chi_{\mathrm{WV}}^{**}(0,s \mid U_X)] = \mathrm{E}[U_X] - \sigma \cdot \frac{1-u_X^*(\gamma/\sigma)}{\gamma}, \tag{2.22}$$

$$\mathrm{E}\left[\chi_{\mathrm{NP}}(U_X)\right] = \mathrm{E}[U_X]. \tag{2.23}$$

We now turn our attention to the waiting time distribution. Consider the censored process obtained by removing all normal service periods. In the resulting process, class k customers arrive according to a Poisson process. Owing to PASTA, the conditional workload in system seen by a randomly chosen class k customer arriving in a working vacation period has the same distribution as U_{WV}. Therefore the conditional waiting time distributions are identical among classes. Similarly, the conditional workload in system seen by class k ($k \in \mathcal{K}$) customers arriving in normal service periods has the same distribution as U_{NP}. Thus, the conditional waiting time distributions are also identical among classes.

Let $W_{\mathrm{WV}}^{(\sigma)}$ (resp. $W_{\mathrm{WV}}^{(1)}$) denote the length of an interval during which a randomly chosen customer waits for his/her service in a working vacation period (resp. normal service period), given that the customer arrived in the working vacation period. By definition, $W_{\mathrm{WV},k} = W_{\mathrm{WV}}^{(\sigma)} + W_{\mathrm{WV}}^{(1)}$ for all k ($k \in \mathcal{K}$). Also, let W_{NP} denote the conditional waiting time of a randomly chosen customer given that the customer arrives in a normal service period. We then define $w_{\mathrm{WV}}^{**}(\omega,s)$ as the joint LST of $W_{\mathrm{WV}}^{(\sigma)}$ and $W_{\mathrm{WV}}^{(1)}$, and $w_{\mathrm{NP}}^*(s)$ as the LST of W_{NP}.

$$w_{\mathrm{WV}}^{**}(\omega,s) = \mathrm{E}\left[\exp[-\omega W_{\mathrm{WV}}^{(\sigma)}] \exp[-s W_{\mathrm{WV}}^{(1)}]\right], \qquad w_{\mathrm{NP}}^*(s) = \mathrm{E}[\exp[-s W_{\mathrm{NP}}]].$$

Theorem 2.2. *$w_{\mathrm{WV}}^{**}(\omega,s)$ and $w_{\mathrm{NP}}^{*}(s)$ are given by*

$$w_{\mathrm{WV}}^{**}(\omega,s)=u_{\mathrm{WV}}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)+\frac{u_{\mathrm{WV}}^{*}(s)-u_{\mathrm{WV}}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega+\gamma)/\sigma-s\}},\qquad w_{\mathrm{NP}}^{*}(s)=u_{\mathrm{NP}}^{*}(s),$$

respectively.

Proof. By definition, $w_{\mathrm{WV}}^{**}(\omega,s)=\chi_{\mathrm{WV}}^{**}(\omega,s\mid U_{\mathrm{WV}})$ and $w_{\mathrm{NP}}^{*}(s)=\chi_{\mathrm{NP}}^{*}(s\mid U_{\mathrm{NP}})$, so that Theorem 2.2 immediately follows from Lemma 2.4. □

Because $W_{\mathrm{WV},k}=W_{\mathrm{WV}}^{(\sigma)}+W_{\mathrm{WV}}^{(1)}$ and $W_{\mathrm{NP},k}=W_{\mathrm{NP}}$ for all k ($k\in\mathcal{K}$),

$$w_{\mathrm{WV},k}^{*}(s)=w_{\mathrm{WV}}^{**}(s,s),\quad w_{\mathrm{NP},k}^{*}(s)=w_{\mathrm{NP}}^{*}(s),\quad \forall k\in\mathcal{K}.$$

Thus the LST $w_k^*(s)$ ($k\in\mathcal{K}$) of the waiting time distribution of class k customers is obtained by (2.16). In particular, the mean waiting time is given by

$$\mathrm{E}[W_k]=P_{\mathrm{WV},k}^{\mathrm{A}}\cdot\left[\frac{(1-\sigma)(1-u_{\mathrm{WV}}^{*}(\gamma/\sigma))}{\gamma}+\mathrm{E}[U_{\mathrm{WV}}]\right]+P_{\mathrm{NP},k}^{\mathrm{A}}\cdot\mathrm{E}[U_{\mathrm{NP}}],$$

where $\mathrm{E}[U_{\mathrm{NP}}]$ denotes the mean conditional workload in system given the system being in a normal service period and it is obtained from (2.3) and Lemma 2.3.

$$\mathrm{E}[U_{\mathrm{NP}}]=\frac{\rho_{\mathrm{WV}}-\sigma}{\gamma}+\frac{\lambda_{\mathrm{WV}}\mathrm{E}[H_{\mathrm{WV}}^{2}]}{2(\rho_{\mathrm{WV}}-\sigma\nu)}+\frac{\lambda_{\mathrm{NP}}\mathrm{E}[H_{\mathrm{NP}}^{2}]}{2(1-\rho_{\mathrm{NP}})}.$$

Next we consider the sojourn time distribution. For each k ($k\in\mathcal{K}$), let $\overline{W}_{\mathrm{WV},k}^{(\sigma)}$ (resp. $\overline{W}_{WV,k}^{(1)}$) denote the length of time during which a randomly chosen class k customer spends in a working vacation period (resp. a normal service period), given that the customer arrives in the working vacation period. By definition, $\overline{W}_{\mathrm{WV},k}^{(\sigma)}>0$, $\overline{W}_{\mathrm{WV},k}^{(1)}\ge 0$, and $\overline{W}_{\mathrm{WV},k}=\overline{W}_{\mathrm{WV},k}^{(\sigma)}+\overline{W}_{\mathrm{WV},k}^{(1)}$. We define $\overline{w}_{\mathrm{WV},k}^{**}(\omega,s)$ ($k\in\mathcal{K}$) as the joint LST of $\overline{W}_{\mathrm{WV},k}^{(\sigma)}$ and $\overline{W}_{WV,k}^{(1)}$, and $\overline{w}_{\mathrm{NP},k}^{*}(s)$ ($k\in\mathcal{K}$) as the LST of $\overline{W}_{\mathrm{NP},k}$.

$$\overline{w}_{\mathrm{WV},k}^{**}(\omega,s)=\mathrm{E}\left[\exp[-\omega\overline{W}_{\mathrm{WV},k}^{(\sigma)}]\exp[-s\overline{W}_{\mathrm{WV},k}^{(1)}]\right],\qquad \overline{w}_{\mathrm{NP},k}^{*}(s)=\mathrm{E}\left[\exp[-s\overline{W}_{\mathrm{NP},k}]\right].$$

Theorem 2.3. *$\overline{w}_{\mathrm{WV},k}^{**}(\omega,s)$ and $\overline{w}_{\mathrm{NP},k}^{*}(s)$ ($k\in\mathcal{K}$) are given by*

$$\begin{aligned}\overline{w}_{\mathrm{WV},k}^{**}(\omega,s) &= u_{\mathrm{WV}}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)h_{\mathrm{WV},k}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)\\ &\quad+\frac{u_{\mathrm{WV}}^{*}(s)h_{\mathrm{WV},k}^{*}(s)-u_{\mathrm{WV}}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)h_{\mathrm{WV},k}^{*}\left(\frac{\omega+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega+\gamma)/\sigma-s\}},\\ \overline{w}_{\mathrm{NP},k}^{*}(s) &= u_{\mathrm{NP}}^{*}(s)\cdot h_{\mathrm{NP},k}^{*}(s),\end{aligned}$$

respectively.

Proof. By definition, $\overline{w}^{**}_{\mathrm{WV},k}(\omega,s) = \chi^{**}_{\mathrm{WV}}(\omega,s \mid U_{\mathrm{WV}} + H_{\mathrm{WV},k})$, and $\overline{w}^{*}_{\mathrm{NP},k}(s) = \chi^{*}_{\mathrm{NP}}(s \mid U_{\mathrm{NP}} + H_{\mathrm{NP},k})$. Theorem 2.3 then follows from Lemma 2.4. □

Note that $\overline{w}^{*}_{\mathrm{WV},k}(s) = \overline{w}^{**}_{\mathrm{WV},k}(s,s)$ $(k \in \mathscr{K})$. Thus the LST $\overline{w}^{*}_{k}(s)$ $(k \in \mathscr{K})$ of the sojourn time distribution of class k customers is obtained by (2.17). In particular, the mean sojourn time is given by

$$\mathrm{E}[\overline{W}_k] = P^{\mathrm{A}}_{\mathrm{WV},k} \cdot \mathrm{E}[\overline{W}_{\mathrm{WV},k}] + P^{\mathrm{A}}_{\mathrm{NP},k} \cdot \mathrm{E}[\overline{W}_{\mathrm{NP},k}],$$

where

$$\begin{aligned}\mathrm{E}[\overline{W}_{\mathrm{WV},k}] &= \frac{(1-\sigma)(1-u^{*}_{\mathrm{WV}}(\gamma/\sigma)h^{*}_{\mathrm{WV},k}(\gamma/\sigma))}{\gamma} + \mathrm{E}[U_{\mathrm{WV}}] + \mathrm{E}[H_{\mathrm{WV},k}],\\ \mathrm{E}[\overline{W}_{\mathrm{NP},k}] &= \mathrm{E}[U_{\mathrm{NP}}] + \mathrm{E}[H_{\mathrm{NP},k}].\end{aligned}$$

2.4 Joint distribution of queue lengths and workloads in system

In this section, we consider the joint distribution of the numbers of customers and the workloads in system in respective classes. To do so, we first derive the joint LST of the attained waiting time and the remaining amount of service requirement of a class k customer being served. With this result, the joint distributions are derived.

For each k $(k \in \mathscr{K})$, let $\overline{\rho}^{(\sigma)}_{\mathrm{WV},k}$ (resp. $\overline{\rho}^{(1)}_{\mathrm{WV},k}$) denote the time-average probability that class k customers, who arrived in working vacation periods, are being served in working vacation periods (resp. in normal service periods). Also, let $\overline{\rho}^{(1)}_{\mathrm{NP},k}$ $(k \in \mathscr{K})$ denote the time-average probability that class k customers arriving in normal service periods are being served.

Lemma 2.5. *$\overline{\rho}^{(\sigma)}_{\mathrm{WV},k}$, $\overline{\rho}^{(1)}_{\mathrm{WV},k}$, and $\overline{\rho}^{(1)}_{\mathrm{NP},k}$ $(k \in \mathscr{K})$ are given by*

$$\overline{\rho}^{(\sigma)}_{\mathrm{WV},k} = P_{\mathrm{WV}} \cdot \nu \cdot \frac{\lambda_{\mathrm{WV},k}\big(1-h^{*}_{\mathrm{WV},k}(\gamma/\sigma)\big)}{\lambda_{\mathrm{WV}}\big(1-h^{*}_{\mathrm{WV}}(\gamma/\sigma)\big)}, \tag{2.24}$$

$$\overline{\rho}^{(1)}_{\mathrm{WV},k} = P_{\mathrm{WV}} \cdot \left[\rho_{WV,k} - \sigma\nu \cdot \frac{\lambda_{\mathrm{WV},k}\big(1-h^{*}_{\mathrm{WV},k}(\gamma/\sigma)\big)}{\lambda_{\mathrm{WV}}\big(1-h^{*}_{\mathrm{WV}}(\gamma/\sigma)\big)}\right], \tag{2.25}$$

$$\overline{\rho}^{(1)}_{\mathrm{NP},k} = P_{\mathrm{NP}} \cdot \rho_{\mathrm{NP},k}, \tag{2.26}$$

respectively.

The proof of Lemma 2.5 is given in Appendix 2.D.

Remark 2.6. *Let $\overline{\rho}$ denote the utilization factor, i.e., the time-average probability that customers are being served. Recall that ν in (2.8) represents the conditional probability of the server being busy given that the server is on working vacation. We then have*

$$\overline{\rho} = 1 - P_{\mathrm{WV}} \cdot (1-\nu) = P_{\mathrm{WV}} \cdot \nu + P_{\mathrm{NP}} = \frac{(1-\rho_{\mathrm{NP}})\nu + \rho_{\mathrm{WV}} - \sigma\nu}{1-\rho_{\mathrm{NP}} + \rho_{\mathrm{WV}} - \sigma\nu}.$$

Furthermore, using Lemma 2.5, we can verify

$$\sum_{k\in\mathcal{K}} \overline{\rho}^{(\sigma)}_{\mathrm{WV},k} = P_{\mathrm{WV}} \cdot \nu, \qquad \sum_{k\in\mathcal{K}} (\overline{\rho}^{(1)}_{\mathrm{WV},k} + \overline{\rho}^{(1)}_{\mathrm{NP},k}) = P_{\mathrm{NP}}.$$

We now consider the attained waiting time [Sen89], which is defined as the length of time spent by a customer being served (if any) in the system. When the system is empty, the attained waiting time is defined to be zero. Note that under the FCFS service discipline, all *waiting* customers in the system arrived in the attained waiting time.

For later use, we divide the attained waiting time into two parts: One is the (sub)interval in working vacation periods and the other is the (sub)interval in normal service periods. Let $A^{(\sigma)}_{\mathrm{WV},k}$ ($k \in \mathcal{K}$) denote the length of time in the attained waiting time, during which the server was on working vacation, given that a class k customer is being served. Furthermore, for each k ($k \in \mathcal{K}$), let $A^{(1)}_{\mathrm{WV},k}$ (resp. $A^{(1)}_{\mathrm{NP},k}$) denote the length of time in the attained waiting time, during which the server worked in a normal service period, given that a class k customer, who arrived in a working vacation period (resp. a normal service period), is being served. For a class k ($k \in \mathcal{K}$) customer being served, let $\tilde{H}_k$ denote the remaining amount of his/her service requirement. We then define the following joint LSTs:

$$\begin{aligned}
&a^{**}_{\mathrm{WV,WV},k}(\omega_k, \alpha_k) \\
&\quad = \mathrm{E}\Big[\exp[-\omega_k A^{(\sigma)}_{\mathrm{WV},k}]\exp[-\alpha_k \tilde{H}_k] \,\Big|\, \text{a class } k \text{ customer is being served} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{at processing rate } \sigma\Big], \\
&a^{***}_{\mathrm{WV,NP},k}(\omega_k, s_k, \alpha_k) \\
&\quad = \mathrm{E}\Big[\exp[-\omega_k A^{(\sigma)}_{\mathrm{WV},k}]\exp[-s_k A^{(1)}_{\mathrm{WV},k}]\exp[-\alpha_k \tilde{H}_k] \\
&\qquad\qquad \Big|\, \text{a class } k \text{ customer who arrived in a working vacation period} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{is being served at processing rate one}\Big], \\
&a^{**}_{\mathrm{NP},k}(s_k, \alpha_k) \\
&\quad = \mathrm{E}\Big[\exp[-s_k A^{(1)}_{\mathrm{NP},k}]\exp[-\alpha_k \tilde{H}_k] \,\Big|\, \text{a class } k \text{ customer, who arrived in a} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{normal service period, is being served}\Big].
\end{aligned}$$

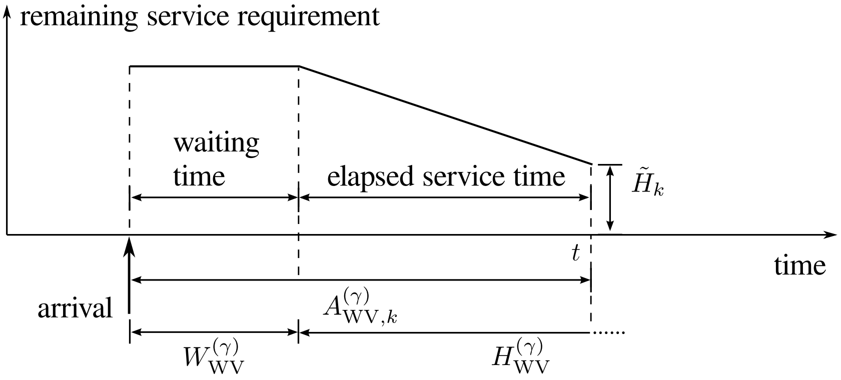
remaining service requirement
waiting time
elapsed service time
$\tilde{H}_k$
t
time
arrival
$A^{(\gamma)}_{\mathrm{WV},k}$
$W^{(\gamma)}_{\mathrm{WV}}$
$H^{(\gamma)}_{\mathrm{WV}}$

Figure 2.1: Attained waiting time of a class k customer in a working vacation period.

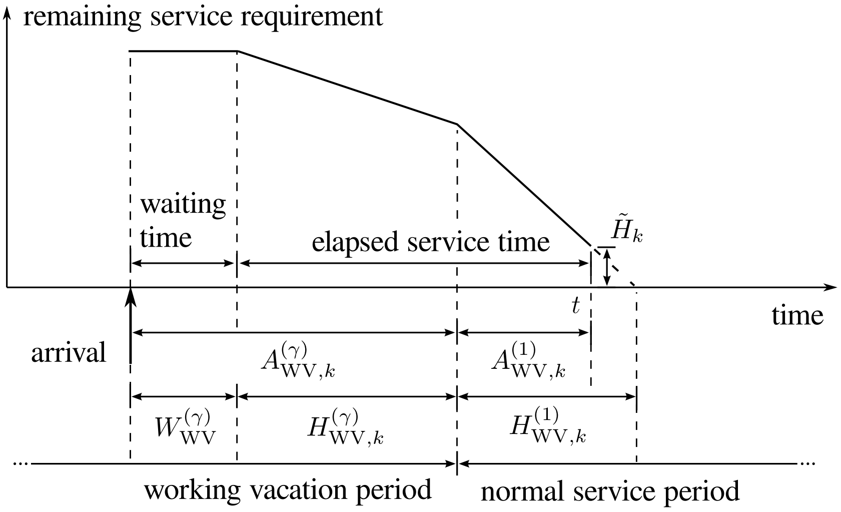
remaining service requirement
waiting time
elapsed service time
$\tilde{H}_k$
t
time
arrival
$A^{(\gamma)}_{\mathrm{WV},k}$
$A^{(1)}_{\mathrm{WV},k}$
$W^{(\gamma)}_{\mathrm{WV}}$
$H^{(\gamma)}_{\mathrm{WV},k}$
$H^{(1)}_{\mathrm{WV},k}$
working vacation period
normal service period

Figure 2.2: Attained waiting time of a class k customer who started his/her service in a working vacation period and will end his/her service in a normal service period.

See Figures 2.1–2.4. Figure 2.1 corresponds to $a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k)$, Figures 2.2 and 2.3 correspond to $a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k)$, and Figure 2.4 corresponds to $a^{**}_{\mathrm{NP},k}(s_k,\alpha_k)$.

Moreover, for each k ($k \in \mathscr{K}$), let $H^{(\sigma)}_{\mathrm{WV},k}$ (resp. $H^{(1)}_{\mathrm{WV},k}$) denote the lengths of time during which a class k customer, who started his/her service in a working vacation period, is served in the working vacation period (resp. the subsequent normal service period). We then define $\hat{h}^{**}_{\mathrm{WV},k}(\omega,s)$ as the joint LST of $H^{(\sigma)}_{\mathrm{WV},k}$ and $H^{(1)}_{\mathrm{WV},k}$. Using Lemma 2.4, we obtain

$$\hat{h}^{**}_{\mathrm{WV},k}(\omega,s) = \mathrm{E}\left[\exp[-\omega H^{(\sigma)}_{\mathrm{WV},k}]\exp[-sH^{(1)}_{\mathrm{WV},k}]\right] = \chi^{**}_{\mathrm{WV}}(\omega,s \mid H_{\mathrm{WV},k})$$

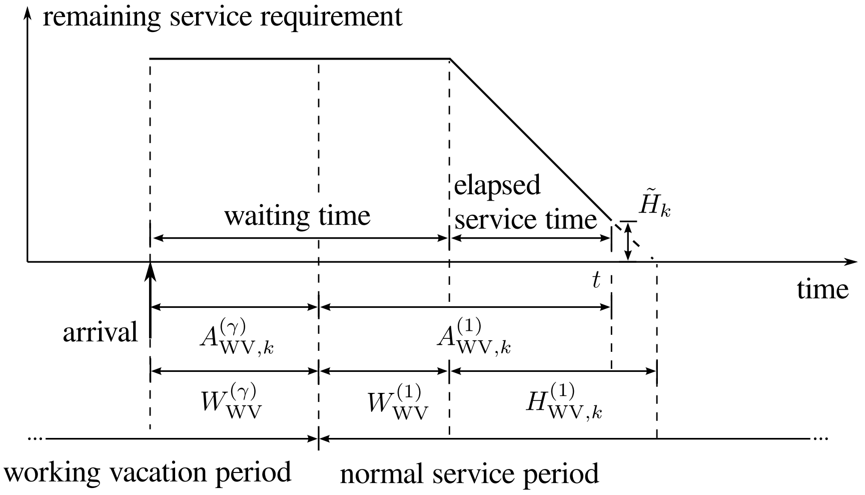
remaining service requirement
elapsed
service time
waiting time
$\tilde{H}_k$
t
time
arrival
$A^{(\gamma)}_{\mathrm{WV},k}$
$A^{(1)}_{\mathrm{WV},k}$
$W^{(\gamma)}_{\mathrm{WV}}$
$W^{(1)}_{\mathrm{WV}}$
$H^{(1)}_{\mathrm{WV},k}$
working vacation period
normal service period

Figure 2.3: Attained waiting time of a class k customer who arrived in a working vacation period and started his/her service in a normal service period.

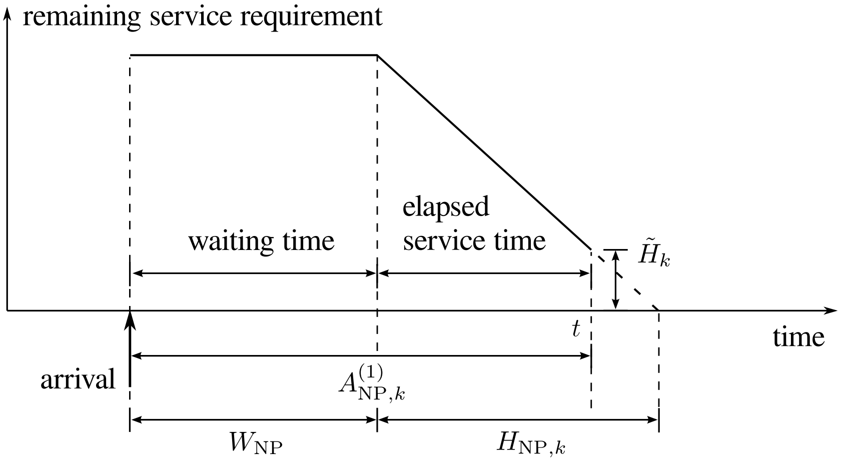

Figure 2.4: Attained waiting time of a class k customer who arrived in a normal service period.

$$= h^*_{\mathrm{WV},k}\left(\frac{\omega+\gamma}{\sigma}\right) + \frac{h^*_{\mathrm{WV},k}(s) - h^*_{\mathrm{WV},k}\left(\frac{\omega+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega+\gamma)/\sigma - s\}}.$$

We then have the following theorem, whose proof is provided in Appendix 2.E.

Theorem 2.4. *$a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k)$, $a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k)$, and $a^{**}_{\mathrm{NP},k}(s_k,\alpha_k)$ are given by*

$$a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k) = \frac{(1/\gamma)u^*_{\mathrm{WV}}\left(\frac{\omega_k+\gamma}{\sigma}\right)}{\mathrm{E}\left[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k}\right]} \cdot \frac{h^*_{\mathrm{WV},k}(\alpha_k) - h^*_{\mathrm{WV},k}\left(\frac{\omega_k+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega_k+\gamma)/\sigma - \alpha_k\}}, \tag{2.27}$$

$$a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k) = \frac{1}{\mathrm{E}\left[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k}\right]} \tag{2.28}$$

$$\cdot \left[u^*_{\mathrm{WV}}\left(\frac{\omega_k+\gamma}{\sigma}\right)\frac{\hat{h}^{**}_{\mathrm{WV},k}(\omega_k,\alpha_k) - \hat{h}^{**}_{\mathrm{WV},k}(\omega_k,s_k)}{s_k-\alpha_k} + \frac{u^*_{\mathrm{WV}}(s_k) - u^*_{\mathrm{WV}}\left(\frac{\omega_k+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega_k+\gamma)/\sigma - s_k\}} \cdot \frac{h^*_{\mathrm{WV},k}(\alpha_k) - h^*_{\mathrm{WV},k}(s_k)}{s_k-\alpha_k}\right], \tag{2.29}$$

$$a^{**}_{\mathrm{NP},k}(s_k,\alpha_k) = u^*_{\mathrm{NP}}(s_k)\cdot\frac{h^*_{\mathrm{NP},k}(\alpha_k) - h^*_{\mathrm{NP},k}(s_k)}{\mathrm{E}[H_{\mathrm{NP},k}](s_k-\alpha_k)}, \tag{2.30}$$

respectively, where $\mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k}]$ and $\mathrm{E}[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k}]$ are given in (2.48) and (2.49), respectively.

With Theorems 2.2, 2.3, and 2.4, we can verify that the LSTs of conditional attained waiting times $a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k)$ and $a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k)$ are represented in terms of $w^{**}_{\mathrm{WV}}(\omega,s)$ and $\overline{w}^{**}_{\mathrm{WV},k}(\omega,s)$.

Corollary 2.1. *$a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k)$ and $a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k)$ are given by*

$$a^{**}_{\mathrm{WV,WV},k}(\omega_k,\alpha_k) = \frac{w^{**}_{\mathrm{WV}}(\omega_k,\alpha_k)h^*_{\mathrm{WV},k}(\alpha_k) - \overline{w}^{**}_{\mathrm{WV},k}(\omega_k,\alpha_k)}{(\omega_k-\sigma\alpha_k)\mathrm{E}\left[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k}\right]},$$

$$a^{***}_{\mathrm{WV,NP},k}(\omega_k,s_k,\alpha_k) = \frac{\left(w^{**}_{\mathrm{WV}}(\omega_k,s_k) - w^{**}_{\mathrm{WV}}(\omega_k,\alpha_k)\right)h^*_{\mathrm{WV},k}(\alpha_k) - \left(\overline{w}^{**}_{\mathrm{WV},k}(\omega_k,s_k) - \overline{w}^{**}_{\mathrm{WV},k}(\omega_k,\alpha_k)\right)}{(s_k-\alpha_k)\mathrm{E}\left[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k}\right]},$$

respectively.

Let $\overline{L}_{\mathrm{WV},k}$ (resp. $\overline{L}_{\mathrm{NP},k}$) ($k \in \mathcal{K}$) denote the number of class k customers in the system, who arrived in working vacation periods (resp. normal service periods). Also, let $\overline{U}_{\mathrm{WV},k}$ (resp. $\overline{U}_{\mathrm{NP},k}$) ($k \in \mathcal{K}$) denote the workload in system, which is brought by class k customers who arrived in working vacation periods (resp. normal service periods). We then define the joint transform $\psi(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \boldsymbol{s}_{\mathrm{WV}}, \boldsymbol{s}_{\mathrm{NP}})$ as

$$\begin{aligned}
&\psi(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \boldsymbol{s}_{\mathrm{WV}}, \boldsymbol{s}_{\mathrm{NP}}) \\
&\qquad = \mathrm{E}\left[\prod_{k \in \mathcal{K}} \left(z_{\mathrm{WV},k}^{\overline{L}_{\mathrm{WV},k}} \cdot z_{\mathrm{NP},k}^{\overline{L}_{\mathrm{NP},k}} \cdot \exp[-s_{\mathrm{WV},k} \overline{U}_{\mathrm{WV},k}] \cdot \exp[-s_{\mathrm{NP},k} \overline{U}_{\mathrm{NP},k}] \right)\right],
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{z}_{\mathrm{WV}} &= (z_{\mathrm{WV},1}, z_{\mathrm{WV},2}, \dots, z_{\mathrm{WV},K}), \qquad \boldsymbol{z}_{\mathrm{NP}} = (z_{\mathrm{NP},1}, z_{\mathrm{NP},2}, \dots, z_{\mathrm{NP},K}), \\
\boldsymbol{s}_{\mathrm{WV}} &= (s_{\mathrm{WV},1}, s_{\mathrm{WV},2}, \dots, s_{\mathrm{WV},K}), \qquad \boldsymbol{s}_{\mathrm{NP}} = (s_{\mathrm{NP},1}, s_{\mathrm{NP},2}, \dots, s_{\mathrm{NP},K}).
\end{aligned}$$

Theorem 2.5. *$\psi(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \boldsymbol{s}_{\mathrm{WV}}, \boldsymbol{s}_{\mathrm{NP}})$ is given by*

$$\begin{aligned}
&\psi(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \boldsymbol{s}_{\mathrm{WV}}, \boldsymbol{s}_{\mathrm{NP}}) \\
&\quad = (1-\nu) P_{\mathrm{WV}} \\
&\qquad + \sum_{k \in \mathcal{K}} z_{\mathrm{WV},k} \overline{\rho}^{(\sigma)}_{\mathrm{WV},k} a^{**}_{\mathrm{WV},\mathrm{WV},k}\Big(\sum_{i \in \mathcal{K}} [\lambda_{\mathrm{WV},i} - \lambda_{\mathrm{WV},i} z_{\mathrm{WV},i} h^*_{\mathrm{WV},i}(s_{\mathrm{WV},i})], s_{\mathrm{WV},k}\Big) \\
&\qquad + \sum_{k \in \mathcal{K}} z_{\mathrm{WV},k} \overline{\rho}^{(1)}_{\mathrm{WV},k} a^{***}_{\mathrm{WV},\mathrm{NP},k}\Big(\sum_{i \in \mathcal{K}} [\lambda_{\mathrm{WV},i} - \lambda_{\mathrm{WV},i} z_{\mathrm{WV},i} h^*_{\mathrm{WV},i}(s_{\mathrm{WV},i})], \\
&\qquad\qquad\qquad\qquad \sum_{i \in \mathcal{K}} [\lambda_{\mathrm{NP},i} - \lambda_{\mathrm{NP},i} z_{\mathrm{NP},i} h^*_{\mathrm{NP},i}(s_{\mathrm{NP},i})], s_{\mathrm{WV},k}\Big) \\
&\qquad + \sum_{k \in \mathcal{K}} z_{\mathrm{NP},k} \overline{\rho}^{(1)}_{\mathrm{NP},k} a^{**}_{\mathrm{NP},k}\Big(\sum_{i \in \mathcal{K}} [\lambda_{\mathrm{NP},i} - \lambda_{\mathrm{NP},i} z_{\mathrm{NP},i} h^*_{\mathrm{NP},i}(s_{\mathrm{NP},i})], s_{\mathrm{NP},k}\Big).
\end{aligned}$$

Proof. Note first that the system is empty with probability $1 - \overline{\rho} = (1-\nu)P_{\mathrm{WV}}$ (see Remark 2.6). Furthermore, when a customer is being served, all waiting customers arrived in the attained waiting time, as noted at the beginning of this section. Theorem 2.5 immediately follows from those observations. □

Remark 2.7. *Let $\overline{L}_{\mathrm{WV}}$ (resp. $\overline{L}_{\mathrm{NP}}$) denote the total number of customers in the system, who arrived in working vacation periods (resp. normal service periods). Also, let $\overline{U}_{\mathrm{WV}}$ (resp. $\overline{U}_{\mathrm{NP}}$) denote the total workload in system, which was brought by customers who arrived in working vacation periods (resp. normal service periods). As stated in Remark 2.2, we can obtain those by considering the single-class system with λ_{WV}, $h^*_{\mathrm{WV}}(s)$, λ_{NP}, and $h^*_{\mathrm{NP}}(s)$. Therefore Theorem 2.5 also provides the formula for the joint transform of $\overline{L}_{\mathrm{WV}}$, $\overline{L}_{\mathrm{NP}}$, $\overline{U}_{\mathrm{WV}}$, and $\overline{U}_{\mathrm{NP}}$ implicitly, because it corresponds the case of $K = 1$.*

Taking the partial derivatives of $\psi(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \boldsymbol{s}_{\mathrm{WV}}, \boldsymbol{s}_{\mathrm{NP}})$, we can obtain the moments of $\overline{L}_{\mathrm{WV},k}$, $\overline{L}_{\mathrm{NP},k}$, $\overline{U}_{\mathrm{WV},k}$, and $\overline{U}_{\mathrm{NP},k}$ ($k \in \mathscr{K}$). In particular, we have

$$\begin{aligned}
\mathrm{E}[\overline{L}_{\mathrm{WV},k}] &= \lambda_{\mathrm{WV},k} P_{\mathrm{WV}} \cdot \mathrm{E}[\overline{W}_{\mathrm{WV},k}], \qquad \mathrm{E}[\overline{L}_{\mathrm{NP},k}] = \lambda_{\mathrm{NP},k} P_{\mathrm{NP}} \cdot \mathrm{E}[\overline{W}_{\mathrm{NP},k}], \\
\mathrm{E}[\overline{U}_{\mathrm{WV},k}] &= P_{\mathrm{WV}} \rho_{\mathrm{WV},k} \left(\mathrm{E}[U_{\mathrm{WV}}] + \frac{\mathrm{E}[H_{\mathrm{WV},k}^2]}{2\mathrm{E}[H_{\mathrm{WV},k}]} + \frac{1}{\gamma} \right) - \frac{\sigma}{\gamma} \left(\overline{\rho}_{\mathrm{WV},k}^{(\sigma)} + \overline{\rho}_{\mathrm{WV},k}^{(1)} \right), \\
\mathrm{E}[\overline{U}_{\mathrm{NP},k}] &= P_{\mathrm{NP}} \rho_{\mathrm{NP},k} \left(\frac{\mathrm{E}[H_{\mathrm{NP},k}^2]}{2\mathrm{E}[H_{\mathrm{NP},k}]} + \mathrm{E}[U_{\mathrm{NP}}] \right).
\end{aligned}$$

2.5 Busy cycle

The busy cycle is defined as the interval between ends of successive busy periods. In order to analyze the busy cycle and related quantities, we first consider the first passage time to the empty system. More specifically, we define $T_{\mathrm{E|WV}}$ (resp. $T_{\mathrm{E|NP}}$) as the first passage time to the empty system given that the server is on working vacation (resp. in a normal service period) at time 0. We divide $T_{\mathrm{E|WV}}$ into two parts: $T_{\mathrm{E|WV}}^{(\sigma)}$ and $T_{\mathrm{E|WV}}^{(1)}$, where $T_{\mathrm{E|WV}}^{(\sigma)}$ (resp. $T_{\mathrm{E|WV}}^{(1)}$) denotes the length of a subinterval during which the server is on working vacation (resp. in a normal service period). By definition, $T_{\mathrm{E|WV}} = T_{\mathrm{E|WV}}^{(\sigma)} + T_{\mathrm{E|WV}}^{(1)}$. Furthermore, for each k ($k \in \mathscr{K}$), we define $N_{\mathrm{WV},k}^{(\sigma)}$ (resp. $N_{\mathrm{WV},k}^{(1)}$) as the number of class k customers arriving in $T_{\mathrm{E|WV}}^{(\sigma)}$ (resp. $T_{\mathrm{E|WV}}^{(1)}$). Similarly, we define $N_{\mathrm{NP},k}$ ($k \in \mathscr{K}$) as the number of class k customers arriving in $T_{\mathrm{E|NP}}$. Let $S(t)$ ($t \geq 0$) denote the state of the server at time t, i.e., and $S(t) = \mathrm{WV}$ if the server is on working vacation at time t, and otherwise $S(t) = \mathrm{NP}$. We define $U(t)$ ($t \geq 0$) as the total workload at time t. We are interested in the following joint transforms.

$$\begin{aligned}
&\zeta_{\mathrm{WV}}^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{WV}}, s_{\mathrm{NP}} \mid x) \\
&\quad = \mathrm{E}\left[\left(\prod_{k \in \mathscr{K}} z_{\mathrm{WV},k}^{N_{\mathrm{WV},k}^{(\sigma)}} \cdot z_{\mathrm{NP},k}^{N_{\mathrm{WV},k}^{(1)}} \right) \exp\left[-s_{\mathrm{WV}} T_{\mathrm{E|WV}}^{(\sigma)}\right] \exp\left[-s_{\mathrm{NP}} T_{\mathrm{E|WV}}^{(1)}\right] \,\middle|\, U(0) = x, S(0) = \mathrm{WV} \right], \\
&\zeta_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}} \mid x) \\
&\quad = \mathrm{E}\left[\left(\prod_{k \in \mathscr{K}} z_{\mathrm{NP},k}^{N_{\mathrm{NP},k}} \right) \exp\left[-s_{\mathrm{NP}} T_{\mathrm{E|NP}}\right] \,\middle|\, U(0) = x, S(0) = \mathrm{NP} \right],
\end{aligned}$$

where $\boldsymbol{z}_{\mathrm{WV}} = (z_{\mathrm{WV},1}, z_{\mathrm{WV},2}, \ldots, z_{\mathrm{WV},K})$ and $\boldsymbol{z}_{\mathrm{NP}} = (z_{\mathrm{NP},1}, z_{\mathrm{NP},2}, \ldots, z_{\mathrm{NP},K})$.

Lemma 2.6. *$\zeta_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}} \mid x)$ is given by*

$$\zeta_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}} \mid x) = \exp\left[-q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}}) x\right], \tag{2.31}$$

where $q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}})$ is defined as

$$q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}}) = s_{\mathrm{NP}} + \lambda_{\mathrm{NP}} - \sum_{k \in \mathscr{K}} z_{\mathrm{NP},k} \lambda_{\mathrm{NP},k} \int_0^\infty \zeta_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}} \mid y) dH_{\mathrm{NP},k}(y), \tag{2.32}$$

and it is given by

$$q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}}) = s_{\mathrm{NP}} + \lambda_{\mathrm{NP}} - \sum_{k\in\mathcal{K}} z_{\mathrm{NP},k}\lambda_{\mathrm{NP},k}h^*_{\mathrm{NP},k}\big(q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}})\big). \tag{2.33}$$

The proof of Lemma 2.6 is given in Appendix 2.F.

Next, we consider the joint transform $\zeta^*_{\mathrm{WV}}(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{WV}}, s_{\mathrm{NP}} \mid x)$. Given $S(0) = \mathrm{WV}$, let T_{WV} denote the time instant when the server ends the current working vacation for the first time after time 0. Because of the memoryless property, T_{WV} is exponentially distributed with parameter γ. We classify the first passage time $T_{\mathrm{E|WV}}$ to the empty system into two cases, $T_{\mathrm{E|WV}} \le T_{\mathrm{WV}}$ and $T_{\mathrm{E|WV}} > T_{\mathrm{WV}}$, and we define $\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}} \mid x)$ and $\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha \mid x)$ as

$$\begin{aligned}
\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}} \mid x) &= \mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathcal{K}} z_{\mathrm{WV},k}^{N^{(\sigma)}_{\mathrm{WV},k}}\Bigg)\exp[-s_{\mathrm{WV}}T^{(\sigma)}_{\mathrm{E|WV}}] \\
&\qquad\qquad \Bigg|\, U(0)=x, S(0)=\mathrm{WV}, T_{\mathrm{E|WV}} \le T_{\mathrm{WV}}\Bigg], \\
\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha \mid x) &= \mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathcal{K}} z_{\mathrm{WV},k}^{N^{(\sigma)}_{\mathrm{WV},k}}\Bigg)\exp[-s_{\mathrm{WV}}T_{\mathrm{WV}}]\exp[-\alpha U(T_{\mathrm{WV}})] \\
&\qquad\qquad \Bigg|\, U(0)=x, S(0)=\mathrm{WV}, T_{\mathrm{E|WV}} > T_{\mathrm{WV}}\Bigg],
\end{aligned}$$

respectively. Note here that (2.31) implies

$$\begin{aligned}
&\mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathcal{K}} z_{\mathrm{NP},k}^{N^{(1)}_{\mathrm{WV},k}}\Bigg)\exp[-s_{\mathrm{NP}}T^{(1)}_{\mathrm{E|WV}}]\,\Bigg|\, U(0)=x, S(0)=\mathrm{WV}, T_{\mathrm{E|WV}} > T_{\mathrm{WV}}, U(T_{\mathrm{WV}})=y\Bigg] \\
&\quad = \mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathcal{K}} z_{\mathrm{NP},k}^{N_{\mathrm{NP},k}}\Bigg)\cdot\exp[-s_{\mathrm{NP}}T_{\mathrm{E|NP}}]\,\Bigg|\, U(0)=y, S(0)=\mathrm{NP}\Bigg] \\
&\quad = \zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}} \mid y) \\
&\quad = \exp[-q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}})y].
\end{aligned}$$

We then have

$$\begin{aligned}
\zeta^*_{\mathrm{WV}}(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{WV}}, s_{\mathrm{NP}} \mid x) &= P_{\mathrm{WV,WV}}(x)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}} \mid x) \\
&\quad + P_{\mathrm{WV,NP}}(x)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}}) \mid x),
\end{aligned} \tag{2.34}$$

where $P_{\mathrm{WV,WV}}(x)$ and $P_{\mathrm{WV,NP}}(x)$ are defined as

$$\begin{aligned}
P_{\mathrm{WV,WV}}(x) &= \Pr(T_{\mathrm{E|WV}} \le T_{\mathrm{WV}} \mid U(0)=x, S(0)=\mathrm{WV}), \\
P_{\mathrm{WV,NP}}(x) &= \Pr(T_{\mathrm{E|WV}} > T_{\mathrm{WV}} \mid U(0)=x, S(0)=\mathrm{WV}).
\end{aligned}$$

Lemma 2.7. *The following equations hold.*

$$P_{\mathrm{WV,WV}}(x)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x) = \exp[-q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})x], \tag{2.35}$$

$$P_{\mathrm{WV,NP}}(x)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x) = \frac{\exp[-\alpha x]-\exp[-q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})x]}{q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})-\alpha} \cdot q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha), \tag{2.36}$$

where $q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})$ *and* $q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha)$ *are defined as*

$$\begin{aligned}q^*_{\mathrm{WV,WV}}&(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})\\ &= \frac{s_{\mathrm{WV}}}{\sigma}+\frac{\gamma}{\sigma}+\frac{\lambda_{\mathrm{WV}}}{\sigma}\\ &\quad -\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\sigma}\int_0^\infty P_{\mathrm{WV,WV}}(y)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid y)dH_{\mathrm{WV},k}(y),\end{aligned} \tag{2.37}$$

$$\begin{aligned}q^*_{\mathrm{WV,NP}}&(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha)\\ &= \gamma/\sigma+\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\sigma}\int_0^\infty P_{\mathrm{WV,NP}}(y)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid y)dH_{\mathrm{WV},k}(y),\end{aligned} \tag{2.38}$$

and they satisfy

$$\begin{aligned}q^*_{\mathrm{WV,WV}}&(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})\\ &= \frac{s_{\mathrm{WV}}}{\sigma}+\frac{\gamma}{\sigma}+\frac{\lambda_{\mathrm{WV}}}{\sigma}-\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\sigma}\cdot h^*_{\mathrm{WV},k}\big(q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})\big),\end{aligned} \tag{2.39}$$

$$\begin{aligned}q^*_{\mathrm{WV,NP}}&(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha)\\ &= \frac{\gamma}{\sigma}+\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\sigma}\cdot\frac{h^*_{\mathrm{WV},k}(\alpha)-h^*_{\mathrm{WV},k}\big(q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})\big)}{q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})-\alpha}\\ &\quad\cdot q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha).\end{aligned} \tag{2.40}$$

The proof of Lemma 2.7 is given in Appendix 2.G.

It follows from (2.34), (2.35), and (2.36) that $\zeta^*_{\mathrm{WV}}(\boldsymbol{z}_{\mathrm{WV}},\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{WV}},s_{\mathrm{NP}}\mid x)$ is given by

$$\begin{aligned}\zeta^*_{\mathrm{WV}}(\boldsymbol{z}_{\mathrm{WV}},\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{WV}},s_{\mathrm{NP}}\mid x) &= \exp[-q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})x]\\ &\quad+\frac{\exp[-q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}})x]-\exp[-q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})x]}{q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})-q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}})}\\ &\quad\cdot q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}})).\end{aligned} \tag{2.41}$$

With (2.37) and (2.38), we define $q_{\mathrm{WV,WV}}$ and $q_{\mathrm{WV,NP}}$ as

$$q_{\mathrm{WV,WV}} = q^*_{\mathrm{WV,WV}}(\boldsymbol{e},0), \qquad q_{\mathrm{WV,NP}} = q^*_{\mathrm{WV,NP}}(\boldsymbol{e},0,0),$$

where $\boldsymbol{e}$ denotes a vector whose elements are all equal to one.

Lemma 2.8. *$q_{\mathrm{WV,WV}}$ and $q_{\mathrm{WV,NP}}$ are given by*

$$q_{\mathrm{WV,WV}} = q_{\mathrm{WV,NP}} = \frac{\gamma/\sigma}{1-\nu}, \tag{2.42}$$

and $P_{\mathrm{WV,WV}}(x)$ and $P_{\mathrm{WV,NP}}(x)$ are given by

$$P_{\mathrm{WV,WV}}(x) = \exp[-q_{\mathrm{WV,WV}} \cdot x], \qquad P_{\mathrm{WV,NP}}(x) = 1 - \exp[-q_{\mathrm{WV,WV}} \cdot x]. \tag{2.43}$$

The proof of Lemma 2.8 is given in Appendix 2.H.

We then consider the busy cycle. Recall that the server is always on working vacation at the beginning of busy cycle. Let Φ denote the length of a randomly chosen busy cycle. We divide Φ into two parts, and let $\Phi^{(\sigma)}$ (resp. $\Phi^{(1)}$) denote the length of the subinterval during which the server is on working vacation (resp. in a normal service period). Furthermore, we divide $\Phi^{(\sigma)}$ into two parts, and let $\Phi_{\mathrm{E}}^{(\sigma)}$ (resp. $\Phi_{\mathrm{B}}^{(\sigma)}$) denote the length of the subinterval during which the server is idle (resp. busy). By definition, $\Phi = \Phi_{\mathrm{E}}^{(\sigma)} + \Phi_{\mathrm{B}}^{(\sigma)} + \Phi^{(1)}$. For each k ($k \in \mathscr{K}$), let $\overline{N}_k^{(\sigma)}$ (resp. $\overline{N}_k^{(1)}$) denote the number of class k customers arriving in $\Phi^{(\sigma)}$ (resp. $\Phi^{(1)}$). We define the joint transform of those quantities as follows.

$$\begin{aligned}&\phi^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \omega, s_{\mathrm{WV}}, s_{\mathrm{NP}}) \\ &\quad = \mathrm{E}\left[\left(\prod_{k \in \mathscr{K}} z_{\mathrm{WV},k}^{\overline{N}_k^{(\sigma)}} \cdot z_{\mathrm{NP},k}^{\overline{N}_k^{(1)}}\right) \cdot \exp\big[-\omega \Phi_{\mathrm{E}}^{(\sigma)}\big] \cdot \exp\big[-s_{\mathrm{WV}} \Phi_{\mathrm{B}}^{(\sigma)}\big] \cdot \exp\big[-s_{\mathrm{NP}} \Phi^{(1)}\big]\right].\end{aligned}$$

By definition, $\phi^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \omega, s_{\mathrm{WV}}, s_{\mathrm{NP}})$ satisfies

$$\begin{aligned}&\phi^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \omega, s_{\mathrm{WV}}, s_{\mathrm{NP}}) \\ &\quad = \frac{\lambda_{\mathrm{WV}}}{\omega + \lambda_{\mathrm{WV}}} \sum_{k \in \mathscr{K}} \frac{z_{\mathrm{WV},k} \lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}} \int_0^\infty \zeta_{\mathrm{WV}}^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{WV}}, s_{\mathrm{NP}} \mid y) dH_{\mathrm{WV},k}(y).\end{aligned}$$

Therefore, with (2.41), we obtain the following theorem.

Theorem 2.6. *$\phi^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \omega, s_{\mathrm{WV}}, s_{\mathrm{NP}})$ is given by*

$$\begin{aligned}&\phi^*(\boldsymbol{z}_{\mathrm{WV}}, \boldsymbol{z}_{\mathrm{NP}}, \omega, s_{\mathrm{WV}}, s_{\mathrm{NP}}) \\ &\quad = \frac{\lambda_{\mathrm{WV}}}{\omega + \lambda_{\mathrm{WV}}} \sum_{k \in \mathscr{K}} \frac{z_{\mathrm{WV},k} \lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}} \Bigg[h_{\mathrm{WV},k}^*\big(q_{\mathrm{WV,WV}}^*(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}})\big) \\ &\qquad + \frac{h_{\mathrm{WV},k}^*(q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}})) - h_{\mathrm{WV},k}^*\big(q_{\mathrm{WV,WV}}^*(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}})\big)}{q_{\mathrm{WV,WV}}^*(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}) - q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}})} \\ &\qquad\qquad \cdot q_{\mathrm{WV,NP}}^*(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, q_{\mathrm{NP}}^*(\boldsymbol{z}_{\mathrm{NP}}, s_{\mathrm{NP}}))\Bigg].\end{aligned}$$

Remark 2.8. *It is clear from the derivation of Theorem 2.6 that*

$$\begin{aligned}&\Pr(\textit{A randomly chosen busy period ends while the server is on working vacation})\\&\quad= \lim_{\omega\to 0+}\frac{\lambda_{\mathrm{WV}}}{\omega+\lambda_{\mathrm{WV}}}\sum_{k\in\mathcal{K}}\frac{\lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}}\cdot h^*_{\mathrm{WV},k}\big(q^*_{\mathrm{WV,WV}}(\mathbf{0},0)\big) = h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}}) = r,\end{aligned}$$

where we use (2.54). This result is consistent with Remark 2.3. Furthermore, using (2.39) and (2.40), we obtain an alternative expression for $\phi^*(\boldsymbol{z}_{\mathrm{WV}},\boldsymbol{z}_{\mathrm{NP}},\omega,s_{\mathrm{WV}},s_{\mathrm{NP}})$.

$$\begin{aligned}\phi^*(\boldsymbol{z}_{\mathrm{WV}},\boldsymbol{z}_{\mathrm{NP}},\omega,s_{\mathrm{WV}},s_{\mathrm{NP}}) \;=\; \frac{\lambda_{\mathrm{WV}}}{\omega+\lambda_{\mathrm{WV}}}\bigg[\frac{1}{\lambda_{\mathrm{WV}}}\Big\{&s_{\mathrm{WV}}+\lambda_{\mathrm{WV}}-\sigma q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}})\\&+\sigma q^*_{\mathrm{WV,NP}}\big(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}})\big)\Big\}\bigg].\end{aligned}$$

Taking the partial derivatives of $\phi^*(\boldsymbol{z}_{\mathrm{WV}},\boldsymbol{z}_{\mathrm{NP}},\omega,s_{\mathrm{WV}},s_{\mathrm{NP}})$, we can obtain the moments of $\overline{N}_k^{(\sigma)}$, $\overline{N}_k^{(1)}$, $\Phi_{\mathrm{B}}^{(\sigma)}$, and $\Phi^{(1)}$. In particular,

$$\begin{aligned}\mathrm{E}[\overline{N}_k^{(\sigma)}] &= \lambda_{\mathrm{WV},k}\cdot\left(\frac{1}{\lambda_{\mathrm{WV}}}+\mathrm{E}[\Phi_{\mathrm{B}}^{(\sigma)}]\right), &\qquad \mathrm{E}[\overline{N}_k^{(1)}] &= \lambda_{\mathrm{NP},k}\cdot\mathrm{E}[\Phi^{(1)}],\\ \mathrm{E}[\Phi_{\mathrm{B}}^{(\sigma)}] &= \frac{\sigma q_{\mathrm{WV,WV}}}{\gamma\lambda_{\mathrm{WV}}}-\frac{1}{\lambda_{\mathrm{WV}}}, & \mathrm{E}[\Phi^{(1)}] &= (1-r)\cdot\frac{\mathrm{E}[U_{\mathrm{WV}}]/\nu}{1-\rho_{\mathrm{NP}}}.\end{aligned}$$

2.6 Conclusion

In this chapter, we considered the stationary multi-class FCFS M/G/1 queue with exponential working vacations. We first analyzed the stationary workload in system. Based on it, we derived the LSTs of the stationary waiting time and sojourn time in each class, and further obtained the joint transform of the queue lengths and the workloads in respective classes. Moreover, we derived the joint transform associated with the busy cycle.

As stated in Section 2.1, if we delete time intervals in normal service periods from the time axis, the resulting process can be viewed as a multi-class FCFS M/G/1 queue with Poisson disasters, where the processing rate is equal to σ. Appendix 2.I summarizes the analytical results for the multi-class FCFS M/G/1 queue with Poisson disasters, all of which are immediately obtained from the results in this chapter. These results are generalized to the case of the multi-class FCFS MAP/G/1 queue in Chapter 3.

Appendices

2.A Proof of Lemma 2.1

We define $U_{\mathrm{NP}}^{\mathrm{B}}$ as the total workload in system at the beginning of a normal service period. Note that $U_{\mathrm{NP}}^{\mathrm{B}}$ is a conditional random variable of $U_{\mathrm{WV}}^{\mathrm{E}}$ given that the server is busy at the end of a working vacation. Let $u^*_{\mathrm{NP,B}}(s)$ denote the LST of $U^{\mathrm{B}}_{\mathrm{NP}}$. We then have

$$u^*_{\mathrm{NP,B}}(s) = \mathrm{E}\left[\exp[-sU^{\mathrm{E}}_{\mathrm{WV}}] \mid U^{\mathrm{E}}_{\mathrm{WV}} > 0\right] = \frac{u^*_{\mathrm{WV,E}}(s) - \Pr(U^{\mathrm{E}}_{\mathrm{WV}} = 0)}{1 - \Pr(U^{\mathrm{E}}_{\mathrm{WV}} = 0)}, \qquad (2.44)$$

$$\mathrm{E}\left[U^{\mathrm{B}}_{\mathrm{NP}}\right] = \frac{\mathrm{E}\left[U^{\mathrm{E}}_{\mathrm{WV}}\right]}{1 - \Pr(U^{\mathrm{E}}_{\mathrm{WV}} = 0)}. \qquad (2.45)$$

Consider a censored workload process by removing all working vacation periods from the time axis. In steady state, the censored process has the same distribution as U_{NP}. Also, the censored process can be viewed as the conditional workload process of the M/G/1 vacation queue with exhaustive services, given that the server is busy. Therefore, it follows from (5.6) in [Dos90] that $u^*_{\mathrm{NP}}(s)$ is given by

$$u^*_{\mathrm{NP}}(s) = \frac{1 - u^*_{\mathrm{NP,B}}(s)}{s\mathrm{E}\left[U^{\mathrm{B}}_{\mathrm{NP}}\right]} \cdot u^*_{\mathrm{M/G/1}}(s).$$

Note here that (2.2), (2.44), and (2.45) imply

$$\frac{1 - u^*_{\mathrm{NP,B}}(s)}{s\mathrm{E}\left[U^{\mathrm{B}}_{\mathrm{NP}}\right]} = \frac{1 - u^*_{\mathrm{WV,E}}(s)}{s\mathrm{E}\left[U^{\mathrm{E}}_{\mathrm{WV}}\right]} = \frac{1 - u^*_{\mathrm{WV}}(s)}{s\mathrm{E}[U_{\mathrm{WV}}]},$$

which completes the proof.

2.B Proof of Lemma 2.2

We regard an interval between successive ends of working vacations as a cycle. Let C_{WV} (resp. C_{NP}) denote the length of an interval during which the server is on working vacation (resp. in a normal service period) in a randomly chosen cycle. Owing to the renewal reward theorem, we have

$$p_{\mathrm{WV}} = \frac{\mathrm{E}[C_{\mathrm{WV}}]}{\mathrm{E}[C_{\mathrm{WV}}] + \mathrm{E}[C_{\mathrm{NP}}]}, \qquad p_{\mathrm{NP}} = \frac{\mathrm{E}[C_{\mathrm{NP}}]}{\mathrm{E}[C_{\mathrm{WV}}] + \mathrm{E}[C_{\mathrm{NP}}]}. \qquad (2.46)$$

Because C_{WV} is equivalent to the working vacation length V, we have $\mathrm{E}[C_{\mathrm{WV}}] = \mathrm{E}[V]$. On the other hand, $\mathrm{E}[C_{\mathrm{NP}}]$ equals to the mean first passage time to the empty system in the corresponding ordinary M/G/1 queue with initial workload of $U^{\mathrm{B}}_{\mathrm{NP}}$. Noting that $C_{\mathrm{NP}} = 0$ if the system is empty at the end of the working vacation, we have

$$\mathrm{E}[C_{\mathrm{NP}}] = \Pr(U^{\mathrm{E}}_{\mathrm{WV}} = 0) \cdot 0 + \{1 - \Pr(U^{\mathrm{E}}_{\mathrm{WV}} = 0)\} \cdot \frac{\mathrm{E}[U^{\mathrm{B}}_{\mathrm{NP}}]}{1 - \rho_{\mathrm{NP}}} = \frac{\mathrm{E}[U_{\mathrm{WV}}]}{1 - \rho_{\mathrm{NP}}}, \qquad (2.47)$$

where we use (2.2) and (2.45). (2.5) now follows from (2.46), (2.47), and $\mathrm{E}[C_{\mathrm{WV}}] = \mathrm{E}[V] = 1/\gamma$.

2.C Proof of Lemma 2.3

The censored process obtained by removing all normal service periods is considered as an M/G/1 queue with Poisson disasters with rate γ, where the system becomes empty when disasters occur. The M/G/1 queue with Poisson disasters has already been studied in [JS96, YKC02], where the processing rate is assumed to be one. In order to apply the results in [JS96, YKC02] to our system, we consider the new process created by extending the time axis of the workload process in working vacation periods σ times so that the processing rate becomes one. Note that the time-average quantities of the new censored process are identical to those of the original process. In the new process, the arrival rate of customers is equal to $\lambda_{\mathrm{WV}}/\sigma$ and lengths of working vacations are exponentially distributed with parameter γ/σ. $u^*_{\mathrm{WV}}(s)$ in (2.7) then immediately follows from Proposition 1 in [JS96]. We also obtain (2.8) by substituting 0 to the repair time in (2.1a) in [YKC02]. The existence of the unique real root of (2.9) is shown in Remark 2.2 in [YKC02]. See Remark 2.3 for the positivity of r. Furthermore, taking the derivative of $u^*_{\mathrm{WV}}(s)$ in (2.7) and evaluating at $s=0$ yields $\mathrm{E}[U_{\mathrm{WV}}]$ in (2.7).

2.D Proof of Lemma 2.5

We first consider $\overline{\rho}^{(\sigma)}_{\mathrm{WV},k}$. Note that all customers being served in working vacation periods arrived in working vacation periods. Thus, from Little's law, we have $\overline{\rho}^{(\sigma)}_{\mathrm{WV},k} = \lambda_{\mathrm{WV},k}P_{\mathrm{WV}} \cdot \mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV}}]$. Furthermore, with Lemma 2.3 and (2.21), $\mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV}}]$ is obtained to be

$$\begin{aligned}\mathrm{E}\left[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV}}\right] &= \mathrm{E}\left[T^{(\sigma)}_{\mathrm{WV}}(U_{\mathrm{WV}}+H_{\mathrm{WV},k})\right] - \mathrm{E}\left[T^{(\sigma)}_{\mathrm{WV}}(U_{\mathrm{WV}})\right] \\ &= \frac{u^*_{\mathrm{WV}}(\gamma/\sigma)\big(1-h^*_{\mathrm{WV},k}(\gamma/\sigma)\big)}{\gamma} = \frac{\nu}{\lambda_{\mathrm{WV}}} \cdot \frac{1-h^*_{\mathrm{WV},k}(\gamma/\sigma)}{1-h^*_{\mathrm{WV}}(\gamma/\sigma)}, \end{aligned} \tag{2.48}$$

from which (2.24) follows.

Similarly, $\overline{\rho}^{(1)}_{\mathrm{WV},k}$ follows from $\overline{\rho}^{(1)}_{\mathrm{WV},k} = \lambda_{\mathrm{WV},k}P_{\mathrm{WV}} \cdot \mathrm{E}[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV}}]$ and

$$\begin{aligned}\mathrm{E}\left[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV}}\right] &= \mathrm{E}\left[T^{(1)}_{\mathrm{WV}}(U_{\mathrm{WV}}+H_{\mathrm{WV},k})\right] - \mathrm{E}\left[T^{(1)}_{\mathrm{WV}}(U_{\mathrm{WV}})\right] \\ &= \mathrm{E}[H_{\mathrm{WV},k}] - \frac{\sigma\nu}{\lambda_{\mathrm{WV}}} \cdot \frac{1-h^*_{\mathrm{WV},k}(\gamma/\sigma)}{1-h^*_{\mathrm{WV}}(\gamma/\sigma)}. \end{aligned} \tag{2.49}$$

Finally, we consider $\overline{\rho}^{(1)}_{\mathrm{NP},k}$. Note that all customers arriving in normal service periods are served in normal service periods. Therefore $\overline{\rho}^{(1)}_{\mathrm{NP},k} = \lambda_{\mathrm{NP},k}P_{\mathrm{NP}} \cdot \mathrm{E}[H_{\mathrm{NP},k}] = P_{\mathrm{NP}} \cdot \rho_{\mathrm{NP},k}$, from which (2.26) follows.

2.E Proof of Theorem 2.4

We first consider (2.27). Suppose a class k ($k \in \mathscr{K}$) customer is being served at processing rate σ (i.e., in a working vacation period). Note here that

$$\mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k} \mid \overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k} > 0] = \frac{\mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k}]}{u^*_{\mathrm{WV}}(\gamma/\sigma)}.$$

We thus have

$$\begin{aligned}
&a^{**}_{\mathrm{WV},\mathrm{WV},k}(\omega_k, \alpha_k) \\
&\quad = \frac{1}{\dfrac{\mathrm{E}[\overline{W}^{(\sigma)}_{\mathrm{WV},k} - W^{(\sigma)}_{\mathrm{WV},k}]}{u^*_{\mathrm{WV}}(\gamma/\sigma)}} \cdot \frac{u^*_{\mathrm{WV}}\left(\dfrac{\omega_k + \gamma}{\sigma}\right)}{u^*_{\mathrm{WV}}(\gamma/\sigma)} \\
&\qquad \cdot \int_0^\infty dH_{\mathrm{WV},k}(x) \Bigg[\exp[-\gamma(x/\sigma)] \int_0^{x/\sigma} \exp[-\omega_k t] \exp[-\alpha_k(x - \sigma t)] dt \\
&\qquad\qquad + \int_0^{x/\sigma} \gamma \exp[-\gamma\tau] d\tau \int_0^\tau \exp[-\omega_k t] \exp[-\alpha_k(x - \sigma t)] dt \Bigg],
\end{aligned}$$

from which (2.27) follows.

Next we consider (2.29). Suppose a class k ($k \in \mathscr{K}$) customer, who arrived in a working vacation period, is being served at processing rate one (i.e., in a normal service period). We then have

$$\mathrm{E}[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k} \mid \overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k} > 0] = \frac{\mathrm{E}[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k}]}{1 - u^*_{\mathrm{WV}}(\gamma/\sigma) h^*_{\mathrm{WV},k}(\gamma/\sigma)}.$$

Therefore

$$\begin{aligned}
&a^{***}_{\mathrm{WV},\mathrm{NP},k}(\omega_k, s_k, \alpha_k) \\
&\quad = \frac{1}{\dfrac{\mathrm{E}[\overline{W}^{(1)}_{\mathrm{WV},k} - W^{(1)}_{\mathrm{WV},k}]}{1 - u^*_{\mathrm{WV}}(\gamma/\sigma) h^*_{\mathrm{WV},k}(\gamma/\sigma)}} \\
&\quad \cdot \Bigg[\frac{u^*_{\mathrm{WV}}(\gamma/\sigma)(1 - h^*_{\mathrm{WV},k}(\gamma/\sigma))}{1 - u^*_{\mathrm{WV}}(\gamma/\sigma) h^*_{\mathrm{WV},k}(\gamma/\sigma)} \cdot \frac{u^*_{\mathrm{WV}}\left(\dfrac{\omega_k + \gamma}{\sigma}\right)}{u^*_{\mathrm{WV}}(\gamma/\sigma)} \cdot \frac{1}{1 - h^*_{\mathrm{WV},k}(\gamma/\sigma)} \\
&\qquad \cdot \int_0^\infty dH_{\mathrm{WV},k}(x) \int_0^{x/\sigma} \gamma \exp[-\gamma\tau] d\tau \int_\tau^{\tau + x - \sigma\tau} \exp[-\omega_k \tau] \exp[-s_k(t - \tau)] \\
&\qquad\qquad\qquad\qquad \cdot \exp[-\alpha_k(x - \sigma\tau - (t - \tau))] dt
\end{aligned}$$

$$+\frac{1-u^*_{\mathrm{WV}}(\gamma/\sigma)}{1-u^*_{\mathrm{WV}}(\gamma/\sigma)h^*_{\mathrm{WV},k}(\gamma/\sigma)}\cdot\frac{1}{1-u^*_{\mathrm{WV}}(\gamma/\sigma)}\cdot\frac{u^*_{\mathrm{WV}}(s_k)-u^*_{\mathrm{WV}}\left(\dfrac{\omega_k+\gamma}{\sigma}\right)}{(\sigma/\gamma)\{(\omega_k+\gamma)/\sigma-s_k\}}$$
$$\left.\cdot\int_0^\infty dH_{\mathrm{WV},k}(x)\int_0^x \exp[-s_k t]\exp[-\alpha_k(x-t)]dt\right],$$

from which (2.29) follows.

Finally, $a^{**}_{\mathrm{NP},k}(s_k,\alpha_k)$ is given by

$$a^{**}_{\mathrm{NP},k}(s_k,\alpha_k)=\frac{1}{\mathrm{E}[H_{\mathrm{NP},k}]}\cdot u^*_{\mathrm{NP},k}(s)\int_0^\infty dH_{\mathrm{NP},k}(x)\int_0^x \exp[-s_k t]\exp[-\alpha_k(x-t)]dt,$$

from which (2.30) follows.

2.F Proof of Lemma 2.6

For $x\ge 0$, $y\ge 0$, we have $\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x+y)=\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x)\cdot\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid y)$. Therefore

$$\begin{aligned}
&\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x+\Delta x)\\
&\quad=\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x)\cdot\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid \Delta x)\\
&\quad=\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x)\Big[1-s_{\mathrm{NP}}\Delta x-\lambda_{\mathrm{NP}}\Delta x+\lambda_{\mathrm{NP}}\Delta x\sum_{k\in\mathcal{K}} z_{\mathrm{NP},k}\\
&\qquad\qquad\cdot\frac{\lambda_{\mathrm{NP},k}}{\lambda_{\mathrm{NP}}}\int_0^\infty \zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid y)dH_{\mathrm{NP},k}(y)+o(\Delta x)\Big],
\end{aligned}$$

from which it follows that

$$\frac{\partial}{\partial x}\Big[\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x)\Big]=-\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid x)q^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}).$$

Noting $\zeta^*_{\mathrm{NP}}(\boldsymbol{z}_{\mathrm{NP}},s_{\mathrm{NP}}\mid 0)=1$, we obtain (2.31). Also, substituting (2.31) into (2.32), we obtain (2.33).

2.G Proof of Lemma 2.7

It is easy to see that for $x\ge 0$, $y\ge 0$,

$$\begin{aligned}
&P_{\mathrm{WV,WV}}(x+y)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x+y)\\
&\quad=P_{\mathrm{WV,WV}}(x)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\cdot P_{\mathrm{WV,WV}}(y)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid y),\\
&P_{\mathrm{WV,NP}}(x+y)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x+y)\\
&\quad=P_{\mathrm{WV,NP}}(x)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\cdot\exp[-\alpha y]\\
&\qquad+P_{\mathrm{WV},WV}(x)\cdot\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\cdot P_{\mathrm{WV,NP}}(y)\cdot\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid y).
\end{aligned}$$

Therefore we have

$$
\begin{aligned}
&P_{\mathrm{WV,WV}}(x+\Delta x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x+\Delta x)\\
&\quad= P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\cdot P_{\mathrm{WV,WV}}(\Delta x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid\Delta x)\\
&\quad= P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\\
&\qquad\cdot\Bigg[1-s_{\mathrm{WV}}\frac{\Delta x}{\sigma}-\gamma\frac{\Delta x}{\sigma}-\lambda_{\mathrm{WV}}\frac{\Delta x}{\sigma}\\
&\qquad\quad+\lambda_{\mathrm{WV}}\frac{\Delta x}{\sigma}\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}}\int_0^{\infty}P_{\mathrm{WV,WV}}(y)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid y)dH_{\mathrm{WV},k}(y)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad+o(\Delta x)\Bigg],
\end{aligned}
$$

from which it follows that

$$
\begin{aligned}
&\frac{\partial}{\partial x}\Big[P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\Big]\\
&\qquad= -P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\cdot q^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}).
\end{aligned}\tag{2.50}
$$

(2.35) now follows from (2.50) with $P_{\mathrm{WV,WV}}(0)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid 0)=1$.

Similarly,

$$
\begin{aligned}
&P_{\mathrm{WV,NP}}(x+\Delta x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x+\Delta x)\\
&\quad= P_{\mathrm{WV,NP}}(x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\cdot\exp[-\alpha\Delta x]\\
&\qquad+P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\cdot P_{\mathrm{WV,NP}}(\Delta x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid\Delta x)\\
&\quad= P_{\mathrm{WV,NP}}(x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\cdot(1-\alpha\Delta x)+o(\Delta x)\\
&\qquad+P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)\\
&\qquad\quad\cdot\Bigg[\gamma\frac{\Delta x}{\sigma}+\lambda_{\mathrm{WV}}\frac{\Delta x}{\sigma}\sum_{k\in\mathcal{K}}\frac{z_{\mathrm{WV},k}\lambda_{\mathrm{WV},k}}{\lambda_{\mathrm{WV}}}\\
&\qquad\qquad\qquad\cdot\int_0^{\infty}P_{\mathrm{WV,NP}}(y)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid y)dH_{\mathrm{WV},k}(y)+o(\Delta x)\Bigg],
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&\frac{\partial}{\partial x}\Big[P_{\mathrm{WV,NP}}(x)\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\Big]\\
&\qquad= -\alpha P_{\mathrm{WV,NP}}(x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\\
&\qquad\quad+P_{\mathrm{WV,WV}}(x)\cdot\zeta^{*}_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}}\mid x)q^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha).
\end{aligned}\tag{2.51}
$$

Multiplying both sides of (2.51) by $\exp[\alpha x]$ and using (2.35) yield

$$
\frac{\partial}{\partial x}\Big[P_{\mathrm{WV,NP}}(x)\cdot\zeta^{*}_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}},s_{\mathrm{WV}},\alpha\mid x)\cdot\exp[\alpha x]\Big]
$$

$$= \exp[-q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}})x] \cdot q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha) \cdot \exp[\alpha x].$$

Because $P_{\mathrm{WV,NP}}(0) = 0$, we obtain

$$\begin{aligned} & P_{\mathrm{WV,NP}}(x) \cdot \zeta^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha \mid x) \cdot \exp[\alpha x] \\ & \quad = \int_0^x \exp[-\{q^*_{\mathrm{WV,WV}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}) - \alpha\}y] \cdot q^*_{\mathrm{WV,NP}}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha) dy, \end{aligned}$$

from which (2.36) follows. Substituting (2.35) into (2.37) yields (2.39), and substituting (2.36) into (2.38) yields (2.40).

2.H Proof of Lemma 2.8

(2.43) follows from $\zeta^*_{\mathrm{WV,WV}}(\boldsymbol{e}, 0 \mid x) = 1$, $P_{\mathrm{WV,WV}}(x) + P_{\mathrm{WV,NP}}(x) = 1$, and (2.35). We thus consider (2.42) below. Note that $\zeta^*_{\mathrm{WV,NP}}(\boldsymbol{e}, 0, 0 \mid x) = 1$. Therefore, taking the limits $\alpha \to 0+$ and $s \to 0+$ in (2.36), we obtain

$$P_{\mathrm{WV,NP}}(x) = \frac{q_{\mathrm{WV,NP}}}{q_{\mathrm{WV,WV}}} \cdot (1 - \exp[-q_{\mathrm{WV,WV}} \cdot x]),$$

from which and (2.43), we have $q_{\mathrm{WV,WV}} = q_{\mathrm{WV,NP}}$.

It is readily seen from (2.39) that $q_{\mathrm{WV,WV}}$ satisfies

$$q_{\mathrm{WV,WV}} = \gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma) h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}}), \tag{2.52}$$

and $h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}}) = h^*_{\mathrm{WV}}\big(\gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma) h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}})\big)$. Furthermore, it follows from (2.37) that

$$\begin{aligned} q_{\mathrm{WV,WV}} &= \gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - \sum_{k \in K} (\lambda_{\mathrm{WV},k}/\sigma) \int_0^\infty P_{\mathrm{WV,WV}}(y) dH_{\mathrm{WV},k}(y) \\ &= \gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma) \int_0^\infty P_{\mathrm{WV,WV}}(y) dH_{\mathrm{WV}}(y) \geq \gamma/\sigma > 0, \qquad (2.53) \end{aligned}$$

so that $|h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}})| < 1$. As a result, $h^*_{\mathrm{WV}}(q_{\mathrm{WV,WV}})$ is identical to the minimum non-negative root r of (2.9). Finally, from (2.8) and (2.52), we obtain

$$\gamma/\sigma + \lambda_{\mathrm{WV}}/\sigma - (\lambda_{\mathrm{WV}}/\sigma) r = \frac{\gamma/\sigma}{1 - \nu}, \tag{2.54}$$

which completes the proof.

2.I The multi-class FCFS M/G/1 queue with Poisson disasters

In this Appendix, we summarize the results of the stationary multi-class FCFS M/G/1 queue with Poisson disasters, where the processing rate is equal to one. We can readily obtain those results by considering the conditional counterparts in the multi-class FCFS M/G/1 with exponential working vacations and $\sigma = 1$, given that the server is on working vacation.

2.I.1 Model

Consider a stationary multi-class FCFS M/G/1 queue with Poisson disasters. Class k ($k \in \mathcal{K}$) customers arrive according to a Poisson process with rate λ_k. Let $h_k(x)$ and $h_k^*(s)$ ($k \in \mathcal{K}$) denote the distribution function of service times H_k of class k customers and its LST, respectively. Disasters occur according to a Poisson process with rate γ ($\gamma > 0$), and the system becomes empty when disasters occur. We define λ and $h^*(s)$ as

$$\lambda = \sum_{k \in \mathcal{K}} \lambda_k, \qquad h^*(s) = \sum_{k \in \mathcal{K}} \frac{\lambda_k}{\lambda} \cdot h_k^*(s).$$

Note that if we ignore customer classes, the system can be regarded as a single-class FCFS M/G/1 queue with Poisson disasters. Note also that the system is stable regardless of values of system parameters.

2.I.2 Results

The LST $u^*(s)$ of the workload in system is given by [JS96, YKC02] (cf. Lemma 2.3 and its proof)

$$u^*(s) = \frac{(1-\nu)s - \gamma}{s - \lambda + \lambda h^*(s) - \gamma}.$$

Note that ν denotes the stationary probability of the server being busy.

$$\nu = \frac{(1-r)\lambda}{(1-r)\lambda + \gamma}, \tag{2.55}$$

where r denotes the minimum non-negative root of the following equation.

$$z = h^*(\gamma + \lambda - \lambda z), \qquad |z| < 1. \tag{2.56}$$

We denote the workload in system seen by a randomly chosen customer on arrival by U_A, and the length of the interval from the arrival of this customer to the occurrence of the next disaster by $\tilde{D}_\mathrm{A}$. Owing to the memoryless property, $\tilde{D}_\mathrm{A}$ is exponentially distributed with parameter γ. We define W_k and $\overline{W}_k$ ($k \in \mathcal{K}$) as the waiting time and sojourn time, respectively, of class k customers, i.e., $W_k = \min(U_\mathrm{A}, \tilde{D}_\mathrm{A})$ and $\overline{W}_k = \min(U_\mathrm{A} + H_k, \tilde{D}_\mathrm{A})$. Note that owing to PASTA, W_k ($k \in \mathcal{K}$) is identical to the waiting time W of a randomly chosen customer. Furthermore, we define

$$\begin{aligned} P_\mathrm{N}^\mathrm{W} &= \Pr(U_\mathrm{A} \le \tilde{D}_\mathrm{A}), & P_\mathrm{D}^\mathrm{W} &= \Pr(U_\mathrm{A} > \tilde{D}_\mathrm{A}), \\ w_\mathrm{N}^*(s) &= \mathrm{E}\left[\exp[-sW] \mid U_\mathrm{A} \le \tilde{D}_\mathrm{A}\right], & w_\mathrm{D}^*(s) &= \mathrm{E}\left[\exp[-sW] \mid U_\mathrm{A} > \tilde{D}_\mathrm{A}\right], \end{aligned}$$

and for each k ($k \in \mathcal{K}$)

$$P_{\mathrm{N},k}^\mathrm{Q} = \Pr(U_\mathrm{A} + H_k \le \tilde{D}_\mathrm{A}), \qquad P_{\mathrm{D},k}^\mathrm{Q} = \Pr(U_\mathrm{A} + H_k > \tilde{D}_\mathrm{A}),$$

$$\overline{w}^*_{\mathrm{N},k}(s) = \mathrm{E}\left[\exp[-s\overline{W}_k] \mid U_{\mathrm{A}} + H_k \le \tilde{D}_{\mathrm{A}}\right], \quad \overline{w}^*_{\mathrm{D}}(s) = \mathrm{E}\left[\exp[-s\overline{W}_k] \mid U_{\mathrm{A}} + H_k > \tilde{D}_{\mathrm{A}}\right].$$

By definition, we have

$$\begin{aligned} w^*(s) &= \mathrm{E}[\exp[-sW]] = P^{\mathrm{W}}_{\mathrm{N}} w^*_{\mathrm{N}}(s) + P^{\mathrm{W}}_{\mathrm{D}} w^*_{\mathrm{D}}(s), \\ \overline{w}^*_k(s) &= \mathrm{E}[\exp[-s\overline{W}_k]] = P^{\mathrm{Q}}_{\mathrm{N},k} \overline{w}^*_{\mathrm{N},k}(s) + P^{\mathrm{Q}}_{\mathrm{D},k} \overline{w}^*_{\mathrm{D},k}(s). \end{aligned}$$

Because W corresponds to $W^{(\sigma)}_{\mathrm{WV}}$ in the queue with working vacations, we obtain from Theorem 2.2

$$w^*(s) = u^*(s+\gamma) + \frac{1 - u^*(s+\gamma)}{(1/\gamma)(s+\gamma)}. \tag{2.57}$$

Note here that $u^*(s+\gamma) = P^{\mathrm{W}}_{\mathrm{N}} w^*_{\mathrm{N}}(s)$. Therefore the second term on the right hand side of (2.57) represents $P^{\mathrm{W}}_{\mathrm{D}} w^*_{\mathrm{D}}(s)$. It then follows that

$$\begin{aligned} w^*_{\mathrm{N}}(s) &= \frac{u^*(s+\gamma)}{u^*(\gamma)}, \qquad w^*_{\mathrm{D}}(s) = \frac{1}{1-u^*(\gamma)} \cdot \frac{1-u^*(s+\gamma)}{(1/\gamma)(s+\gamma)}, \\ P^{\mathrm{W}}_{\mathrm{N}} &= u^*(\gamma), \qquad P^{\mathrm{W}}_{\mathrm{D}} = 1 - u^*(\gamma). \end{aligned}$$

Similarly, it follows from Theorem 2.3 that

$$\overline{w}^*_k(s) = u^*(s+\gamma)h^*_k(s+\gamma) + \frac{1 - u^*(s+\gamma)h^*_k(s+\gamma)}{(1/\gamma)(s+\gamma)}, \qquad k \in \mathscr{K},$$

and therefore for each k ($k \in \mathscr{K}$)

$$\begin{aligned} \overline{w}^*_{\mathrm{N},k}(s) &= \frac{u^*(s+\gamma)h^*_k(s+\gamma)}{u^*(\gamma)h^*_k(\gamma)}, \qquad \overline{w}^*_{\mathrm{D},k}(s) = \frac{1}{1-u^*(\gamma)h^*_k(\gamma)} \cdot \frac{1-u^*(s+\gamma)h^*_k(s+\gamma)}{(1/\gamma)(s+\gamma)}, \\ P^{\mathrm{Q}}_{\mathrm{N},k} &= u^*(\gamma)h^*_k(\gamma), \qquad P^{\mathrm{Q}}_{\mathrm{D},k} = 1 - u^*(\gamma)h^*_k(\gamma). \end{aligned}$$

Let $\overline{\rho}_k$ ($k \in \mathscr{K}$) denote the probability of a class k customer being served, which corresponds to $\overline{\rho}^{(\sigma)}_{\mathrm{WV},k}/P_{\mathrm{WV}}$ in the queue with working vacations. It follows from (2.24) that

$$\overline{\rho}_k = \nu \cdot \frac{\lambda_k(1-h^*_k(\gamma))}{\lambda(1-h^*(\gamma))},$$

where ν is given in (2.55).

Let A_k ($k \in \mathscr{K}$) denote the conditional attained waiting time given that a class k customer is being served and let $\tilde{H}_k$ ($k \in \mathscr{K}$) denote the remaining service time of class k customer being served. We then define $a^{**}_k(s_k, \alpha_k)$ ($k \in \mathscr{K}$) as

$$a^{**}_k(s_k, \alpha_k) = \mathrm{E}\left[\exp[-s_k A_k] \cdot \exp[-\alpha_k \tilde{H}_k] \mid \text{a class } k \text{ customer is being served}\right].$$

Note that $a_k^{**}(s_k,\alpha_k)$ corresponds to $a_{\mathrm{WV,WV},k}^{**}(\omega_k,\alpha_k)$ in the queue with working vacations. Moreover, $\overline{W}_k$ and W_k corresponds to $\overline{W}_{\mathrm{WV},k}^{(\sigma)}$ and $W_{\mathrm{WV},k}^{(\sigma)}$, respectively. It then follows from (2.27) that

$$a_k^{**}(s_k,\alpha_k) = \frac{u^*(s_k+\gamma)}{\mathrm{E}[\overline{W}_k - W_k]} \cdot \frac{h_k^*(\alpha_k) - h_k^*\left(s_k+\gamma\right)}{s_k+\gamma-\alpha_k},$$

where $\mathrm{E}[\overline{W}_k - W_k]$ is obtained from (2.48).

$$\mathrm{E}[\overline{W}_k - W_k] = \frac{\nu}{\lambda} \cdot \frac{1-h_k^*(\gamma)}{1-h^*(\gamma)}.$$

Let L_k ($k \in \mathscr{K}$) denote the number of class k customers in the system and let U_k ($k \in \mathscr{K}$) denote the workload in system belonging to class k. We then define the joint transform $\psi(\boldsymbol{z},\boldsymbol{s})$ as

$$\psi(\boldsymbol{z},\boldsymbol{s}) = \mathrm{E}\left[\prod_{k\in\mathscr{K}} z_k^{L_k} \cdot \exp[-s_k U_k]\right],$$

where $\boldsymbol{z} = (z_1, z_2, \ldots, z_K)$ and $\boldsymbol{s} = (s_1, s_2, \ldots, s_K)$. We then have

$$\psi(\boldsymbol{z},\boldsymbol{s}) = 1-\nu+\sum_{k\in\mathscr{K}} z_k \overline{\rho}_k a_k^{**}\Big(\sum_{i\in\mathscr{K}}[\lambda_i - \lambda_i z_i h_i^*(s_i)], s_k\Big),$$

which corresponds to Theorem 2.5.

Finally, we consider the busy cycle, which is defined as the interval between successive ends of busy periods. Let Φ denote the length of a randomly chosen busy cycle. We divide Φ into two parts, and let Φ_{E} (resp. Φ_{B}) denote the length of the subinterval during which the server is idle (resp. busy). We define $\overline{N}_k$ ($k \in \mathscr{K}$) as the number of class k customers arriving in Φ. Let $\tilde{U}_{\mathrm{L}}$ denote the workload in system that is lost due to disasters. We then define joint transforms $\phi_{\mathrm{N}}^*(\boldsymbol{z},s)$ and $\phi_{\mathrm{D}}^*(\boldsymbol{z},s,\alpha)$ as follows.

$$\begin{aligned}\phi_{\mathrm{N}}^*(\boldsymbol{z},\omega,s) &= \mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathscr{K}} z^{\overline{N}_k}\Bigg)\cdot\exp[-\omega\Phi_{\mathrm{E}}]\cdot\exp[-s\Phi_{\mathrm{B}}] \\ &\qquad\qquad \Bigg|\,\text{a busy period ends without disasters}\Bigg], \\ \phi_{\mathrm{D}}^*(\boldsymbol{z},\omega,s,\alpha) &= \mathrm{E}\Bigg[\Bigg(\prod_{k\in\mathscr{K}} z^{\overline{N}_k}\Bigg)\cdot\exp[-\omega\Phi_{\mathrm{E}}]\cdot\exp[-s\Phi_{\mathrm{B}}]\cdot\exp[-\alpha\tilde{U}_{\mathrm{L}}] \\ &\qquad\qquad \Bigg|\,\text{a busy period ends with disasters}\Bigg].\end{aligned}$$

We also define $P_{\mathrm{N}}^{\mathrm{B}}$ as

$$P_{\mathrm{N}}^{\mathrm{B}} = \Pr(\text{a busy period ends without disasters}),$$

and let $P_{\mathrm{D}}^{\mathrm{B}} = 1 - P_{\mathrm{N}}^{\mathrm{B}}$. It then follows from Lemma 2.7 that

$$\begin{aligned}
P_{\mathrm{N}}^{\mathrm{B}} \cdot \phi_{\mathrm{N}}^{*}(\boldsymbol{z},\omega,s) &= \frac{\lambda}{\lambda+\omega} \sum_{k\in\mathcal{K}} z_k \cdot \frac{\lambda_k}{\lambda} \int_0^{\infty} \exp[-q_{\mathrm{N}}^{*}(\boldsymbol{z},s)y] dH_k(y) \\
&= \frac{\lambda}{\lambda+\omega} \sum_{k\in\mathcal{K}} z_k \cdot \frac{\lambda_k h_k^{*}\big(q_{\mathrm{N}}^{*}(\boldsymbol{z},s)\big)}{\lambda} \qquad (2.58) \\
&= \frac{\lambda}{\lambda+\omega} \cdot \frac{s+\gamma+\lambda-q_{\mathrm{N}}^{*}(\boldsymbol{z},s)}{\lambda}, \qquad (2.59) \\
P_{\mathrm{D}}^{\mathrm{B}} \cdot \phi_{\mathrm{N}}^{*}(\boldsymbol{z},\omega,s,\alpha) &= \frac{\lambda}{\lambda+\omega} \sum_{k\in\mathcal{K}} z_k \cdot \frac{\lambda_k}{\lambda} \int_0^{\infty} \frac{\exp[-\alpha y] - \exp[-q_{\mathrm{N}}^{*}(\boldsymbol{z},s)y]}{q_{\mathrm{N}}^{*}(\boldsymbol{z},s)-\alpha} \\
&\qquad \cdot q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha) dH_k(y) \\
&= \frac{\lambda}{\lambda+\omega} \sum_{k\in\mathcal{K}} z_k \cdot \frac{\lambda_k}{\lambda} \cdot \frac{h_k^{*}(\alpha) - h_k^{*}\big(q_{\mathrm{N}}^{*}(\boldsymbol{z},s)\big)}{q_{\mathrm{N}}^{*}(\boldsymbol{z},s)-\alpha} \cdot q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha) \\
&= \frac{\lambda}{\lambda+\omega} \cdot \frac{q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha)-\gamma}{\lambda}, \qquad (2.60)
\end{aligned}$$

where $q_{\mathrm{N}}^{*}(\boldsymbol{z},s)$ and $q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha)$ satisfy

$$\begin{aligned}
q_{\mathrm{N}}^{*}(\boldsymbol{z},s) &= s+\gamma+\lambda - \sum_{k\in\mathcal{K}} z_k \lambda_k h_k^{*}(q_{\mathrm{N}}^{*}(\boldsymbol{z},s)), \\
q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha) &= \gamma + \sum_{k\in\mathcal{K}} z_k \lambda_k \cdot \frac{h_k^{*}(\alpha) - h_k^{*}(q_{\mathrm{N}}^{*}(\boldsymbol{z},s))}{q_{\mathrm{N}}^{*}(\boldsymbol{z},s)-\alpha} \cdot q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha),
\end{aligned}$$

which correspond to $q_{\mathrm{WV,WV}}^{*}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}})$ in (2.37) and $q_{\mathrm{WV,NP}}^{*}(\boldsymbol{z}_{\mathrm{WV}}, s_{\mathrm{WV}}, \alpha)$ in (2.38), respectively.

We define q_{N} and q_{D} as

$$q_{\mathrm{N}} = q_{\mathrm{N}}^{*}(\boldsymbol{e},0), \qquad q_{\mathrm{D}} = q_{\mathrm{D}}^{*}(\boldsymbol{e},0,0).$$

We then have (cf. Lemma 2.8 and its proof)

$$q_{\mathrm{N}} = q_{\mathrm{D}} = \frac{\gamma}{1-\nu}, \qquad h^{*}(q_{\mathrm{N}}) = r,$$

where r is the minimum non-negative root of (2.56). It then follows from (2.58) that

$$P_{\mathrm{N}}^{\mathrm{B}} = r, \qquad P_{\mathrm{D}}^{\mathrm{B}} = 1-r,$$

and from (2.59) and (2.60) that

$$\phi_{\mathrm{N}}^{*}(\boldsymbol{z},\omega,s) = \frac{\lambda}{\lambda+\omega} \cdot \frac{s+\gamma+\lambda-q_{\mathrm{N}}^{*}(\boldsymbol{z},s)}{\gamma+\lambda-q_{\mathrm{N}}}, \qquad \phi_{\mathrm{D}}^{*}(\boldsymbol{z},\omega,s,\alpha) = \frac{\lambda}{\lambda+\omega} \cdot \frac{q_{\mathrm{D}}^{*}(\boldsymbol{z},s,\alpha)-\gamma}{q_{\mathrm{D}}-\gamma}.$$

3 Multi-Class MAP/G/1 Queue with Disasters

3.1 Introduction

In this chapter, we consider the multi-class FCFS MAP/G/1 queue with disasters. When disasters occur, all workload in system is removed instantaneously and the system becomes empty. There are K classes of customers, which are labeled one to K. Let $\mathscr{K}$ denote $\mathscr{K} = \{1,2,\ldots,K\}$. We assume that the disaster occurrence process and the customer arrival process are governed by a common continuous-time underlying Markov chain. We assume that the underlying Markov chain is irreducible and has finite state-space $\mathscr{M} = \{1,2,\ldots,M\}$. The underlying Markov chain stays in state i ($i \in \mathscr{M}$) for an exponential interval of time with mean $1/\sigma_i$, where $\sigma_i > 0$ ($i \in \mathscr{M}$). When the sojourn time in state i is elapsed, the chain changes its state to j ($j \in \mathscr{M}, j \neq i$) with probability $p_{i,j}$ without customer arrivals and disasters. Also, with probability $q_{k,i,j}$ (resp. $\gamma_{i,j}$), it changes its state from i to j ($j \in \mathscr{M}$) and a class k ($k \in \mathscr{K}$) customer arrives (resp. a disaster occurs). We assume that the arrival of a customer and the occurrence of a disaster do not occur simultaneously. We thus have

$$\sum_{j\in\mathscr{M}} \left[p_{i,j} + \Big(\sum_{k\in\mathscr{K}} q_{k,i,j}\Big) + \gamma_{i,j}\right] = 1, \qquad i \in \mathscr{M},$$

where $p_{i,i} = 0$ ($i \in \mathscr{M}$).

We assume that the amounts of service requirements brought by class k ($k \in \mathscr{K}$) customers who arrive with state transition from state i to state j ($i,j \in \mathscr{M}$) are i.i.d. according to a general distribution function $H_{k,i,j}(x)$ ($x \geq 0$).

To deal with the underlying Markov chain described above, we introduce $M \times M$ matrices $\boldsymbol{C}$, $\boldsymbol{D}_k(x)$ ($k \in \mathscr{K}, x \geq 0$), and $\boldsymbol{\Gamma}$, whose (i,j)-th ($i,j \in \mathscr{M}$) elements are given by

$$[\boldsymbol{C}]_{i,j} = \begin{cases} \sigma_i p_{i,j}, & i \neq j, \\ -\sigma_i, & i = j, \end{cases} \qquad [\boldsymbol{D}_k(x)]_{i,j} = \sigma_i q_{k,i,j} H_{k,i,j}(x), \qquad [\boldsymbol{\Gamma}]_{i,j} = \sigma_i \gamma_{i,j},$$

respectively. Let $\boldsymbol{D}_k^*(s)$ ($k \in \mathscr{K}$, $\mathrm{Re}(s) > 0$) denote the LST of $\boldsymbol{D}_k(x)$.

$$\boldsymbol{D}_k^*(s) = \int_0^\infty \exp[-sx] d\boldsymbol{D}_k(x).$$

We define $\boldsymbol{D}_k$ ($k \in \mathscr{K}$) as

$$\boldsymbol{D}_k = \lim_{x\to\infty} \boldsymbol{D}_k(x) = \lim_{s\to 0+} \boldsymbol{D}_k^*(s).$$

Furthermore, we define $\boldsymbol{D}(x)$ ($x \geq 0$) as

$$\boldsymbol{D}(x) = \sum_{k\in\mathscr{K}} \boldsymbol{D}_k(x),$$

and $\boldsymbol{D}^*(s)$ ($\mathrm{Re}(s) > 0$) as its LST.

$$\boldsymbol{D}^*(s) = \int_0^\infty \exp[-sx] d\boldsymbol{D}(x).$$

We also define $\boldsymbol{D}$ as

$$\boldsymbol{D} = \lim_{x\to\infty} \boldsymbol{D}(x) = \lim_{s\to 0+} \boldsymbol{D}^*(s).$$

Note here that

$$\boldsymbol{D}^*(s) = \sum_{k\in\mathscr{K}} \boldsymbol{D}_k^*(s), \qquad \boldsymbol{D} = \sum_{k\in\mathscr{K}} \boldsymbol{D}_k.$$

By definition, $\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{\Gamma}$ represents the infinitesimal generator of the underlying Markov chain, and it satisfies

$$(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{\Gamma})\boldsymbol{e} = \boldsymbol{0}, \tag{3.1}$$

where $\boldsymbol{e}$ denotes an $M \times 1$ vector whose elements are all equal to one. Let $\boldsymbol{\pi}$ denote the stationary probability vector of the underlying Markov chain. Because the underlying Markov chain is irreducible and has finite state-space $\mathscr{M}$, $\boldsymbol{\pi} > \boldsymbol{0}$ and it is determined uniquely by

$$\boldsymbol{\pi}(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{\Gamma}) = \boldsymbol{0}, \qquad \boldsymbol{\pi}\boldsymbol{e} = 1.$$

We assume that $\boldsymbol{D}_k \neq \boldsymbol{0}$ ($k \in \mathscr{K}$) and $\boldsymbol{\Gamma} \neq \boldsymbol{0}$. Because the system becomes empty when a disaster occur, $\boldsymbol{\Gamma} \neq \boldsymbol{0}$ and the irreducibility of the underlying Markov chain ensure the existence of the steady state. We assume that the service discipline is nonpreemptive FCFS, unless otherwise mentioned.

Remark 3.1. *$\boldsymbol{C}$ represents the defective infinitesimal generator of the underlying Markov chain when neither arrivals nor disasters occur. Also, $\boldsymbol{C} + \boldsymbol{D}$ (resp. $\boldsymbol{C} + \boldsymbol{\Gamma}$) represents the defective infinitesimal generator of the underlying Markov chain when no disasters occur (resp. no arrivals occur). Therefore $\boldsymbol{C}$, $\boldsymbol{C} + \boldsymbol{D}$, and $\boldsymbol{C} + \boldsymbol{\Gamma}$ are non-singular, and all eigenvalues of those matrices have strictly negative real parts. It then follows that $s\boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{D})$ and $s\boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{\Gamma})$ are also non-singular for all s (*$\mathrm{Re}(s) >$ *0).*

We first analyze the first passage time to the idle state, the busy cycle, and the total workload process. We then derive two different representations of the LST of the total stationary workload in system and discuss the relation between those. Note that similar formulas are derived in [Tak02] and [TH94] for the ordinary MAP/G/1 queue without disasters. Next, using the results on the total workload distribution, we analyze the waiting time and the sojourn time distributions of each class, and the joint queue length distribution.

The rest of this chapter is organized as follows. In Section 3.2, the first passage time to the idle state is analyzed. In Section 3.3, we analyze the busy cycle and derive the stationary probability that the system is empty. In Section 3.4, the total stationary workload in system is studied. In Section 3.5, we analyze the waiting time and sojourn time distributions. In Section 3.6, we derive the joint probability generating function of the queue length distribution, and consider the computational procedure of the joint queue length distribution. In Section 3.7, we briefly summarize the computational procedure for performance measures of interest and provide an numerical example. Finally, we conclude this chapter in Section 3.8.

3.2 First passage time to the idle state

Let U_t $(t \geq 0)$ denote the total workload in system at time t. We define T_{E} as the first passage time to the idle state after time 0.

$$T_{\mathrm{E}} = \inf\{t; U_t = 0, t > 0\}.$$

Let S_t $(t \geq 0)$ denote the state of the underlying Markov chain at time t. We then define $\boldsymbol{P}(t \mid x)$ $(t \geq 0,\ x \geq 0)$ as an $M \times M$ matrix whose (i,j)-th $(i,j \in \mathscr{M})$ element represents the conditional joint probability that the first passage time is not greater than t and the underlying Markov chain is in state j at the end of the first passage time, given that at time 0, the total workload in system is equal to x and the underlying Markov chain is in state i.

$$[\boldsymbol{P}(t \mid x)]_{i,j} = \Pr(T_{\mathrm{E}} \leq t, S_{T_{\mathrm{E}}} = j \mid U_0 = x, S_0 = i).$$

Let $\boldsymbol{P}^*(s \mid x)$ $(\mathrm{Re}(s) > 0, x \geq 0)$ denote the LST of $\boldsymbol{P}(t \mid x)$ with respect to t.

$$\boldsymbol{P}^*(s \mid x) = \int_{t=0}^{\infty} \exp[-st] d\boldsymbol{P}(t \mid x),$$

where the (i,j)th $(i,j \in \mathscr{M})$ element of $d\boldsymbol{P}(t \mid x)$ represents $\Pr(t < T_{\mathrm{E}} \leq t + dt, S_{T_{\mathrm{E}}} = j \mid U_0 = x, S_0 = i)$. We classify the first passage process by whether or not a disaster occurs in the first passage time. Let T_{D} denote the time when a disaster occurs for the first time after time 0. We define $\boldsymbol{P}_{\mathrm{N}}(t \mid x)$ $(t \geq 0,\ x \geq 0)$ as an $M \times M$ matrix whose (i,j)-th $(i,j \in \mathscr{M})$ element represents the conditional joint probability that no disaster occurs in the first passage time, the first passage time is not greater than t, and the underlying Markov chain is in state j at the end of the first passage time, given that at time 0, the total workload in system is equal to x and the underlying Markov chain is in state i.

$$[\boldsymbol{P}_{\mathrm{N}}(t \mid x)]_{i,j} = \Pr(T_{\mathrm{E}} \le t, T_{\mathrm{E}} < T_{\mathrm{D}}, S_{T_{\mathrm{E}}} = j \mid U_0 = x, S_0 = i).$$

We also define $\boldsymbol{P}_{\mathrm{D}}(t \mid x)$ ($t \ge 0$, $x \ge 0$) as an $M \times M$ matrix whose (i,j)-th $(i,j \in \mathscr{M})$ element represents the conditional joint probability that the first passage time to the idle state ends with a disaster, the first passage time is not greater than t, and the underlying Markov chain is in state j at the end of the first passage time, given that at time 0, the total workload in system is equal to x and the underlying Markov chain is in state i.

$$[\boldsymbol{P}_{\mathrm{D}}(t \mid x)]_{i,j} = \Pr(T_{\mathrm{E}} \le t, T_{\mathrm{E}} = T_{\mathrm{D}}, S_{T_{\mathrm{E}}} = j \mid U_0 = x, S_0 = i).$$

Let $\boldsymbol{P}^*_{\mathrm{N}}(s \mid x)$ and $\boldsymbol{P}^*_{\mathrm{D}}(s \mid x)$ ($\mathrm{Re}(s) > 0, x \ge 0$) denote the LSTs of $\boldsymbol{P}_{\mathrm{N}}(t \mid x)$ and $\boldsymbol{P}_{\mathrm{D}}(t \mid x)$, respectively, with respect to t.

$$\boldsymbol{P}^*_{\mathrm{N}}(s \mid x) = \int_{t=0}^{\infty} \exp[-st] d\boldsymbol{P}_{\mathrm{N}}(t \mid x), \quad \boldsymbol{P}^*_{\mathrm{D}}(s \mid x) = \int_{t=0}^{\infty} \exp[-st] d\boldsymbol{P}_{\mathrm{D}}(t \mid x).$$

By definition, we have $\boldsymbol{P}^*_{\mathrm{N}}(s \mid 0) = \boldsymbol{I}$, $\boldsymbol{P}^*_{\mathrm{D}}(s \mid 0) = \boldsymbol{0}$, and

$$\boldsymbol{P}^*(s \mid x) = \boldsymbol{P}^*_{\mathrm{N}}(s \mid x) + \boldsymbol{P}^*_{\mathrm{D}}(s \mid x). \tag{3.2}$$

Lemma 3.1. *$\boldsymbol{P}^*_{\mathrm{N}}(s \mid x)$ ($\mathrm{Re}(s) > 0$, $x \ge 0$) satisfies*

$$\boldsymbol{P}^*_{\mathrm{N}}(s \mid x) = \exp\left[\boldsymbol{Q}^*_{\mathrm{N}}(s)x\right], \tag{3.3}$$

where $\boldsymbol{Q}^*_{\mathrm{N}}(s)$ ($\mathrm{Re}(s) > 0$) is defined as

$$\boldsymbol{Q}^*_{\mathrm{N}}(s) = -s\boldsymbol{I} + \boldsymbol{C} + \int_0^{\infty} d\boldsymbol{D}(y)\boldsymbol{P}^*_{\mathrm{N}}(s \mid y). \tag{3.4}$$

On the other hand, $\boldsymbol{P}^*_{\mathrm{D}}(s \mid x)$ ($\mathrm{Re}(s) > 0$, $x \ge 0$) satisfies

$$\boldsymbol{P}^*_{\mathrm{D}}(s \mid x) = \int_0^{x} \exp\left[\boldsymbol{Q}^*_{\mathrm{N}}(s)w\right] dw \cdot \boldsymbol{Q}^*_{\mathrm{D}}(s), \tag{3.5}$$

where $\boldsymbol{Q}^*_{\mathrm{D}}(s)$ ($\mathrm{Re}(s) > 0$) is defined as

$$\boldsymbol{Q}^*_{\mathrm{D}}(s) = \boldsymbol{\Gamma} + \int_0^{\infty} d\boldsymbol{D}(y)\boldsymbol{P}^*_{\mathrm{D}}(s \mid y). \tag{3.6}$$

The proof of Lemma 3.1 is given in Appendix 3.A.

We define $\boldsymbol{Q}_{\mathrm{N}}$ and $\boldsymbol{Q}_{\mathrm{D}}$ as $M \times M$ matrices given by

$$\boldsymbol{Q}_{\mathrm{N}} = \lim_{s \to 0+} \boldsymbol{Q}^*_{\mathrm{N}}(s), \qquad \boldsymbol{Q}_{\mathrm{D}} = \lim_{s \to 0+} \boldsymbol{Q}^*_{\mathrm{D}}(s),$$

respectively. It then follows from (3.4) and (3.6) that

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{N}} &= \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(y)\boldsymbol{P}^*_{\mathrm{N}}(0 \mid y), & (3.7) \\ \boldsymbol{Q}_{\mathrm{D}} &= \boldsymbol{\Gamma} + \int_0^\infty d\boldsymbol{D}(y)\boldsymbol{P}^*_{\mathrm{D}}(0 \mid y), & (3.8) \end{aligned}$$

where

$$\boldsymbol{P}^*_{\mathrm{N}}(0 \mid x) = \lim_{s \to 0+} \boldsymbol{P}^*_{\mathrm{N}}(s \mid x), \qquad \boldsymbol{P}^*_{\mathrm{D}}(0 \mid x) = \lim_{s \to 0+} \boldsymbol{P}^*_{\mathrm{D}}(s \mid x).$$

Using (3.3) and (3.5), we rewrite (3.7) and (3.8) as

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{N}} &= \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(y)\exp[\boldsymbol{Q}_{\mathrm{N}} y], & (3.9) \\ \boldsymbol{Q}_{\mathrm{D}} &= \boldsymbol{\Gamma} + \int_0^\infty d\boldsymbol{D}(y) \int_0^y \exp[\boldsymbol{Q}_{\mathrm{N}} w] dw \cdot \boldsymbol{Q}_{\mathrm{D}}, & (3.10) \end{aligned}$$

respectively.

Lemma 3.2. $\boldsymbol{Q}_{\mathrm{N}}$ *is non-singular and* $\boldsymbol{Q}_{\mathrm{D}}$ *is given by*

$$\boldsymbol{Q}_{\mathrm{D}} = (-\boldsymbol{Q}_{\mathrm{N}})[-(\boldsymbol{C} + \boldsymbol{D})]^{-1}\boldsymbol{\Gamma}. \tag{3.11}$$

The proof of Lemma 3.2 is given in Appendix 3.B.

Note here that $[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}$ appeared in (3.11) represents the transition probability matrix of the embedded Markov chain obtained by observing the state of the underlying Markov chain immediately after disasters, and therefore $[-(\boldsymbol{C} + \boldsymbol{D})]^{-1}\boldsymbol{\Gamma}\boldsymbol{e} = \boldsymbol{e}$, which can also be verified with (3.1). We then have from (3.11)

$$(\boldsymbol{Q}_{\mathrm{N}} + \boldsymbol{Q}_{\mathrm{D}})\boldsymbol{e} = \boldsymbol{0}. \tag{3.12}$$

In addition, it can be shown with (3.7) and (3.8) that off-diagonal elements of $\boldsymbol{Q}_{\mathrm{N}}$ and all elements of $\boldsymbol{Q}_{\mathrm{D}}$ are non-negative. Therefore $\boldsymbol{Q}_{\mathrm{N}} + \boldsymbol{Q}_{\mathrm{D}}$ can be interpreted as the infinitesimal generator of a continuous-time Markov chain defined on finite state-space $\mathcal{M}$.

The probabilistic interpretation of $\boldsymbol{Q}_{\mathrm{N}}$ and $\boldsymbol{Q}_{\mathrm{D}}$ is as follows. Consider a censored underlying Markov chain obtained by removing all busy periods from the time axis. The first term $\boldsymbol{C}$ on the right-hand side of (3.7) represents the defective infinitesimal generator of the underlying Markov chain when neither arrivals nor disasters occur. On the other hand, the integral on the right-hand side of (3.7) represents the transition rate matrix when busy periods without disasters are removed. Similarly, the first term $\boldsymbol{\Gamma}$ on the right-hand side of (3.8) represents the transition rate matrix when disasters occur in the idle state, and the integral on the right-hand side of (3.8) represents the transition rate matrix when busy periods ending with disasters are removed. Therefore $\boldsymbol{Q}_{\mathrm{N}}+\boldsymbol{Q}_{\mathrm{D}}$ represents the infinitesimal generator of the censored underlying Markov chain obtained by removing all busy periods from the time axis.

Let $\boldsymbol{\kappa}$ denote a $1 \times M$ vector whose j-th ($j \in \mathcal{M}$) element represents the conditional probability that the underlying Markov chain is in state j, given that the system is empty in steady state.

$$[\boldsymbol{\kappa}]_j = \lim_{t\to\infty} \Pr(S_t = j \mid U_t = 0).$$

By definition, $\boldsymbol{\kappa}$ represents the steady state probability vector of the censored underlying Markov chain obtained by removing all busy periods from the time axis. Because the original underlying Markov chain is irreducible, the censored underlying Markov chain is also irreducible. Therefore $\boldsymbol{\kappa}$ is determined uniquely by

$$\boldsymbol{\kappa}(\boldsymbol{Q}_{\mathrm{N}}+\boldsymbol{Q}_{\mathrm{D}}) = \boldsymbol{0}, \qquad \boldsymbol{\kappa}\boldsymbol{e} = 1, \tag{3.13}$$

and the irreducibility of the censored underlying Markov chain implies

$$\boldsymbol{\kappa} > \boldsymbol{0}. \tag{3.14}$$

We now consider the computational procedure of $\boldsymbol{Q}_{\mathrm{N}}$. For this purpose, we define $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$ ($n = 0,1,\ldots$) as an $M \times M$ matrix defined by the following recursion:

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{N}}^{(0)} &= \boldsymbol{C}, \\ \boldsymbol{Q}_{\mathrm{N}}^{(n)} &= \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(y)\exp[\boldsymbol{Q}_{\mathrm{N}}^{(n-1)}y], \quad n = 1,2,\ldots. \end{aligned} \tag{3.15}$$

Lemma 3.3. $\{\boldsymbol{Q}_{\mathrm{N}}^{(n)};\ n = 0,1,\ldots\}$ *is an elementwise increasing sequence of matrices and* $\boldsymbol{Q}_{\mathrm{N}}$ *is given by*

$$\boldsymbol{Q}_{\mathrm{N}} = \lim_{n\to\infty} \boldsymbol{Q}_{\mathrm{N}}^{(n)}.$$

The proof of Lemma 3.3 is given in Appendix 3.C

Because $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$ is regarded as a defective infinitesimal generator, the integral on the right-hand side of (3.15) can be computed with uniformization [Tij94, Page 154], so that we can numerically obtain $\boldsymbol{Q}_{\mathrm{N}} = \lim_{n\to\infty}\boldsymbol{Q}_{\mathrm{N}}^{(n)}$.

Remark 3.2. *Because* $\boldsymbol{Q}_{\mathrm{N}}$ *represents a defective infinitesimal generator, the stopping criterion for computing* $\boldsymbol{Q}_{\mathrm{N}}$ *is not clear in general. We address this issue under a more general condition in Section 4.4.2.*

Let $\boldsymbol{f}(x)$ $(x \geq 0)$ denote an $M \times 1$ vector whose i-th $(i \in \mathscr{M})$ element represents the mean first passage time to the idle state, given that at time 0, the workload in system is equal to x and the underlying Markov chain is in state i.

$$[\boldsymbol{f}(x)]_i = \mathrm{E}[T_{\mathrm{E}} \mid U_0 = x, S_0 = i].$$

By definition, we have

$$\boldsymbol{f}(x) = (-1) \cdot \lim_{s \to 0+} \frac{\partial}{\partial s} \big[\boldsymbol{P}^*(s \mid x)\big] \boldsymbol{e}. \tag{3.16}$$

Lemma 3.4. $\boldsymbol{f}(x)$ *is given by*

$$\boldsymbol{f}(x) = [\boldsymbol{I} - \exp[\boldsymbol{Q}_{\mathrm{N}} x]][-(\boldsymbol{C} + \boldsymbol{D})]^{-1} \boldsymbol{e}. \tag{3.17}$$

The proof of Lemma 3.4 is given in Appendix 3.D

3.3 Busy cycle

This section considers the busy cycle, which is defined as the interval between successive ends of busy periods. Let Φ denote the length of a busy cycle which starts at time 0. We divide Φ into two parts, i.e., $\Phi = \Phi_{\mathrm{E}} + \Phi_{\mathrm{B}}$, where Φ_{E} (resp. Φ_{B}) denotes the length of a subinterval during which the server is idle (resp. busy).

We define $\boldsymbol{\Phi}(\tau, t)$ as an $M \times M$ matrix whose (i,j)th $(i,j \in \mathscr{M})$ element represents the conditional joint probability that in a busy cycle, the length of the idle period is not greater than τ, the length of the subsequent busy period is not greater than t, and the underlying Markov chain is in state j at the end of the busy cycle, given that the underlying Markov chain is in state i at the beginning of the busy cycle.

$$[\boldsymbol{\Phi}(\tau, t)]_{i,j} = \Pr(\Phi_{\mathrm{E}} \leq \tau, \Phi_{\mathrm{B}} \leq t, S_{\Phi} = j \mid S_0 = i),$$

We then define $\boldsymbol{\Phi}^{**}(\omega, s)$ $(\mathrm{Re}(\omega) > 0$, $\mathrm{Re}(s) > 0)$ as the joint LST of $\boldsymbol{\Phi}(\tau, t)$.

$$\boldsymbol{\Phi}^{**}(\omega, s) = \int_{\tau=0}^{\infty} \int_{t=0}^{\infty} \exp[-\omega\tau] \exp[-st] d\boldsymbol{\Phi}(\tau, t),$$

where the (i,j)th $(i,j \in \mathscr{M})$ element of $d\boldsymbol{\Phi}(\tau, t)$ represents $\Pr(\tau < \Phi_{\mathrm{E}} \leq \tau + d\tau, t < \Phi_{\mathrm{B}} \leq t + dt, S_{\Phi} = j \mid S_0 = i)$. By definition, $\boldsymbol{\Phi}^{**}(\omega, s)$ is given by

$$\begin{aligned} \boldsymbol{\Phi}^{**}(\omega, s) &= \int_0^{\infty} \exp[-\omega\tau] \exp\big[(\boldsymbol{C} + \boldsymbol{\Gamma})\tau\big] d\tau \int_0^{\infty} d\boldsymbol{D}(y) \boldsymbol{P}^*(s \mid y) \\ &= [\omega \boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} \int_0^{\infty} d\boldsymbol{D}(y) \boldsymbol{P}^*(s \mid y). \end{aligned} \tag{3.18}$$

Recall that $\omega \boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{\Gamma})$ $(\mathrm{Re}(\omega) > 0)$ is non-singular (see Remark 3.1).

Let $\boldsymbol{\phi}$ denote an $M \times 1$ vector whose i-th ($i \in \mathcal{M}$) element represents the mean length of a busy cycle, given that the underlying Markov chain is in state i at the beginning of the busy cycle. Also, let $\boldsymbol{\phi}_{\mathrm{E}}$ denote an $M \times 1$ vector whose i-th ($i \in \mathcal{M}$) element represents the mean length of the idle period in a busy cycle, given that the underlying Markov chain is in state i at the beginning of the busy cycle.

$$[\boldsymbol{\phi}]_i = \mathrm{E}[\Phi \mid S_0 = i], \qquad [\boldsymbol{\phi}_{\mathrm{E}}]_i = \mathrm{E}[\Phi_{\mathrm{E}} \mid S_0 = i].$$

By definition, we have

$$\boldsymbol{\phi} = (-1) \cdot \lim_{s \to 0+} \frac{\partial}{\partial s} \big[\boldsymbol{\Phi}^{**}(s,s)\big] \boldsymbol{e}, \tag{3.19}$$

$$\boldsymbol{\phi}_{\mathrm{E}} = (-1) \cdot \lim_{\omega \to 0+} \lim_{s \to 0+} \frac{\partial}{\partial \omega} \big[\boldsymbol{\Phi}^{**}(\omega,s)\big] \boldsymbol{e}. \tag{3.20}$$

Note that the i-th ($i \in \mathcal{M}$) element of $\boldsymbol{\phi}_{\mathrm{E}}$ represents the mean length of time elapsed before the first customer arrives after time 0, given that the underlying Markov chain is in state i at time 0. We thus have

$$\boldsymbol{\phi}_{\mathrm{E}} = \int_0^{\infty} \exp[(\boldsymbol{C} + \boldsymbol{\Gamma})\tau] \boldsymbol{e} d\tau = [-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} \boldsymbol{e}. \tag{3.21}$$

Lemma 3.5. *$\boldsymbol{\phi}$ is given by*

$$\boldsymbol{\phi} = [-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} (-\boldsymbol{Q}_{\mathrm{N}}) [-(\boldsymbol{C} + \boldsymbol{D})]^{-1} \boldsymbol{e}. \tag{3.22}$$

The proof of Lemma 3.5 is given in Appendix 3.E.

Let $\boldsymbol{\eta}$ denote a $1 \times M$ vector whose j-th ($j \in \mathcal{M}$) element represents the probability that the underlying Markov chain is in state j at the beginning of a busy cycle. By definition, $\boldsymbol{\eta}\boldsymbol{e} = 1$ and $\boldsymbol{\eta}$ satisfies

$$\begin{aligned}
\boldsymbol{\eta} &= \boldsymbol{\eta} \lim_{s \to 0+} \lim_{\omega \to 0+} \boldsymbol{\Phi}^{**}(\omega,s) \\
&= \boldsymbol{\eta}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} \int_0^{\infty} d\boldsymbol{D}(y) \big[\boldsymbol{P}_{\mathrm{N}}^{*}(0 \mid y) + \boldsymbol{P}_{\mathrm{D}}^{*}(0 \mid y)\big] \\
&= \boldsymbol{\eta}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} [\boldsymbol{Q}_{\mathrm{N}} + \boldsymbol{Q}_{\mathrm{D}} - (\boldsymbol{C} + \boldsymbol{\Gamma})],
\end{aligned}$$

where we use (3.2), (3.7), (3.8), and (3.18). We then have

$$\boldsymbol{\eta}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1} (\boldsymbol{Q}_{\mathrm{N}} + \boldsymbol{Q}_{\mathrm{D}}) = \boldsymbol{0},$$

which implies that $\boldsymbol{\eta}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]^{-1}$ is a real multiple of $\boldsymbol{\kappa}$ defined in (3.13), and therefore we have

$$\boldsymbol{\eta} = \frac{\boldsymbol{\kappa}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]}{\boldsymbol{\kappa}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]\boldsymbol{e}}. \tag{3.23}$$

It then follows from (3.21), (3.22), and (3.23) that

$$\begin{aligned} \mathrm{E}[\Phi] &= \boldsymbol{\eta}\boldsymbol{\phi} = \frac{\boldsymbol{\kappa}(-\boldsymbol{Q}_{\mathrm{N}})[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}}{\boldsymbol{\kappa}[-(\boldsymbol{C}+\boldsymbol{\Gamma})]\boldsymbol{e}}, \\ \mathrm{E}[\Phi_{\mathrm{E}}] &= \boldsymbol{\eta}\boldsymbol{\phi}_{\mathrm{E}} = \frac{1}{\boldsymbol{\kappa}[-(\boldsymbol{C}+\boldsymbol{\Gamma})]\boldsymbol{e}}. \end{aligned}$$

Let ν denote the time-average probability that the server is busy. Because of the ergodicity of the system, ν is also regarded as the limiting probability of the server being busy.

$$\nu = \lim_{t\to\infty} \Pr(U_t > 0).$$

Owing to the renewal reward theorem [Cin75], we have $\nu = 1 - \mathrm{E}[\Phi_{\mathrm{E}}]/\mathrm{E}[\Phi]$. Therefore we obtain the following lemma.

Lemma 3.6. *ν is given by*

$$\nu = 1 - \frac{1}{\boldsymbol{\kappa}(-\boldsymbol{Q}_{\mathrm{N}})[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}}. \tag{3.24}$$

3.4 Total workload in system

Let $\boldsymbol{u}_t(x)$ ($t \ge 0$, $x \ge 0$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the joint probability that the total workload in system is not greater than x and the underlying Markov chain is in state j at time t.

$$[\boldsymbol{u}_t(x)]_j = \Pr(U_t \le x, S_t = j).$$

We define $\boldsymbol{u}(x)$ ($x \ge 0$) as

$$\boldsymbol{u}(x) = \lim_{t\to\infty} \boldsymbol{u}_t(x),$$

and $\boldsymbol{u}^*(s)$ ($\mathrm{Re}(s) > 0$) as its LST.

$$\boldsymbol{u}^*(s) = \int_0^\infty \exp[-sx] d\boldsymbol{u}(x).$$

Theorem 3.1. *$\boldsymbol{u}^*(s)$ satisfies*

$$\boldsymbol{u}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = s(1-\nu)\boldsymbol{\kappa} - \boldsymbol{\pi}\boldsymbol{\Gamma}, \tag{3.25}$$

where $\boldsymbol{\kappa}$ and ν are given in (3.13) and (3.24), respectively.

The proof of Theorem 3.1 is given in Appendix 3.F.

Differentiating both sides of (3.25) with respect to s, taking the limit $s \to 0+$, and rearranging terms, we have

$$-\boldsymbol{u}'(0) = \left[\boldsymbol{\pi}(-\boldsymbol{D}'(0)) - \{\boldsymbol{\pi} - (1-\nu)\boldsymbol{\kappa}\}\right][-(\boldsymbol{C}+\boldsymbol{D})]^{-1}, \tag{3.26}$$

where

$$\boldsymbol{u}'(0) = \lim_{s\to 0+} \frac{d}{ds}\left[\boldsymbol{u}^*(s)\right], \qquad \boldsymbol{D}'(0) = \lim_{s\to 0+} \frac{d}{ds}\left[\boldsymbol{D}^*(s)\right].$$

Therefore, the mean stationary workload E$[U]$ in system is given by

$$\mathrm{E}[U] = \left[\boldsymbol{\pi}(-\boldsymbol{D}'(0)) - \{\boldsymbol{\pi} - (1-\nu)\boldsymbol{\kappa}\}\right][-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}.$$

Next, we derive an alternative formula for $\boldsymbol{u}^*(s)$, following the same approach as in [Tak01]. For a while, we assume that customers are served under the LCFS-PR basis. Note that this service discipline is work conserving. Let $\hat{L}_t$ $(t \ge 0)$ denote the number of customers in the system at time t under the LCFS-PR service discipline. We define $\boldsymbol{u}_t(x,n)$ $(t \ge 0,\ x \ge 0,\ n = 0,1,\ldots)$ as a $1 \times M$ vector whose j-th $(j \in \mathscr{M})$ element represents the joint probability that at time t, the total workload in system is not greater than x, the number of customers in the system is equal to n, and the underlying Markov chain is in state j.

$$[\boldsymbol{u}_t(x,n)]_j = \Pr(U_t \le x, \hat{L}_t = n, S_t = j).$$

We define $\boldsymbol{u}(x,n)$ $(x \ge 0,\ n = 0,1,\ldots)$ as

$$\boldsymbol{u}(x,n) = \lim_{t\to\infty} \boldsymbol{u}_t(x,n).$$

By definition, we have

$$\boldsymbol{u}(x,0) = \boldsymbol{u}(0) = (1-\nu)\boldsymbol{\kappa}, \quad x \ge 0,$$

and

$$\boldsymbol{u}(x) = \sum_{n=0}^{\infty} \boldsymbol{u}(x,n).$$

Let $\boldsymbol{u}^*(s,n)$ $(\mathrm{Re}(s) > 0,\ n = 0,1,\ldots)$ denote the LST of $\boldsymbol{u}(x,n)$ with respect to x.

$$\boldsymbol{u}^*(s,n) = \int_{x=0}^{\infty} \exp[-sx]d\boldsymbol{u}(x,n),$$

where the i-th $(i \in \mathscr{M})$ element of $d\boldsymbol{u}(x,n)$ represents $\Pr(x < U \le x+dx, L = n)$, where U and L denote the stationary workload in system and the number of customers in the system, respectively.

Lemma 3.7. *$\boldsymbol{u}^*(s,n)$ is given by*

$$\boldsymbol{u}^*(s,n) = (1-\nu)\boldsymbol{\kappa}[\boldsymbol{R}^*(s)]^n, \quad n = 0,1,\ldots, \tag{3.27}$$

where $\boldsymbol{R}^*(s)$ $(\mathrm{Re}(s) > 0)$ denotes an $M \times M$ matrix defined as

$$\boldsymbol{R}^*(s) = \int_0^{\infty} \exp[-sx]dx \int_x^{\infty} d\boldsymbol{D}(y)\exp\left[\boldsymbol{Q}_{\mathrm{N}}(y-x)\right], \tag{3.28}$$

and it satisfies

$$\boldsymbol{R}^*(s)(s\boldsymbol{I} + \boldsymbol{Q}_{\mathrm{N}}) = \boldsymbol{Q}_{\mathrm{N}} - \boldsymbol{C} - \boldsymbol{D}^*(s). \tag{3.29}$$

The proof of Lemma 3.7 is given in Appendix 3.G. Note that (3.29) is equivalent to

$$s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)=[\boldsymbol{I}-\boldsymbol{R}^*(s)](s\boldsymbol{I}+\boldsymbol{Q}_{\mathrm{N}}). \tag{3.30}$$

It follows from Lemma 3.7 that

$$\boldsymbol{u}^*(s,n)=\boldsymbol{u}^*(s,n-1)\boldsymbol{R}^*(s), \qquad n=1,2,\ldots.$$

Summing up both sides of the above equation over all n ($n=1,2,\ldots$) then yields

$$\boldsymbol{u}^*(s)-(1-\nu)\boldsymbol{\kappa}=\boldsymbol{u}^*(s)\boldsymbol{R}^*(s).$$

Therefore, taking the limit $s\to 0+$, we have

$$\boldsymbol{\pi}-(1-\nu)\boldsymbol{\kappa}=\boldsymbol{\pi}\boldsymbol{R}, \tag{3.31}$$

where $\boldsymbol{R}$ is defined as

$$\boldsymbol{R}=\lim_{s\to 0+}\boldsymbol{R}^*(s).$$

It then follows from (3.29) that $\boldsymbol{R}$ is given by

$$\boldsymbol{R}=\boldsymbol{I}+(\boldsymbol{C}+\boldsymbol{D})(-\boldsymbol{Q}_{\mathrm{N}})^{-1}. \tag{3.32}$$

Lemma 3.8. *$\boldsymbol{R}$ is a non-negative matrix and all eigenvalues of $\boldsymbol{R}$ lie inside the unit disk.*

The proof of Lemma 3.8 is given in Appendix 3.H.

We now obtain the alternative formula for $\boldsymbol{u}^*(s)$, whose proof is given in Appendix 3.I.

Theorem 3.2. *$\boldsymbol{u}^*(s)$ (Re$(s)>0$) is given by*

$$\boldsymbol{u}^*(s)=(1-\nu)\boldsymbol{\kappa}\left[\boldsymbol{I}-\boldsymbol{R}^*(s)\right]^{-1}. \tag{3.33}$$

Before closing this section, we consider the relation between two representations for $\boldsymbol{u}^*(s)$ in Theorems 3.1 and 3.2. Recall that $\boldsymbol{R}$ is given by (3.32). Therefore, using $\boldsymbol{\pi}(\boldsymbol{C}+\boldsymbol{D}+\boldsymbol{\Gamma})=\boldsymbol{0}$, we obtain

$$\boldsymbol{\pi}\boldsymbol{R}=\boldsymbol{\pi}-\boldsymbol{\pi}\boldsymbol{\Gamma}(-\boldsymbol{Q}_{\mathrm{N}})^{-1}. \tag{3.34}$$

It then follows from (3.31) and (3.34) that $(1-\nu)\boldsymbol{\kappa}$ is given by

$$(1-\nu)\boldsymbol{\kappa}=\boldsymbol{\pi}\boldsymbol{\Gamma}(-\boldsymbol{Q}_{\mathrm{N}})^{-1}. \tag{3.35}$$

As a result, (3.25) is equivalent to

$$\boldsymbol{u}^*(s)[s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)]=(1-\nu)\boldsymbol{\kappa}(s\boldsymbol{I}+\boldsymbol{Q}_{\mathrm{N}}). \tag{3.36}$$

On the other hand, (3.33) is equivalent to

$$\boldsymbol{u}^*(s)[\boldsymbol{I}-\boldsymbol{R}^*(s)] = (1-\nu)\boldsymbol{\kappa}. \tag{3.37}$$

In what follows, we will show that (3.36) can be derived from (3.37), and vice versa.

Post-multiplying both sides of (3.37) by $s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}$ and using (3.30), we have (3.36). Therefore (3.36) is derived from (3.37).

Conversely, to derive (3.36) from (3.37), we should note that $s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}$ is singular for some s (Re$(s) > 0$). Let Ξ denote the set of s (Re$(s) > 0$) for which $s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}$ is singular.

$$\Xi = \{s; \det(s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}) = 0, \mathrm{Re}(s) > 0\}.$$

Note that Ξ is a (sub)set of eigenvalues of $-\boldsymbol{Q}_{\rm N}$, whose real parts are positive. Because $\boldsymbol{Q}_{\rm N}$ represents a defective infinitesimal generator of the censored underlying Markov chain, all eigenvalues of $-\boldsymbol{Q}_{\rm N}$ have positive real parts. Therefore, Ξ consists of all eigenvalues of $-\boldsymbol{Q}_{\rm N}$, so that the cardinal number of Ξ is not greater than M. It then follows from (3.30) and the elementwise continuity of $\boldsymbol{R}^*(s)$ (Re$(s) > 0$) that

$$\boldsymbol{I}-\boldsymbol{R}^*(s) = [s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)](s\boldsymbol{I}+\boldsymbol{Q}_{\rm N})^{-1}, \quad s \notin \Xi,$$

and

$$\boldsymbol{I}-\boldsymbol{R}^*(s) = \lim_{\alpha\to s}[\alpha\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(\alpha)](\alpha\boldsymbol{I}+\boldsymbol{Q}_{\rm N})^{-1}, \quad s \in \Xi.$$

Remark 3.3. *Because $\boldsymbol{I}-\boldsymbol{R}^*(s)$ (*Re$(s)>0$*) is non-singular, (3.30) implies that $s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)$ (*Re$(s)>0$*) is singular only for $s \in \Xi$.*

Therefore an $M \times M$ matrix $\boldsymbol{X}^*(s)$ (Re$(s) > 0$) defined as

$$\boldsymbol{X}^*(s) = \begin{cases} [s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)](s\boldsymbol{I}+\boldsymbol{Q}_{\rm N})^{-1}, & s \notin \Xi, \\ \lim\limits_{\alpha\to s}\left[[\alpha\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(\alpha)](\alpha\boldsymbol{I}+\boldsymbol{Q}_{\rm N})^{-1}\right], & s \in \Xi, \end{cases}$$

satisfies

$$\boldsymbol{X}^*(s) = \boldsymbol{I}-\boldsymbol{R}^*(s), \quad \mathrm{Re}(s) > 0.$$

Furthermore, because of the elementwise continuity of $\boldsymbol{u}^*(s)$ (Re$(s) > 0$), it follows from (3.36) that

$$\boldsymbol{u}^*(s)\boldsymbol{X}^*(s) = (1-\nu)\boldsymbol{\kappa}, \quad \mathrm{Re}(s) > 0.$$

(3.37) is thus derived from (3.36).

3.5 Waiting time and sojourn time

In this section, we consider the waiting time and sojourn time distributions in steady state. Recall that the service discipline is assumed to be FCFS. Let $U_k^{\rm A}$

($k \in \mathcal{K}$) denote the workload in system seen by a randomly chosen class k customer on arrival and let B_k ($k \in \mathcal{K}$) denote the amount of the service requirement of this customer. Also, let $\tilde{T}^{\mathrm{A}}_{\mathrm{D},k}$ ($k \in \mathcal{K}$) denote the length of the interval from the arrival of this customer to the occurrence of the next disaster. We define W_k and $\overline{W}_k$ ($k \in \mathcal{K}$) as the waiting time and the sojourn time, respectively, of a randomly chosen class k customer.

$$W_k = \min(U^{\mathrm{A}}_k, \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}), \qquad \overline{W}_k = \min(U^{\mathrm{A}}_k + B_k, \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}), \qquad k \in \mathcal{K}.$$

For $k \in \mathcal{K}$ and $x \geq 0$, we define $\boldsymbol{w}_k(x)$ (resp. $\overline{\boldsymbol{w}}_k(x)$) as a $1 \times M$ vector whose j-th ($j \in \mathcal{M}$) element represents the joint probability that the waiting time (resp. sojourn time) of a randomly chosen class k customer does not exceed x and the underlying Markov chain is in state j at the end of the waiting time (resp. sojourn time).

$$\begin{aligned}[\boldsymbol{w}_k(x)]_j &= \Pr(W_k \leq x, S_{t_{\mathrm{A}}+W_k} = j),\\ [\overline{\boldsymbol{w}}_k(x)]_j &= \Pr(\overline{W}_k \leq x, S_{t_{\mathrm{A}}+\overline{W}_k} = j),\end{aligned}$$

where t_{A} denotes the arrival time. Let $\boldsymbol{w}^*_k(s)$ and $\overline{\boldsymbol{w}}^*_k(s)$ ($\mathrm{Re}(s) > 0$) denote the LSTs of $\boldsymbol{w}_k(x)$ and $\overline{\boldsymbol{w}}_k(x)$, respectively.

$$\boldsymbol{w}^*_k(s) = \int_0^\infty \exp[-sx] d\boldsymbol{w}_k(x), \qquad \overline{\boldsymbol{w}}^*_k(s) = \int_0^\infty \exp[-sx] d\overline{\boldsymbol{w}}_k(x).$$

Moreover, for $k \in \mathcal{K}$ and $x \geq 0$, we define $\boldsymbol{w}_{\mathrm{N},k}(x)$ (resp. $\overline{\boldsymbol{w}}_{\mathrm{N},k}(x)$) as a $1 \times M$ vector whose j-th ($j \in \mathcal{M}$) element represents the joint probability that the waiting time (resp. sojourn time) of a randomly chosen class k customer does not exceed x, the underlying Markov chain is in state j at the end of the waiting time (resp. sojourn time), and no disasters occur in the waiting time (resp. sojourn time).

$$\begin{aligned}[\boldsymbol{w}_{\mathrm{N},k}(x)]_j &= \Pr(U^{\mathrm{A}}_k \leq x, U^{\mathrm{A}}_k < \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}, S_{t_{\mathrm{A}}+U^{\mathrm{A}}_k} = j),\\ [\overline{\boldsymbol{w}}_{\mathrm{N},k}(x)]_j &= \Pr(U^{\mathrm{A}}_k + B_k \leq x, U^{\mathrm{A}}_k + B_k < \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}, S_{t_{\mathrm{A}}+U^{\mathrm{A}}_k+B_k} = j).\end{aligned}$$

Also, for $k \in \mathcal{K}$ and $x \geq 0$, we define $\boldsymbol{w}_{\mathrm{D},k}(x)$ (resp. $\overline{\boldsymbol{w}}_{\mathrm{D},k}(x)$) as a $1 \times M$ vector whose j-th ($j \in \mathcal{M}$) element represents the joint probability that the waiting time (resp. sojourn time) of a randomly chosen class k customer does not exceed x, the underlying Markov chain is in state j at the end of the waiting time (resp. sojourn time), and the waiting time (resp. sojourn time) ends with a disaster.

$$\begin{aligned}[\boldsymbol{w}_{\mathrm{D},k}(x)]_j &= \Pr(\tilde{T}^{\mathrm{A}}_{\mathrm{D},k} \leq x, U^{\mathrm{A}}_k \geq \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}, S_{t_{\mathrm{A}}+\tilde{T}^{\mathrm{A}}_{\mathrm{D},k}} = j),\\ [\overline{\boldsymbol{w}}_{\mathrm{D},k}(x)]_j &= \Pr(\tilde{T}^{\mathrm{A}}_{\mathrm{D},k} \leq x, U^{\mathrm{A}}_k + B_k \geq \tilde{T}^{\mathrm{A}}_{\mathrm{D},k}, S_{t_{\mathrm{A}}+\tilde{T}^{\mathrm{A}}_{\mathrm{D},k}} = j).\end{aligned}$$

We define LSTs $\boldsymbol{w}^*_{\mathrm{N},k}(s)$, $\boldsymbol{w}^*_{\mathrm{D},k}(s)$, $\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)$, and $\overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s)$ ($k \in \mathcal{K}$, $\mathrm{Re}(s) > 0$) as

$$\boldsymbol{w}^*_{\mathrm{N},k}(s) = \int_0^\infty \exp[-sx] d\boldsymbol{w}_{\mathrm{N},k}(x), \quad \boldsymbol{w}^*_{\mathrm{D},k}(s) = \int_0^\infty \exp[-sx] d\boldsymbol{w}_{\mathrm{D},k}(x),$$

$$\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s) = \int_0^\infty \exp[-sx]d\overline{\boldsymbol{w}}_{\mathrm{N},k}(x), \quad \overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s) = \int_0^\infty \exp[-sx]d\overline{\boldsymbol{w}}_{\mathrm{D},k}(x).$$

We then have

$$\boldsymbol{w}^*_k(s) = \boldsymbol{w}^*_{\mathrm{N},k}(s) + \boldsymbol{w}^*_{\mathrm{D},k}(s), \qquad \overline{\boldsymbol{w}}^*_k(s) = \overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s) + \overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s). \tag{3.38}$$

Theorem 3.3. *$\boldsymbol{w}^*_{\mathrm{N},k}(s)$ and $\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)$ ($k \in \mathscr{K}$, $\mathrm{Re}(s) > 0$) are given by*

$$\boldsymbol{w}^*_{\mathrm{N},k}(s) = \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}} \int_0^\infty d\boldsymbol{u}(x)\boldsymbol{D}_k \exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big], \tag{3.39}$$

$$\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s) = \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}} \int_0^\infty d\boldsymbol{u}(x) \int_0^\infty d\boldsymbol{D}_k(y) \exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})(x+y)\big], \tag{3.40}$$

respectively. On the other hand, $\boldsymbol{w}^*_{\mathrm{D},k}(s)$ and $\overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s)$ ($k \in \mathscr{K}$, $\mathrm{Re}(s) > 0$) are given by

$$\boldsymbol{w}^*_{\mathrm{D},k}(s) = \left(\frac{\boldsymbol{\pi D}_k}{\boldsymbol{\pi D}_k\boldsymbol{e}} - \boldsymbol{w}^*_{\mathrm{N},k}(s)\right)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}, \tag{3.41}$$

$$\overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s) = \left(\frac{\boldsymbol{\pi D}_k}{\boldsymbol{\pi D}_k\boldsymbol{e}} - \overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)\right)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}, \tag{3.42}$$

respectively.

The proof of Theorem 3.3 is given in Appendix 3.J. Recall that $s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})$ is nonsingular for all s ($\mathrm{Re}(s) > 0$) (see Remark 3.1).

With (3.38) and Theorem 3.3, we have the following corollary.

Corollary 3.1. *$\boldsymbol{w}^*_k(s)$ and $\overline{\boldsymbol{w}}^*_k(s)$ ($k \in \mathscr{K}$, $\mathrm{Re}(s) > 0$) are represented in terms of $\boldsymbol{w}^*_{\mathrm{N},k}(s)$ and $\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)$, respectively.*

$$\boldsymbol{w}^*_k(s) = \boldsymbol{w}^*_{\mathrm{N},k}(s) + \left(\frac{\boldsymbol{\pi D}_k}{\boldsymbol{\pi D}_k\boldsymbol{e}} - \boldsymbol{w}^*_{\mathrm{N},k}(s)\right)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}, \quad k \in \mathscr{K}, \tag{3.43}$$

$$\overline{\boldsymbol{w}}^*_k(s) = \overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s) + \left(\frac{\boldsymbol{\pi D}_k}{\boldsymbol{\pi D}_k\boldsymbol{e}} - \overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)\right)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}, \quad k \in \mathscr{K}. \tag{3.44}$$

We now discuss the computation of the moments of the waiting time and sojourn time distributions, which we can obtain by taking the derivatives of their LSTs. From Theorem 3.3 and Corollary 3.1, we observe that $\boldsymbol{w}^*_{\mathrm{N},k}(s)$ and $\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s)$ are essential quantities for the waiting time and sojourn time distributions, so that we consider those first.

Let θ denote the maximum absolute value of the diagonal elements of matrix $\boldsymbol{C}$. With uniformization at rate θ, (3.39) and (3.40) are rewritten to be

$$\boldsymbol{w}^*_{\mathrm{N},k}(s) = \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}} \sum_{m=0}^\infty \int_0^\infty d\boldsymbol{u}(x) \exp\big[-(s+\theta)x\big] \frac{(\theta x)^m}{m!} \boldsymbol{D}_k[\boldsymbol{I} + \theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m$$

$$
\begin{aligned}
&= \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty} \boldsymbol{u}^{(m)}(s,\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m, \qquad (3.45)\\
\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s) &= \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\int_0^{\infty} \exp[-(s+\theta)x]\frac{(\theta x)^j}{j!}d\boldsymbol{u}(x)\\
&\qquad \cdot \int_0^{\infty}\exp[-(s+\theta)y]\frac{(\theta y)^i}{i!}d\boldsymbol{D}_k(y)\\
&\qquad\qquad \cdot[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^{i+j}\\
&= \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty}\sum_{i=0}^{m} \boldsymbol{u}^{(i)}(s,\theta)\boldsymbol{D}_k^{(m-i)}(s,\theta)[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m, \qquad (3.46)
\end{aligned}
$$

respectively, where

$$
\begin{aligned}
\boldsymbol{u}^{(m)}(s,\theta) &= \int_0^{\infty}\exp\big[-(s+\theta)x\big]\frac{(\theta x)^m}{m!}d\boldsymbol{u}(x), \qquad m=0,1,\ldots,\\
\boldsymbol{D}_k^{(m)}(s,\theta) &= \int_0^{\infty}\exp\big[-(s+\theta)x\big]\frac{(\theta x)^m}{m!}d\boldsymbol{D}_k(x), \qquad k\in\mathcal{K},\ m=0,1,\ldots.
\end{aligned}
$$

We define $\boldsymbol{w}^{(n)}_{\mathrm{N},k}$ and $\overline{\boldsymbol{w}}^{(n)}_{\mathrm{N},k}$ ($k\in\mathcal{K}$, $n=0,1,\ldots$) as

$$
\begin{aligned}
\boldsymbol{w}^{(0)}_{\mathrm{N},k} &= \lim_{s\to0+}\boldsymbol{w}^*_{\mathrm{N},k}(s), & \overline{\boldsymbol{w}}^{(0)}_{\mathrm{N},k} &= \lim_{s\to0+}\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s),\\
\boldsymbol{w}^{(n)}_{\mathrm{N},k} &= \lim_{s\to0+}\frac{\partial^n}{\partial s^n}\boldsymbol{w}^*_{\mathrm{N},k}(s), & \overline{\boldsymbol{w}}^{(n)}_{\mathrm{N},k} &= \lim_{s\to0+}\frac{\partial^n}{\partial s^n}\overline{\boldsymbol{w}}^*_{\mathrm{N},k}(s), \qquad n=1,2,\ldots,
\end{aligned}
$$

It then follows from (3.45) and (3.46) that

$$
\begin{aligned}
\boldsymbol{w}^{(0)}_{\mathrm{N},k} &= \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}}\sum_{m=0}^{\infty}\boldsymbol{u}^{(m)}(\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m,\\
\overline{\boldsymbol{w}}^{(0)}_{\mathrm{N},k} &= \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}}\sum_{m=0}^{\infty}\sum_{i=0}^{m}\boldsymbol{u}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta)[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m,
\end{aligned}
$$

where $\boldsymbol{u}^{(m)}(\theta)$ ($m=0,1,\ldots$) and $\boldsymbol{D}_k^{(m)}(\theta)$ ($k\in\mathcal{K}$, $m=0,1,\ldots$) are defined as

$$
\begin{aligned}
\boldsymbol{u}^{(m)}(\theta) &= \lim_{s\to0+}\boldsymbol{u}^{(m)}(s,\theta)=\int_0^{\infty}\exp[-\theta x]\frac{(\theta x)^m}{m!}d\boldsymbol{u}(x), \qquad (3.47)\\
\boldsymbol{D}_k^{(m)}(\theta) &= \lim_{s\to0+}\boldsymbol{D}_k^{(m)}(s,\theta)=\int_0^{\infty}\exp[-\theta x]\frac{(\theta x)^m}{m!}d\boldsymbol{D}_k(x). \qquad (3.48)
\end{aligned}
$$

Note here that $\boldsymbol{D}_k^{(m)}(\theta)$ ($k\in\mathcal{K}$, $m=0,1,\ldots$) satisfies

$$
\boldsymbol{D}_k^*(\theta-\theta z)=\sum_{m=0}^{\infty}\boldsymbol{D}_k^{(m)}(\theta)z^m, \qquad |z|<1. \qquad (3.49)
$$

Therefore we can obtain $\boldsymbol{D}_k^{(m)}(\theta)$ ($k \in \mathcal{K}$, $m = 0,1,\ldots$) by comparing the coefficient of z^m on both sides of this equation. In a numerical example, we demonstrate how to compute $\boldsymbol{D}_k^{(m)}(\theta)$ when the service requirement distribution is of phase-type.

Note that the probability that no disasters occur in the waiting time of a randomly chosen class k ($k \in \mathcal{K}$) customer is given by

$$\Pr(U_k^{\mathrm{A}} < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}}) = \boldsymbol{w}_{\mathrm{N},k}^{(0)}\boldsymbol{e},$$

and the probability that no disasters occur in the sojourn time of a randomly chosen class k ($k \in \mathcal{K}$) customer is given by

$$\Pr(U_k^{\mathrm{A}} + B_k < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}}) = \overline{\boldsymbol{w}}_{\mathrm{N},k}^{(0)}\boldsymbol{e}.$$

Next we consider $\boldsymbol{w}_{\mathrm{N},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}$ ($k \in \mathcal{K}$, $n = 0,1,\ldots$).

Lemma 3.9. *The limits $s \to 0+$ of the n-th ($n = 1,2,\ldots$) derivatives of $\boldsymbol{u}^{(m)}(s,\theta)$ ($m = 0,1,\ldots$) and $\boldsymbol{D}_k^{(m)}(s,\theta)$ ($k \in \mathcal{K}$, $m = 0,1,\ldots$) with respect to s are given in terms of $\boldsymbol{u}^{(m+n)}(\theta)$ and $\boldsymbol{D}_k^{(m+n)}(\theta)$, respectively.*

$$\begin{aligned}\lim_{s\to 0+}\frac{\partial^n}{\partial s^n}[\boldsymbol{u}^{(m)}(s,\theta)] &= (-1)^n \cdot \frac{(n+m)!}{m!\theta^n} \cdot \boldsymbol{u}^{(n+m)}(\theta), &(3.50)\\ \lim_{s\to 0+}\frac{\partial^n}{\partial s^n}[\boldsymbol{D}_k^{(m)}(s,\theta)] &= (-1)^n \cdot \frac{(n+m)!}{m!\theta^n} \cdot \boldsymbol{D}_k^{(n+m)}(\theta). &(3.51)\end{aligned}$$

The proof of Lemma 3.9 is given in Appendix 3.K.

Differentiating both sides of (3.45) and (3.46) with respect to s, taking the limits $s \to 0$, and using Lemma 3.9, we obtain the following theorem.

Theorem 3.4. *$\boldsymbol{w}_{\mathrm{N},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}$ ($k \in \mathcal{K}$, $n = 0,1,\ldots$) are given by*

$$\begin{aligned}\boldsymbol{w}_{\mathrm{N},k}^{(n)} &= \frac{n!}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}}\left(\frac{-1}{\theta}\right)^n \sum_{m=n}^{\infty}\binom{m}{n}\boldsymbol{u}^{(m)}(\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^{m-n},\\ \overline{\boldsymbol{w}}_{N,k}^{(n)} &= \frac{n!}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}}\left(\frac{-1}{\theta}\right)^n \sum_{l=0}^{n}\sum_{i=l}^{\infty}\binom{i}{l}\boldsymbol{u}^{(i)}(\theta)\sum_{m=0}^{\infty}\binom{n-l+m}{n-l}\boldsymbol{D}_k^{(n-l+m)}(\theta)\\ &\qquad \cdot[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^{i-l+m},\end{aligned}$$

respectively.

The remaining is to show a computational procedure for $\boldsymbol{u}^{(m)}(\theta)$ ($m = 0,1,\ldots$). Note that $\boldsymbol{u}^{(m)}(\theta)$ ($m = 0,1,\ldots$) can be obtained in a way similar to that in [Tak01, Lemma 3]. We define $\boldsymbol{A}_m$ ($m = 0,1,\ldots$) and $\boldsymbol{E}$ as $M \times M$ matrices given by

$$\boldsymbol{A}_0 = \boldsymbol{I} + \theta^{-1}\boldsymbol{C} + \theta^{-1}\boldsymbol{D}^{(0)}(\theta), \qquad (3.52)$$

$$\boldsymbol{A}_m = \theta^{-1}\boldsymbol{D}^{(m)}(\theta), \qquad m = 1,2,\ldots, \tag{3.53}$$
$$\boldsymbol{E} = \theta^{-1}\boldsymbol{\Gamma}, \tag{3.54}$$

respectively, where

$$\boldsymbol{D}^{(m)}(\theta) = \sum_{k\in\mathcal{K}} \boldsymbol{D}_k^{(m)}(\theta), \qquad m = 0,1,\ldots.$$

We also define $\boldsymbol{A}$ as an $M \times M$ matrix given by

$$\boldsymbol{A} = \sum_{m=0}^{\infty} \boldsymbol{A}_m = \boldsymbol{I} + \theta^{-1}(\boldsymbol{C}+\boldsymbol{D}).$$

Note that $\boldsymbol{I} - \boldsymbol{A} = -\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})$ is non-singular.

Lemma 3.10. *$\{\boldsymbol{u}^{(m)}(\theta),\ m = 0,1,\ldots\}$ is identical to the stationary distribution of a Markov chain, whose transition probability matrix $\boldsymbol{T}$ is given by*

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{A}_0+\boldsymbol{A}_1+\boldsymbol{E} & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{A}_0+\boldsymbol{E} & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ \boldsymbol{E} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ \boldsymbol{E} & \boldsymbol{0} & \boldsymbol{A}_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

i.e.,

$$(\boldsymbol{u}^{(0)}(\theta), \boldsymbol{u}^{(1)}(\theta),\ldots)\boldsymbol{T} = (\boldsymbol{u}^{(0)}(\theta), \boldsymbol{u}^{(1)}(\theta),\ldots), \tag{3.55}$$
$$\sum_{m=0}^{\infty} \boldsymbol{u}^{(m)}(\theta)\boldsymbol{e} = 1. \tag{3.56}$$

The proof of Lemma 3.10 is given in Appendix 3.L.

Lemma 3.10 shows that $\boldsymbol{u}^{(m)}(\theta)$ $(m = 0,1,\ldots)$ can be computed by using an algorithm developed in [DS04, Ram88]. To that end, we need to find the boundary vector $\boldsymbol{u}^{(0)}(\theta)$ first. Let $\boldsymbol{G}_{\mathrm{N}}$ denote an $M \times M$ substochastic matrix given by the minimum non-negative solution of the following equation.

$$\boldsymbol{G}_{\mathrm{N}} = \sum_{m=0}^{\infty} \boldsymbol{A}_m \boldsymbol{G}_{\mathrm{N}}^m.$$

We define $\boldsymbol{G}_{\mathrm{D}}$ as an $M \times M$ substochastic matrix given by

$$\boldsymbol{G}_{\mathrm{D}} = \boldsymbol{E} + \sum_{m=1}^{\infty} \boldsymbol{A}_m \sum_{i=0}^{m-1} \boldsymbol{G}_{\mathrm{N}}^i \boldsymbol{G}_{\mathrm{D}}.$$

It can be verified that

$$\boldsymbol{G}_{\mathrm{D}} = (\boldsymbol{I} - \boldsymbol{G}_{\mathrm{N}})(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{E},$$

and

$$(\boldsymbol{G}_{\mathrm{N}} + \boldsymbol{G}_{\mathrm{D}})\boldsymbol{e} = \boldsymbol{e}.$$

Remark 3.4. *It can be shown that* $\boldsymbol{G}_{\mathrm{N}}$ *and* $\boldsymbol{G}_{\mathrm{D}}$ *are given by*

$$\boldsymbol{G}_{\mathrm{N}} = \boldsymbol{I} + \theta^{-1}\boldsymbol{Q}_{\mathrm{N}}, \qquad \boldsymbol{G}_{\mathrm{D}} = \theta^{-1}\boldsymbol{Q}_{\mathrm{D}}, \tag{3.57}$$

respectively.

We define $\boldsymbol{K}$ as an $M \times M$ stochastic matrix given by

$$\boldsymbol{K} = \boldsymbol{A}_0 + \sum_{m=1}^{\infty} \boldsymbol{A}_m \boldsymbol{G}_{\mathrm{N}}^{m-1} + \boldsymbol{E} + \sum_{m=2}^{\infty} \boldsymbol{A}_m \sum_{i=0}^{m-2} \boldsymbol{G}_{\mathrm{N}}^{i} \boldsymbol{G}_{\mathrm{D}}. \tag{3.58}$$

Note that $\boldsymbol{K}$ represents the transition probability matrix of a censored Markov chain obtained by observing the evolution of the Markov chain characterized by $\boldsymbol{T}$ only when it stays in level zero, so that

$$\boldsymbol{u}^{(0)}(\theta) = \boldsymbol{u}^{(0)}(\theta)\boldsymbol{K}. \tag{3.59}$$

We then introduce the invariant probability vector $\hat{\boldsymbol{\kappa}}$ of $\boldsymbol{K}$.

$$\hat{\boldsymbol{\kappa}}\boldsymbol{K} = \hat{\boldsymbol{\kappa}}, \qquad \hat{\boldsymbol{\kappa}}\boldsymbol{e} = 1.$$

Lemma 3.11. $\boldsymbol{u}^{(0)}(\theta)$ *is given by*

$$\boldsymbol{u}^{(0)}(\theta) = \frac{1 - \nu - \boldsymbol{\pi}\boldsymbol{E}\boldsymbol{e}}{\hat{\boldsymbol{\kappa}}\boldsymbol{A}_0\boldsymbol{e}} \cdot \hat{\boldsymbol{\kappa}}.$$

The proof of Lemma 3.11 is given in Appendix 3.M. Once we obtain $\boldsymbol{u}^{(0)}(\theta)$, $\boldsymbol{u}^{(m)}(\theta)$ ($m = 1, 2, \ldots$) can be computed successively in the same way as in [DS04, Theorem 1], i.e.,

$$\boldsymbol{u}^{(m)}(\theta) = \sum_{i=0}^{m-1} \boldsymbol{u}^{(i)}(\theta)\overline{\boldsymbol{A}}_{m-i+1}(\boldsymbol{I} - \overline{\boldsymbol{A}}_1)^{-1}, \qquad m = 1, 2, \ldots,$$

where

$$\overline{\boldsymbol{A}}_i = \sum_{j=i}^{\infty} \boldsymbol{A}_j \boldsymbol{G}_{\mathrm{N}}^{j-i}, \qquad i = 1, 2, \ldots.$$

So far, we provide the numerically feasible formulas for $\boldsymbol{w}_{\mathrm{N},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}$. Next we consider $\boldsymbol{w}_{\mathrm{D},k}^{*}(s)$ and $\overline{\boldsymbol{w}}_{\mathrm{D},k}^{*}(s)$ in Theorem 3.3. We define $\boldsymbol{w}_{\mathrm{D},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}$ ($k \in \mathscr{K}$, $n = 0, 1, \ldots$) as

$$\boldsymbol{w}_{\mathrm{D},k}^{(0)} = \lim_{s \to 0+} \boldsymbol{w}_{\mathrm{D},k}^{*}(s), \qquad \overline{\boldsymbol{w}}_{\mathrm{D},k}^{(0)} = \lim_{s \to 0+} \overline{\boldsymbol{w}}_{\mathrm{D},k}^{*}(s),$$

$$\boldsymbol{w}_{\mathrm{D},k}^{(n)} = \lim_{s \to 0+} \frac{\partial^n}{\partial s^n} \boldsymbol{w}_{\mathrm{D},k}^{*}(s), \qquad \overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)} = \lim_{s \to 0+} \frac{\partial^n}{\partial s^n} \overline{\boldsymbol{w}}_{\mathrm{D},k}^{*}(s), \qquad n = 1, 2, \ldots,$$

Theorem 3.5. *$\boldsymbol{w}_{\mathrm{D},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}$ ($k \in \mathscr{K}$, $n = 0, 1, \ldots$) are given by*

$$\boldsymbol{w}_{\mathrm{D},k}^{(n)} = \boldsymbol{\beta}_{\mathrm{D},k}^{(n)} \boldsymbol{\Gamma}, \qquad \overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)} = \overline{\boldsymbol{\beta}}_{\mathrm{D},k}^{(n)} \boldsymbol{\Gamma},$$

respectively, where $\boldsymbol{\beta}_{\mathrm{D},k}^{(n)}$ and $\overline{\boldsymbol{\beta}}_{\mathrm{D},k}^{(n)}$ are determined successively by

$$\begin{aligned}
\boldsymbol{\beta}_{\mathrm{D},k}^{(0)} &= \left(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \boldsymbol{w}_{\mathrm{N},k}^{(0)}\right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}, \\
\boldsymbol{\beta}_{\mathrm{D},k}^{(n)} &= (-1)\left(n\boldsymbol{\beta}_{\mathrm{D},k}^{(n-1)} + \boldsymbol{w}_{\mathrm{N},k}^{(n)}\right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}, \qquad n = 1, 2, \ldots, \\
\overline{\boldsymbol{\beta}}_{\mathrm{D},k}^{(0)} &= \left(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \overline{\boldsymbol{w}}_{\mathrm{N},k}^{(0)}\right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}, \\
\overline{\boldsymbol{\beta}}_{\mathrm{D},k}^{(n)} &= (-1)\left(n\overline{\boldsymbol{\beta}}_{\mathrm{D},k}^{(n-1)} + \overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}\right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}, \qquad n = 1, 2, \ldots,
\end{aligned}$$

respectively.

The proof of Theorem 3.5 is given in Appendix 3.N.

In summary, we can compute the moments of waiting time and sojourn time distributions based on Theorems 3.4 and 3.5, i.e., for $n = 1, 2, \ldots$,

$$\mathrm{E}[W_k^n \mid \text{no disasters occur in the waiting time}] = (-1)^n \frac{\boldsymbol{w}_{\mathrm{N},k}^{(n)}\boldsymbol{e}}{\boldsymbol{w}_{\mathrm{N},k}^{(0)}\boldsymbol{e}}, \tag{3.60}$$

$$\mathrm{E}[W_k^n \mid \text{the waiting time ends with a disaster}] = (-1)^n \frac{\boldsymbol{w}_{\mathrm{D},k}^{(n)}\boldsymbol{e}}{\boldsymbol{w}_{\mathrm{D},k}^{(0)}\boldsymbol{e}}, \tag{3.61}$$

$$\mathrm{E}[W_k^n] = (-1)^n \left(\boldsymbol{w}_{\mathrm{N},k}^{(n)} + \boldsymbol{w}_{\mathrm{D},k}^{(n)}\right)\boldsymbol{e}, \tag{3.62}$$

$$\mathrm{E}[\overline{W}_k^n \mid \text{no disasters occur in the sojourn time}] = (-1)^n \frac{\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}\boldsymbol{e}}{\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(0)}\boldsymbol{e}}, \tag{3.63}$$

$$\mathrm{E}[\overline{W}_k^n \mid \text{the sojourn time ends with a disaster}] = (-1)^n \frac{\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}\boldsymbol{e}}{\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(0)}\boldsymbol{e}}, \tag{3.64}$$

$$\mathrm{E}[\overline{W}_k^n] = (-1)^n \left(\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)} + \overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}\right)\boldsymbol{e}. \tag{3.65}$$

In particular, we have

$$\begin{aligned}
&\mathrm{E}[W_k \mid \text{no disasters occur in the waiting time}] \\
&\quad = \frac{1}{\Pr(U_k^{\mathrm{A}} < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}})} \cdot \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty} \frac{m+1}{\theta} \cdot \boldsymbol{u}^{(m+1)}(\theta)\boldsymbol{D}_k[\boldsymbol{I} + \theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \boldsymbol{e},
\end{aligned}$$

$$\mathrm{E}[W_k \mid \text{the waiting time ends with a disaster}]$$

$$\begin{aligned}
&= \frac{1}{1-\Pr(U_k^{\mathrm{A}} < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}})} \\
&\quad \cdot \Bigg[(-1)\cdot \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \left(\sum_{m=0}^{\infty} \frac{m+1}{\theta} \cdot \boldsymbol{u}^{(m+1)}(\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \boldsymbol{e} \right) \\
&\quad + \left(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty} \boldsymbol{u}^{(m)}(\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}\Bigg],
\end{aligned}$$

$$\mathrm{E}[W_k] = \left(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty} \boldsymbol{u}^{(m)}(\theta)\boldsymbol{D}_k[\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \right)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e},$$

$$\begin{aligned}
&\mathrm{E}[\overline{W}_k \mid \text{no disasters occur in the sojourn time}] \\
&\quad = \frac{1}{\Pr(U_k^{\mathrm{A}} + B_k < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}})} \\
&\qquad \cdot \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty}\sum_{i=0}^{m} \frac{m+1}{\theta} \left(\boldsymbol{u}^{(i+1)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta) + \boldsymbol{u}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i+1)}(\theta) \right) \\
&\qquad\qquad \cdot [\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \boldsymbol{e},
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}[\overline{W}_k \mid \text{the sojourn time ends with a disaster}] \\
&\quad = \frac{1}{1-\Pr(U_k^{\mathrm{A}} + B_k < \tilde{T}_{\mathrm{D},k}^{\mathrm{A}})} \\
&\qquad \cdot \Bigg[(-1)\cdot \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \Bigg\{ \sum_{m=0}^{\infty}\sum_{i=0}^{m} \frac{m+1}{\theta} \Big(\boldsymbol{u}^{(i+1)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta) \\
&\qquad\qquad + \boldsymbol{u}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i+1)}(\theta) \Big) [\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \boldsymbol{e} \Bigg\} \\
&\qquad + \Bigg(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty}\sum_{i=0}^{m} \boldsymbol{u}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta) \\
&\qquad\qquad \cdot [\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \Bigg)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}\Bigg],
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\overline{W}_k] = {}& \Bigg(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \frac{1}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \sum_{m=0}^{\infty}\sum_{i=0}^{m} \boldsymbol{u}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta) \\
&\qquad \cdot [\boldsymbol{I}+\theta^{-1}(\boldsymbol{C}+\boldsymbol{D})]^m \Bigg)[-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{e}.
\end{aligned}$$

3.6 Joint queue length distribution

In this section, we consider the joint distribution of the numbers of customers in respective classes. Let $\overline{L}_{k,t}$ ($k \in \mathscr{K}$, $t \geq 0$) denote the number of class k customers in the system at time t, and $L_{k,t}$ ($k \in \mathscr{K}$, $t \geq 0$) denote the number of waiting customers of class k at time t. We define $\boldsymbol{n}$ as a $1 \times K$ non-negative integer vector whose i-th ($i \in \mathscr{K}$) element is denoted by n_i.

$$\boldsymbol{n} = (n_1, n_2, \ldots, n_K).$$

Let $\mathscr{N}$ denote the set of all $1 \times K$ non-negative integer vectors.

$$\mathscr{N} = \{(n_1, n_2, \ldots n_K);\ n_k \in \{0, 1, \ldots\}\ (k \in \mathscr{K})\}.$$

Let $\boldsymbol{y}(t,\boldsymbol{n})$ ($t > 0$, $\boldsymbol{n} \in \mathscr{N}$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the joint probability that for each $k \in \mathscr{K}$, the number of class k customers in the system is equal to n_k and the underlying Markov chain is in state j at time t.

$$[\boldsymbol{y}(t,\boldsymbol{n})]_j = \Pr(\overline{L}_{1,t} = n_1, \overline{L}_{2,t} = n_2, \ldots, \overline{L}_{K,t} = n_K, S_t = j).$$

Similarly, let $\boldsymbol{x}(t,\boldsymbol{n})$ ($t > 0$, $\boldsymbol{n} \in \mathscr{N}$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the joint probability that for each $k \in \mathscr{K}$, the number of waiting customers of class k is equal to n_k and the underlying Markov chain is in state j at time t.

$$[\boldsymbol{x}(t,\boldsymbol{n})]_j = \Pr(L_{1,t} = n_1, L_{2,t} = n_2, \ldots, L_{K,t} = n_K, S_t = j).$$

We then define $\boldsymbol{y}(\boldsymbol{n})$ and $\boldsymbol{x}(\boldsymbol{n})$ ($\boldsymbol{n} \in \mathscr{N}$) as

$$\boldsymbol{y}(\boldsymbol{n}) = \lim_{t\to\infty} \boldsymbol{y}(t,\boldsymbol{n}), \qquad \boldsymbol{x}(\boldsymbol{n}) = \lim_{t\to\infty} \boldsymbol{x}(t,\boldsymbol{n}),$$

respectively. By definition, $\boldsymbol{y}(\boldsymbol{0}) = (1-\nu)\boldsymbol{\kappa}$. For complex numbers z_k ($k \in \mathscr{K}$) such that $|z_k| < 1$, we define $\boldsymbol{z}$ as a $1 \times K$ vector given by

$$\boldsymbol{z} = (z_1, z_2, \ldots, z_K).$$

Let $\mathscr{Z}$ denote the set of all $1 \times K$ vectors whose elements are complex numbers with moduli less than 1.

$$\mathscr{Z} = \{(z_1, z_2, \ldots, z_K);\ |z_k| < 1\ (k \in \mathscr{K})\}.$$

Let $\boldsymbol{y}^*(\boldsymbol{z})$ and $\boldsymbol{x}^*(\boldsymbol{z})$ ($\boldsymbol{z} \in \mathscr{Z}$) denote the joint probability generating functions of $\boldsymbol{y}(\boldsymbol{n})$ and $\boldsymbol{x}(\boldsymbol{n})$, respectively.

$$\boldsymbol{y}^*(\boldsymbol{z}) = \sum_{n \in \mathscr{N}} \boldsymbol{y}(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K},$$

$$\boldsymbol{x}^*(\boldsymbol{z}) = \sum_{n \in \mathscr{N}} \boldsymbol{x}(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K}.$$

Furthermore, let $\boldsymbol{x}_k(t,\boldsymbol{n})$ ($k \in \mathscr{K}$, $t > 0$, $\boldsymbol{n} \in \mathscr{N}$) denote a $1 \times M$ vector whose j-th ($j \in \mathscr{M}$) element represents the joint probability that a class k customer is being served, the numbers of waiting customers of class i ($i \in \mathscr{K}$) is equal to n_i, and the underlying Markov chain is in state j at time t.

$$\begin{aligned}[\boldsymbol{x}_k(t,\boldsymbol{n})]_j \;=\; & \Pr(\text{a class } k \text{ customer is being served at time } t, \\ & \qquad L_{1,t} = n_1, L_{2,t} = n_2, \ldots, L_{K,t} = n_K, S_t = j).\end{aligned}$$

We define $\boldsymbol{x}_k(\boldsymbol{n})$ ($k \in \mathscr{K}$, $\boldsymbol{n} \in \mathscr{N}$) as

$$\boldsymbol{x}_k(\boldsymbol{n}) = \lim_{t \to \infty} \boldsymbol{x}_k(t,\boldsymbol{n}).$$

Let $\boldsymbol{x}_k^*(\boldsymbol{z})$ ($k \in \mathscr{K}$, $\boldsymbol{z} \in \mathscr{Z}$) denote the joint probability generating function of $\boldsymbol{x}(\boldsymbol{n})$.

$$\boldsymbol{x}_k^*(\boldsymbol{z}) = \sum_{n \in \mathscr{N}} \boldsymbol{x}_k(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K}.$$

We then have

$$\boldsymbol{y}^*(\boldsymbol{z}) = (1-\nu)\boldsymbol{\kappa} + \sum_{k \in \mathscr{K}} z_k \boldsymbol{x}_k^*(\boldsymbol{z}), \qquad \boldsymbol{z} \in \mathscr{Z}, \tag{3.66}$$

$$\boldsymbol{x}^*(\boldsymbol{z}) = (1-\nu)\boldsymbol{\kappa} + \sum_{k \in \mathscr{K}} \boldsymbol{x}_k^*(\boldsymbol{z}), \qquad \boldsymbol{z} \in \mathscr{Z}. \tag{3.67}$$

Theorem 3.6. *$\boldsymbol{x}_k^*(\boldsymbol{z})$ ($k \in \mathscr{K}$, $\boldsymbol{z} \in \mathscr{Z}$) is given by*

$$\boldsymbol{x}_k^*(\boldsymbol{z}) = \int_0^\infty d\boldsymbol{u}(x) \int_0^\infty d\boldsymbol{D}_k(y) \int_0^y \exp\Big[\Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l\Big)(x+t)\Big] dt. \tag{3.68}$$

The proof of Theorem 3.6 is given in Appendix 3.O.

Next we discuss the computational procedure of $\boldsymbol{x}_k(\boldsymbol{n})$ ($k \in \mathscr{K}$, $\boldsymbol{n} \in \mathscr{N}$). We first rewrite (3.68) by using the uniformization technique.

$$\begin{aligned}
\boldsymbol{x}_k^*(\boldsymbol{z}) &= \sum_{m=0}^\infty \int_0^\infty d\boldsymbol{u}(x) \int_0^\infty d\boldsymbol{D}_k(y) \int_0^y \exp[-\theta(x+t)] \frac{(\theta(x+t))^m}{m!} dt \\
&\qquad \cdot \Big[\boldsymbol{I} + \theta^{-1}\Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l\Big)\Big]^m \\
&= \sum_{m=0}^\infty \sum_{i=0}^m \int_0^\infty \exp[-\theta x] \frac{(\theta x)^i}{i!} d\boldsymbol{u}(x) \int_0^\infty d\boldsymbol{D}_k(y) \int_0^y \exp[-\theta t] \frac{(\theta t)^{(m-i)}}{(m-i)!} dt \\
&\qquad \cdot \Big[\boldsymbol{I} + \theta^{-1}\Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l\Big)\Big]^m \\
&= \sum_{m=0}^\infty \sum_{i=0}^m \boldsymbol{u}^{(i)}(\theta) \tilde{\boldsymbol{D}}_k^{(m-i)}(\theta) \Big[\boldsymbol{I} + \theta^{-1}\Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l\Big)\Big]^m,
\end{aligned} \tag{3.69}$$

where $\boldsymbol{u}^{(m)}(\theta)$ $(m = 0, 1, \ldots)$ is given in Lemma 3.10 and $\tilde{\boldsymbol{D}}_k^{(m)}(\theta)$ $(k \in \mathscr{K},\ m = 0, 1, \ldots)$ is defined as

$$\tilde{\boldsymbol{D}}_k^{(m)}(\theta) = \int_0^\infty d\boldsymbol{D}_k(y) \int_0^y \exp[-\theta t] \frac{(\theta t)^m}{m!} dt.$$

Lemma 3.12. *$\tilde{\boldsymbol{D}}_k^{(m)}(\theta)$ $(k \in \mathscr{K},\ m = 0, 1, \ldots)$ is given by*

$$\tilde{\boldsymbol{D}}_k^{(m)}(\theta) = \frac{1}{\theta} \Big[\boldsymbol{D}_k - \sum_{i=0}^{m} \boldsymbol{D}_k^{(i)}(\theta) \Big]. \tag{3.70}$$

The proof of Lemma 3.12 is given in Appendix 3.P.

We define $\boldsymbol{x}_k^{(m)}(\theta)$ $(k \in \mathscr{K},\ m = 0, 1, \ldots)$ as

$$\boldsymbol{x}_k^{(m)}(\theta) = \sum_{i=0}^{m} \boldsymbol{u}^{(i)}(\theta) \tilde{\boldsymbol{D}}_k^{(m-i)}(\theta). \tag{3.71}$$

It then follows from (3.69) that

$$\boldsymbol{x}_k^*(\boldsymbol{z}) = \sum_{m=0}^{\infty} \boldsymbol{x}_k^{(m)}(\theta) \Big[\boldsymbol{I} + \theta^{-1} \Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l \Big) \Big]^m, \qquad k \in \mathscr{K}, \boldsymbol{z} \in \mathscr{Z}. \tag{3.72}$$

Let $|\boldsymbol{n}|$ $(\boldsymbol{n} \in \mathscr{N})$ denote the sum of elements of $\boldsymbol{n}$.

$$|\boldsymbol{n}| = \sum_{k \in \mathscr{K}} n_k.$$

We then define $\boldsymbol{F}_m(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathscr{N},\ |\boldsymbol{n}| \le m,\ m = 0, 1, \ldots)$ as an $M \times M$ matrix that satisfies

$$\Big[\boldsymbol{I} + \theta^{-1} \Big(\boldsymbol{C} + \sum_{l \in \mathscr{K}} z_l \boldsymbol{D}_l \Big) \Big]^m = \sum_{n \in \mathscr{N}, |n| \le m} \boldsymbol{F}_m(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K}. \tag{3.73}$$

As shown in [MT03, Lemma IV.2], $\boldsymbol{F}_m(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathscr{N},\ |\boldsymbol{n}| \le m,\ m = 0, 1, \ldots)$ can be computed recursively by

$$\boldsymbol{F}_0(\boldsymbol{0}) = \boldsymbol{I}, \tag{3.74}$$

$$\begin{aligned} \boldsymbol{F}_m(\boldsymbol{n}) &= \boldsymbol{F}_{m-1}(\boldsymbol{n})(\boldsymbol{I} + \theta^{-1}\boldsymbol{C}) \\ &\quad + \sum_{k \in \mathscr{K}, n_k \ge 1} \boldsymbol{F}_{m-1}(\boldsymbol{n} - \boldsymbol{\epsilon}_k)\theta^{-1}\boldsymbol{D}_k, \quad \boldsymbol{n} \in \mathscr{N}, |n| = 1, 2, \ldots, m-1, \end{aligned} \tag{3.75}$$

$$\boldsymbol{F}_m(\boldsymbol{n}) = \sum_{k \in \mathscr{K}, n_k \ge 1} \boldsymbol{F}_{m-1}(\boldsymbol{n} - \boldsymbol{\epsilon}_k)\theta^{-1}\boldsymbol{D}_k, \qquad \boldsymbol{n} \in \mathscr{N}, \boldsymbol{n} \ne \boldsymbol{0}, |n| = m, \tag{3.76}$$

where $\boldsymbol{\epsilon}_k$ $(k \in \mathscr{K})$ denotes a $1 \times K$ unit vector whose k-th element is equal to one and all other elements are equal to zero.

It now follows from (3.72) and (3.73) that

$$\boldsymbol{x}_k^*(\boldsymbol{z}) = \sum_{m=0}^{\infty} \boldsymbol{x}_k^{(m)}(\theta) \sum_{n \in \mathscr{N}, |n| \le m} \boldsymbol{F}_m(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K}$$

$$= \sum_{n \in \mathcal{N}} \sum_{m=|n|}^{\infty} \boldsymbol{x}_k^{(m)}(\theta)\boldsymbol{F}_m(\boldsymbol{n}) z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K}, \qquad k \in \mathcal{K}, \boldsymbol{z} \in \mathcal{Z}.$$

Therefore, $\boldsymbol{x}_k(\boldsymbol{n})$ ($k \in \mathcal{K}$, $\boldsymbol{n} \in \mathcal{N}$) can be computed by

$$\boldsymbol{x}_k(\boldsymbol{n}) = \sum_{m=|n|}^{\infty} \boldsymbol{x}_k^{(m)}(\theta)\boldsymbol{F}_m(\boldsymbol{n}). \tag{3.77}$$

In summary, we have the following theorem.

Theorem 3.7. *$\boldsymbol{y}(\boldsymbol{n})$ and $\boldsymbol{x}(\boldsymbol{n})$ ($\boldsymbol{n} \in \mathcal{N}$) are given by*

$$\begin{aligned}
\boldsymbol{y}(\boldsymbol{0}) &= (1-\nu)\boldsymbol{\kappa}, \\
\boldsymbol{y}(\boldsymbol{n}) &= \sum_{k \in \mathcal{K}, n_k \geq 1} \boldsymbol{x}_k(\boldsymbol{n} - \boldsymbol{\epsilon}_k), \ \boldsymbol{n} \neq \boldsymbol{0}, \\
\boldsymbol{x}(\boldsymbol{0}) &= (1-\nu)\boldsymbol{\kappa} + \sum_{k \in \mathcal{K}} \boldsymbol{x}_k(\boldsymbol{0}), \\
\boldsymbol{x}(\boldsymbol{n}) &= \sum_{k \in \mathcal{K}} \boldsymbol{x}_k(\boldsymbol{n}), \ \boldsymbol{n} \neq \boldsymbol{0},
\end{aligned}$$

respectively, where $\boldsymbol{x}_k(\boldsymbol{n})$ ($k \in \mathcal{K}$, $\boldsymbol{n} \in \mathcal{N}$) is given in (3.77).

3.7 Numerical examples

In this section, we show some numerical examples. Figure 3.1 shows a brief summary of the computational procedure.

We assume that the amounts of service requirements of class k ($k \in \mathcal{K}$) customers are i.i.d. according to a phase-type distribution with representation $(\boldsymbol{\alpha}_k, \boldsymbol{B}_k)$. Let $B_k^*(s)$ ($k \in \mathcal{K}$, Re$(s) > 0$) denote the LST of the amounts of service requirements of class k customers.

$$B_k^*(s) = \boldsymbol{\alpha}_k(-\boldsymbol{B}_k)(s\boldsymbol{I} - \boldsymbol{B}_k)^{-1}\boldsymbol{e}.$$

In this case, $\boldsymbol{D}_k^{(m)}(\theta)$ ($k \in \mathcal{K}$, $m = 0, 1, \ldots$) can be computed as follows. It follows from (3.49) that

$$\sum_{m=0}^{\infty} \boldsymbol{D}_k^{(m)}(\theta) z^m = B_k^*(\theta - \theta z) \cdot \boldsymbol{D}_k = \boldsymbol{\alpha}_k(-\boldsymbol{B}_k)[(\theta - \theta z)\boldsymbol{I} - \boldsymbol{B}_k]^{-1}\boldsymbol{e} \cdot \boldsymbol{D}_k.$$

We then define $\boldsymbol{H}_k(z)$ ($k \in \mathcal{K}$, $|z| < 1$) as

$$\boldsymbol{H}_k(z) = [(\theta - \theta z)\boldsymbol{I} - \boldsymbol{B}_k]^{-1}.$$

Let $\boldsymbol{H}_{k,m}$ ($k \in \mathcal{K}$, $m = 0, 1, \ldots$) denote the coefficient matrix of z^m in $\boldsymbol{H}_k(z)$. Because

$$\boldsymbol{H}(z)[(\theta\boldsymbol{I} - \boldsymbol{B}_k) - \theta\boldsymbol{I}z] = \boldsymbol{I},$$

Input : $\boldsymbol{C}$, $\boldsymbol{D}_k(x)$ $(k \in \mathscr{K})$, and $\boldsymbol{\Gamma}$.

Output : The moments of waiting times and sojourn times in (3.60)–(3.65), and the joint queue length distributions $\boldsymbol{x}(\boldsymbol{n})$ and $\boldsymbol{y}(\boldsymbol{n})$.

1. Computation of Fundamental Quantities
   - (a) Compute $\boldsymbol{Q}_{\mathrm{N}}$ by Lemma 3.3 and compute $\boldsymbol{Q}_{\mathrm{D}}$ in (3.11).
   - (b) Compute $\boldsymbol{G}_{\mathrm{N}}$ and $\boldsymbol{G}_{\mathrm{D}}$ in (3.57).
   - (c) Compute $\boldsymbol{\kappa}$ in (3.13) and ν in (3.24).
   - (d) Compute $\boldsymbol{D}_k^{(m)}(\theta)$ $(k \in \mathscr{K}, m = 0, 1, \ldots)$ in (3.48).
   - (e) Compute $\tilde{\boldsymbol{D}}_k^{(m)}(\theta)$ in (3.70).
   - (f) Compute $\boldsymbol{A}_m$ $(m = 0, 1, \ldots)$ in (3.52) and (3.53), and $\boldsymbol{E}$ in (3.54).
   - (g) Compute $\boldsymbol{K}$ in (3.58) and its invariant probability vector $\hat{\boldsymbol{\kappa}}$.
   - (h) Compute $\boldsymbol{u}^{(m)}(\theta)$ $(m = 0, 1, \ldots)$ in Lemma 3.10 and Lemma 3.11.
   - (i) Compute $\boldsymbol{x}_k^{(m)}(\theta)$ $(k \in \mathscr{K},\ m = 0, 1, \ldots)$ in (3.71).
   - (j) Compute $\boldsymbol{F}_m(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathscr{N},\ |\boldsymbol{n}| \le m,\ m = 0, 1, \ldots)$ in (3.74), (3.75), and (3.76).
2. Computation of Waiting Time and Sojourn Time Moments
   - (a) Compute $\boldsymbol{w}_{\mathrm{N},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{N},k}^{(n)}$ $(k \in \mathscr{K},\ n = 0, 1, \ldots)$ in Theorem 3.4.
   - (b) Compute $\boldsymbol{w}_{\mathrm{D},k}^{(n)}$ and $\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}$ $(k \in \mathscr{K},\ n = 0, 1, \ldots)$ in Theorem 3.5.
   - (c) Compute the moments of waiting time and sojourn time distributions in (3.60)–(3.65).
3. Computation of the Joint Queue Length Distribution
   - (a) Compute $\boldsymbol{x}_k(\boldsymbol{n})$ $(k \in \mathscr{K}, \boldsymbol{n} \in \mathscr{N})$ in (3.77).
   - (b) Compute $\boldsymbol{x}(\boldsymbol{n})$ and $\boldsymbol{y}(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathscr{N})$ in Theorem 3.7.

Figure 3.1: Brief summary of the computational procedure.

Table 3.1: Joint queue length distribution $\boldsymbol{y}(n_1, n_2)\boldsymbol{e}$ $(a = 0.1)$.

n_2 n_1	0	1	2	3	4
0	0.30776170	0.05304471	0.01592757	0.00575379	0.00216682
1	0.12508144	0.04994412	0.02100711	0.00939074	0.00425087
2	0.05167721	0.03280510	0.01803788	0.00972928	0.00514504
3	0.02431553	0.02079512	0.01377549	0.00851632	0.00506414
4	0.01251500	0.01328265	0.01010444	0.00690758	0.00448230

we have

$$\begin{aligned} \boldsymbol{H}_0 &= (\theta \boldsymbol{I} - \boldsymbol{B}_k)^{-1}, \\ \boldsymbol{H}_m &= \boldsymbol{H}_{m-1}(\boldsymbol{I} - \theta^{-1}\boldsymbol{B}_k)^{-1}, \qquad m = 1, 2, \ldots. \end{aligned}$$

Therefore $\boldsymbol{D}_k^{(m)}(\theta)$ $(m = 0, 1, \ldots)$ is given by

$$\boldsymbol{D}_k^{(m)}(\theta) = \boldsymbol{\alpha}_k(-\boldsymbol{B}_k)\boldsymbol{H}_m \boldsymbol{e} \cdot \boldsymbol{D}_k$$

In the rest of this section, we assume $\mathscr{K} = \{1, 2\}$ and consider the case that $\boldsymbol{C}$, $\boldsymbol{D}_k$ $(k = 1, 2)$, and $\boldsymbol{\Gamma}$ are given by

$$\boldsymbol{C} = \begin{pmatrix} -0.65 & 0.05 \\ 0.1 & -1.4 - a \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 0.3 & 0 \\ 0 & 1.0 \end{pmatrix}, \quad \boldsymbol{D}_2 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix},$$

respectively, where a $(a > 0)$ is a parameter. We also assume that $\boldsymbol{\alpha}_k$ and $\boldsymbol{B}_k$ $(k = 1, 2)$ are given by

$$\boldsymbol{\alpha}_1 = (1, 0), \boldsymbol{B}_1 = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}, \qquad \boldsymbol{\alpha}_2 = (0.2, 0.8), \boldsymbol{B}_2 = \begin{pmatrix} -0.25 & 0 \\ 0 & -4 \end{pmatrix},$$

respectively. We thus have $\boldsymbol{\pi} = (2/3, 1/3)$. Note that the mean amount of service requirements is equal to one in both classes, so that the traffic intensity is about 0.83.

Table 3.1 shows the joint queue length distribution $\boldsymbol{y}(n_1, n_2)\boldsymbol{e}$ when $a = 0.1$. We observe that in steady state, the system is busy with probability about $1 - 0.31 = 0.69$, while the traffic intensity is about 0.83. The discrepancy between those is due to disasters. We also observe that when we fix one of n_1 and n_2, $\boldsymbol{y}(n_1, n_2)$ is not necessarily a decreasing function of the other. For example, $\boldsymbol{y}(n_1, 4)$ takes its maximum value for $n_1 = 2$. This indicates that the queue lengths are (positively) correlated.

Finally, Figure 3.2 shows the distribution mass function of the total number of customers, where a is set to be 1, 0.1, 0.01, 0.001, and 0.0001. The frequency of disasters decreases with a, and the queue length distribution converges to that in the corresponding queue without disasters.

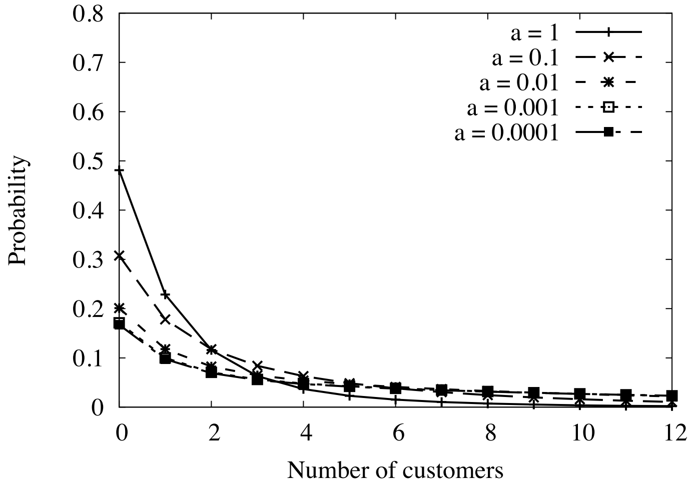

Figure 3.2: Total queue length mass function.

3.8 Conclusion

In this chapter, we considered the multi-class FCFS MAP/G/1 queue with disasters. We assumed that arrivals of customers and occurrences of disasters are governed by a common underlying Markov chain, and the amounts of service requirements brought by arriving customers follow general distributions, which depend on the customer class and the states of the underlying Markov chain immediately before and after arrivals.

We first analyzed the first passage time to the idle state and the busy cycle. We then derived two different formulas for the LST of the stationary total workload in system, and showed that those formulas are equivalent in a sense that one can be derived from another. Furthermore, using the result on the workload distribution, we analyzed the waiting time and sojourn time distributions of each class. We also analyzed the joint queue length distribution, and provided its computational procedure. We also showed some numerical examples.

The model we analyzed in this chapter is a generalization of the multi-class FCFS M/G/1 queue with Poisson disasters considered in Appendix 2.I. As mentioned in Chapter 1, this model is closely related to the multi-class FCFS MAP/G/1 queue with working vacations. Analytical results for the corresponding queueing model with working vacations can be derived from the results in this chapter, noting that the censored process obtained by removing all normal service periods in the working vacation model is equivalent to the disaster model considered in this chapter.

Appendices

3.A Proof of Lemma 3.1

It is easy to see that for $x \geq 0$ and $y \geq 0$,

$$\boldsymbol{P}^*_{\mathrm{N}}(s \mid x+y) = \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{P}^*_{\mathrm{N}}(s \mid y).$$

We thus have

$$\begin{aligned}\boldsymbol{P}^*_{\mathrm{N}}(s \mid x+\Delta x) &= \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{P}^*_{\mathrm{N}}(s \mid \Delta x)\\ &= \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\Big[\boldsymbol{I} - s\boldsymbol{I}\Delta x + \boldsymbol{C}\Delta x + \int_0^\infty d\boldsymbol{D}(y)\Delta x \boldsymbol{P}^*_{\mathrm{N}}(s \mid y) + \boldsymbol{o}(\Delta x)\Big].\end{aligned}$$

It then follows that

$$\frac{\partial}{\partial x}\left[\boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\right] = \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{Q}^*_{\mathrm{N}}(s), \tag{3.78}$$

where $\boldsymbol{Q}^*_{\mathrm{N}}(s)$ is given by (3.4). (3.3) now follows from (3.78) and $\boldsymbol{P}^*_{\mathrm{N}}(s \mid 0) = \boldsymbol{I}$.

Similarly, it is readily verified that

$$\boldsymbol{P}^*_{\mathrm{D}}(s \mid x+y) = \boldsymbol{P}^*_{\mathrm{D}}(s \mid x) + \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{P}^*_{\mathrm{D}}(s \mid y),$$

and therefore

$$\begin{aligned}\boldsymbol{P}^*_{\mathrm{D}}(s \mid x+\Delta x) &= \boldsymbol{P}^*_{\mathrm{D}}(s \mid x) + \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{P}^*_{\mathrm{D}}(s \mid \Delta x)\\ &= \boldsymbol{P}^*_{\mathrm{D}}(s \mid x) + \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\Big[\Gamma\Delta x + \int_0^\infty d\boldsymbol{D}(y)\Delta x \boldsymbol{P}^*_{\mathrm{D}}(s \mid y) + \boldsymbol{o}(\Delta x)\Big].\end{aligned}$$

We thus have

$$\frac{\partial}{\partial x}\left[\boldsymbol{P}^*_{\mathrm{D}}(s \mid x)\right] = \boldsymbol{P}^*_{\mathrm{N}}(s \mid x)\boldsymbol{Q}^*_{\mathrm{D}}(s) = \exp\left[\boldsymbol{Q}^*_{\mathrm{N}}(s)x\right]\boldsymbol{Q}^*_{\mathrm{D}}(s), \tag{3.79}$$

where $\boldsymbol{Q}^*_{\mathrm{D}}(s)$ is given by (3.6). (3.5) now follows from (3.79) and $\boldsymbol{P}^*_{\mathrm{D}}(s \mid 0) = \boldsymbol{0}$. □

3.B Proof of Lemma 3.2

We define $\boldsymbol{J}$ as an $M \times M$ matrix given by

$$\boldsymbol{J} = \boldsymbol{I} - \int_0^\infty d\boldsymbol{D}(y)\int_0^y \exp[\boldsymbol{Q}_{\mathrm{N}}w]dw. \tag{3.80}$$

By definition, $\boldsymbol{J}$ satisfies

$$\begin{aligned}\boldsymbol{J}\boldsymbol{Q}_{\mathrm{N}} &= \boldsymbol{Q}_{\mathrm{N}} - \int_0^\infty d\boldsymbol{D}(y)\left[\exp[\boldsymbol{Q}_{\mathrm{N}}y] - \boldsymbol{I}\right] = \boldsymbol{Q}_{\mathrm{N}} - \left[(\boldsymbol{Q}_{\mathrm{N}} - \boldsymbol{C}) - \boldsymbol{D}\right]\\ &= \boldsymbol{C} + \boldsymbol{D}.\end{aligned} \tag{3.81}$$

Because $\boldsymbol{C}+\boldsymbol{D}$ is non-singular (see Remark 3.1),

$$\det(\boldsymbol{J}\boldsymbol{Q}_{\mathrm{N}}) = \det(\boldsymbol{J})\det(\boldsymbol{Q}_{\mathrm{N}}) = \det(\boldsymbol{C}+\boldsymbol{D}) \neq 0.$$

We thus have $\det(\boldsymbol{J}) \neq 0$ and $\det(\boldsymbol{Q}_{\mathrm{N}}) \neq 0$, and therefore $\boldsymbol{J}$ and $\boldsymbol{Q}_{\mathrm{N}}$ are non-singular. Using (3.10) and (3.80), we have

$$\boldsymbol{Q}_{\mathrm{D}} = \boldsymbol{J}^{-1}\boldsymbol{\Gamma},$$

from which and (3.81), (3.11) follows. □

3.C Proof of Lemma 3.3

We prove the lemma by providing a probabilistic interpretation of $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$. Suppose customers are served under the LCFS-PR basis. Note that this service discipline is work conserving. Let $\mathscr{A}^{(k)}$ $(k = 0,1,\ldots)$ denote the set of customers who find k customers in the system on arrival. Due to the LCFS-PR service discipline, all customers in $\mathscr{A}^{(k)}$ $(k = 1,2,\ldots)$ arrive in service periods of customers in $\mathscr{A}^{(k-1)}$. Let $\mathscr{B}^{(n)}$ $(n = 1,2,\ldots)$ denote the set of busy periods in which no customers in $\mathscr{A}^{(l)}$ $(l = n, n+1,\ldots)$ arrive. Furthermore, let $\mathscr{B}_{\mathrm{N}}^{(n)}$ $(n = 1,2,\ldots)$ denote a subset of busy periods in $\mathscr{B}^{(n)}$, in which no disasters occur. By definition, $\mathscr{B}_{\mathrm{N}}^{(1)} \subset \mathscr{B}_{\mathrm{N}}^{(2)} \subset \cdots \subset \mathscr{B}_{\mathrm{N}}^{(\infty)}$, where $\mathscr{B}_{\mathrm{N}}^{(\infty)}$ denotes the set of all busy periods in which no disasters occur.

It is clear that $\boldsymbol{Q}_{\mathrm{N}}^{(0)} = \boldsymbol{C}$ represents the infinitesimal generator of the censored underlying Markov chain when there are neither arrivals nor disasters. Also, $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$ $(n = 1,2,\ldots)$ in (3.15) represents the infinitesimal generator of the censored underlying Markov chain when (i) there are neither arrivals nor disasters, or (ii) busy periods removed from the time axis belong to $\mathscr{B}_{\mathrm{N}}^{(n)}$. This implies that $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$ $(n = 0,1,\ldots)$ increases elementwise with n and converges to $\boldsymbol{Q}_{\mathrm{N}}$. Note that a similar observation has been made in [TH94] for the ordinary MAP/G/1 queue without disasters.

3.D Proof of Lemma 3.4

Pre-multiplying both sides of (3.5) by $-\boldsymbol{Q}_{\mathrm{N}}^{*}(s)$ and calculating the integral on the right-hand side yield

$$\left(-\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right)\boldsymbol{P}_{\mathrm{D}}^{*}(s \mid x) = \left[\boldsymbol{I} - \exp\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)x\right]\right]\boldsymbol{Q}_{\mathrm{D}}^{*}(s). \tag{3.82}$$

Using (3.2), (3.3), and (3.82), we have

$$\begin{aligned}\left(-\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right)\boldsymbol{P}^{*}(s \mid x) &= \left(-\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right)\exp\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)x\right] + \left[\boldsymbol{I} - \exp\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right]\right]\boldsymbol{Q}_{\mathrm{D}}^{*}(s) \\ &= \boldsymbol{Q}_{\mathrm{D}}^{*}(s) - \exp\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right]\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s) + \boldsymbol{Q}_{\mathrm{D}}^{*}(s)\right],\end{aligned} \tag{3.83}$$

where we use $\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\exp[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)] = \exp[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)]\boldsymbol{Q}_{\mathrm{N}}^{*}(s)$. Post-multiplying both sides of (3.83) by $\boldsymbol{e}$, taking the partial derivative with respect to s, and taking the limit $s \to 0+$, we obtain

$$\begin{aligned}\left(-\boldsymbol{Q}'_{\mathrm{N}}(0)\right)\boldsymbol{e} + (-\boldsymbol{Q}_{\mathrm{N}}) \cdot \lim_{s\to 0+} \frac{\partial}{\partial s}\left[\boldsymbol{P}^{*}(s \mid x)\right]\boldsymbol{e}& \\ = \boldsymbol{Q}'_{\mathrm{D}}(0)\boldsymbol{e} - \lim_{s\to 0+}\frac{\partial}{\partial s}&\left[\exp\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)x\right]\right][\boldsymbol{Q}_{\mathrm{N}} + \boldsymbol{Q}_{\mathrm{D}}]\boldsymbol{e} \\ &- \exp[\boldsymbol{Q}_{\mathrm{N}}x]\left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e},\end{aligned}$$

where

$$\boldsymbol{Q}'_{\mathrm{N}}(0) = \lim_{s\to 0+}\frac{d}{ds}\left[\boldsymbol{Q}_{\mathrm{N}}^{*}(s)\right], \quad \boldsymbol{Q}'_{\mathrm{D}}(0) = \lim_{s\to 0+}\frac{d}{ds}\left[\boldsymbol{Q}_{\mathrm{D}}^{*}(s)\right].$$

It then follows from (3.12) and (3.16) that

$$\boldsymbol{f}(x) = (-\boldsymbol{Q}_{\mathrm{N}})^{-1}[\boldsymbol{I} - \exp[\boldsymbol{Q}_{\mathrm{N}}x]]\left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e}. \tag{3.84}$$

Furthermore, with (3.4) and (3.6), we have

$$\begin{aligned}\left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e} &= \boldsymbol{e} + \int_0^{\infty} d\boldsymbol{D}(y)\boldsymbol{f}(y) \\ &= \boldsymbol{e} + \int_0^{\infty} d\boldsymbol{D}(y)(-\boldsymbol{Q}_{\mathrm{N}})^{-1}[\boldsymbol{I} - \exp[\boldsymbol{Q}_{\mathrm{N}}y]] \cdot \left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e} \\ &= \boldsymbol{e} + [\boldsymbol{D} - (\boldsymbol{Q}_{\mathrm{N}} - \boldsymbol{C})](-\boldsymbol{Q}_{\mathrm{N}})^{-1} \cdot \left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e} \\ &= \boldsymbol{e} + [\boldsymbol{I} + (\boldsymbol{C} + \boldsymbol{D})(-\boldsymbol{Q}_{\mathrm{N}})^{-1}]\left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e},\end{aligned}$$

which implies

$$\left[\boldsymbol{Q}'_{\mathrm{N}}(0) + \boldsymbol{Q}'_{\mathrm{D}}(0)\right]\boldsymbol{e} = (-\boldsymbol{Q}_{\mathrm{N}})\left[-(\boldsymbol{C} + \boldsymbol{D})\right]^{-1}\boldsymbol{e}.$$

(3.17) then follows from (3.84) and $[\boldsymbol{I} - \exp[\boldsymbol{Q}_{\mathrm{N}}x]](-\boldsymbol{Q}_{\mathrm{N}}) = (-\boldsymbol{Q}_{\mathrm{N}})[\boldsymbol{I} - \exp[\boldsymbol{Q}_{\mathrm{N}}x]]$. □

3.E Proof of Lemma 3.5

Pre-multiplying both sides of (3.18) by $\omega\boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{\Gamma})$, setting $\omega = s$, and taking the partial derivative with respect to s, we have

$$[s\boldsymbol{I} - (\boldsymbol{C} + \boldsymbol{\Gamma})]\frac{d}{ds}\left[\boldsymbol{\Phi}^{**}(s,s)\right] + \boldsymbol{\Phi}^{**}(s,s) = \int_0^{\infty} d\boldsymbol{D}(y)\frac{\partial}{\partial s}\left[\boldsymbol{P}^{*}(s \mid y)\right]. \tag{3.85}$$

Furthermore, post-multiplying both sides of (3.85) by $\boldsymbol{e}$, taking the limit $s \to 0+$, and rearranging terms yield

$$\begin{aligned}[-(\boldsymbol{C} + \boldsymbol{\Gamma})]\boldsymbol{\phi} &= \boldsymbol{e} + \int_0^{\infty} d\boldsymbol{D}(y)\boldsymbol{f}(y) = \boldsymbol{e} + [\boldsymbol{D} - (\boldsymbol{Q}_{\mathrm{N}} - \boldsymbol{C})][-(\boldsymbol{C} + \boldsymbol{D})]^{-1}\boldsymbol{e} \\ &= (-\boldsymbol{Q}_{\mathrm{N}})[-(\boldsymbol{C} + \boldsymbol{D})]^{-1}\boldsymbol{e},\end{aligned}$$

where we use (3.9), (3.16), (3.17), and (3.19). (3.22) now follows immediately. □

3.F Proof of Theorem 3.1

Considering the transition from time t to time $t+\Delta t$, we have

$$\boldsymbol{u}_{t+\Delta t}(x) = \boldsymbol{u}_t(x+\Delta t)(\boldsymbol{I}+\boldsymbol{C}\Delta t)+\int_0^x \boldsymbol{u}_t(x-y+\Delta t)d\boldsymbol{D}(y)\Delta t+\boldsymbol{u}_t(\infty)\boldsymbol{\Gamma}\Delta t+\boldsymbol{o}(\Delta t),$$

where $\boldsymbol{u}_t(\infty)=\lim_{x\to\infty}\boldsymbol{u}_t(x)$. It then follows that

$$\frac{\partial}{\partial t}\left[\boldsymbol{u}_t(x)\right]=\frac{\partial}{\partial x}\left[\boldsymbol{u}_t(x)\right]+\boldsymbol{u}_t(x)\boldsymbol{C}+\int_0^x \boldsymbol{u}_t(x-y)d\boldsymbol{D}(y)+\boldsymbol{u}_t(\infty)\boldsymbol{\Gamma}. \tag{3.86}$$

Because the system is stable, taking the limit $t\to\infty$ in (3.86) yields

$$0=\frac{d}{dx}\left[\boldsymbol{u}(x)\right]+\boldsymbol{u}(x)\boldsymbol{C}+\int_0^\infty \boldsymbol{u}(x-y)d\boldsymbol{D}(y)+\boldsymbol{\pi}\boldsymbol{\Gamma}, \tag{3.87}$$

where we use $\lim_{t\to\infty}\boldsymbol{u}_t(\infty)=\boldsymbol{\pi}$. Furthermore, taking the LST on both sides of (3.87), we obtain

$$0=\boldsymbol{u}^*(s)-\boldsymbol{u}(0)+\boldsymbol{u}^*(s)\boldsymbol{C}/s+\boldsymbol{u}^*(s)\boldsymbol{D}^*(s)/s+\boldsymbol{\pi}\boldsymbol{\Gamma}/s.$$

Note here that $\boldsymbol{u}(0)=\lim_{t\to\infty}\Pr(U_t=0)$ is given by

$$\boldsymbol{u}(0)=(1-\nu)\boldsymbol{\kappa}.$$

(3.25) now follows from the above two equations. □

3.G Proof of Lemma 3.7

We can prove Lemma 3.7 in the same way as in [Tak02], where a similar result for the ordinary MAP/G/1 queue is discussed. Therefore we provide only the outline of the proof. Let $\hat{H}_j(x,n)$ ($j\in\mathcal{M}$, $x\ge 0$, $n=1,2,\ldots$) denote the time-average joint probability that a customer who found $n-1$ customers in the system on arrival is being served, the workload in system is not greater than x, and the underlying Markov chain is in state j. Also, let $\hat{Y}_j(x,n)$ ($j\in\mathcal{M}, x\ge 0, n=1,2,\ldots$) denote the mean length of time in which a randomly chosen customer, who found $n-1$ customers in the system on arrival, is served, the workload in system is not greater than x, and the underlying Markov chain is in state j.

By definition, $\hat{H}_j(x,n)$ ($j\in\mathcal{M}$) is identical to the j-th element of $\boldsymbol{u}(x,n)$. On the other hand, $\hat{Y}_j(x,n)$ is given by

$$\hat{Y}_j(x,n)=\left[\int_0^x \frac{\boldsymbol{u}(dw,n-1)}{\boldsymbol{u}(\infty,n-1)\boldsymbol{D}\boldsymbol{e}}\int_0^{x-w}dt\int_t^\infty d\boldsymbol{D}(y)\exp\left[\boldsymbol{Q}_{\mathrm{N}}(y-t)\right]\right]_j,$$

where $\boldsymbol{u}(\infty,n-1)=\lim_{x\to\infty}\boldsymbol{u}(x,n-1)$ ($n=1,2,\ldots$). Note that the arrival rate of customers who find $n-1$ customers in the system on arrival is given by $\boldsymbol{u}(\infty,n-$

1)$\boldsymbol{De}$. Owing to the relation $H = \lambda G$ between time and customer averages [HS80, Tak02], we have

$$\boldsymbol{u}(x,n) = \int_0^x \boldsymbol{u}(dw, n-1) \int_0^{x-w} dt \int_t^\infty d\boldsymbol{D}(y) \exp\big[\boldsymbol{Q}_{\rm N}(y-t)\big].$$

Therefore, taking the LST with respect to x yields

$$\boldsymbol{u}^*(s,n) = \boldsymbol{u}^*(s,n-1)\boldsymbol{R}^*(s), \quad n = 1,2,\dots, \tag{3.88}$$

from which and $\boldsymbol{u}^*(s,0) = (1-\nu)\boldsymbol{\kappa}$, (3.27) follows. Furthermore, (3.29) is derived from (3.28) as follows.

$$\begin{aligned}\boldsymbol{R}^*(s)(s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}) &= \int_0^\infty d\boldsymbol{D}(y)\exp[\boldsymbol{Q}_{\rm N}y]\int_0^y \exp\big[-(s\boldsymbol{I}+\boldsymbol{Q}_{\rm N})x\big]dx \cdot (s\boldsymbol{I}+\boldsymbol{Q}_{\rm N}) \\ &= \int_0^\infty d\boldsymbol{D}(y)\big[\exp[\boldsymbol{Q}_{\rm N}y] - \exp[-sy]\boldsymbol{I}\big] \\ &= \boldsymbol{Q}_{\rm N} - \boldsymbol{C} - \boldsymbol{D}^*(s).\end{aligned}$$

□

3.H Proof of Lemma 3.8

With (3.3) and (3.28), we have

$$\boldsymbol{R} = \int_0^\infty d\boldsymbol{D}(y) \int_0^y \boldsymbol{P}^*(0 \mid y-x)dx. \tag{3.89}$$

Because $\boldsymbol{D} \ge 0$ and $\boldsymbol{D} \ne \boldsymbol{0}$, (3.89) implies $\boldsymbol{R} \ge \boldsymbol{0}$ and $\boldsymbol{R} \ne \boldsymbol{0}$. It then follows from Theorem 3 in [Gan59, Page 66] that $\boldsymbol{R}$ has the maximum modulus real eigenvalue $\mu \ge 0$ and a non-negative right eigenvector $\boldsymbol{v} \ne \boldsymbol{0}$ associated with μ.

With (3.14) and (3.31), we have $\boldsymbol{\pi R} < \boldsymbol{\pi}$. Post-multiplying both sides of this inequality by $\boldsymbol{v}$ yields

$$\boldsymbol{\pi R v} = \mu \boldsymbol{\pi v} < \boldsymbol{\pi v}.$$

Because $\boldsymbol{\pi v} > 0$, we have $\mu < 1$, from which Lemma 3.8 follows. □

3.I Proof of Theorem 3.2

Using (3.27), we have

$$\boldsymbol{u}^*(s) = \sum_{n=0}^\infty (1-\nu)\boldsymbol{\kappa}[\boldsymbol{R}^*(s)]^n, \qquad \mathrm{Re}(s) > 0. \tag{3.90}$$

Because $\boldsymbol{R} \ge \boldsymbol{0}$, the (i,j)th $(i,j \in \mathscr{M})$ element of $\boldsymbol{R}^*(s)$ satisfies

$$|[\boldsymbol{R}^*(s)]_{i,j}| \le [\boldsymbol{R}]_{i,j}, \qquad \mathrm{Re}(s) > 0,\ i,j \in \mathscr{M}.$$

Furthermore, Lemma 3.8 implies that $\sum_{n=0}^\infty \boldsymbol{R}^n$ converges. Therefore $\sum_{n=0}^\infty [\boldsymbol{R}^*(s)]^n$ (Re$(s) > 0$) also converges, so that $\boldsymbol{I} - \boldsymbol{R}^*(s)$ (Re$(s) > 0$) is non-singular. (3.33) then follows from (3.90). □

3.J Proof of Theorem 3.3

By definition, $\boldsymbol{w}_{\mathrm{N},k}(x)$ and $\overline{\boldsymbol{w}}_{\mathrm{N},k}(x)$ are given by

$$\begin{aligned}
\boldsymbol{w}_{\mathrm{N},k}(x) &= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^x d\boldsymbol{u}(y)\boldsymbol{D}_k\exp\big[(\boldsymbol{C}+\boldsymbol{D})y\big],\\
\overline{\boldsymbol{w}}_{\mathrm{N},k}(x) &= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^x d\boldsymbol{u}(y)\int_0^{x-y} d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D})(y+v)\big],
\end{aligned}$$

respectively. Taking the LST on both sides of these equations, we have (3.39) and (3.40). Similarly, $\boldsymbol{w}_{\mathrm{D},k}(x)$ and $\overline{\boldsymbol{w}}_{\mathrm{D},k}(x)$ are given by

$$\boldsymbol{w}_{\mathrm{D},k}(x) = \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^x dt\int_t^\infty d\boldsymbol{u}(y)\boldsymbol{D}_k\exp\big[(\boldsymbol{C}+\boldsymbol{D})t\big]\boldsymbol{\Gamma}, \tag{3.91}$$

$$\begin{aligned}
\overline{\boldsymbol{w}}_{\mathrm{D},k}(x) = \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^x dt\Big[&\int_0^\infty d\boldsymbol{u}(y)\int_0^\infty d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D})t\big]\boldsymbol{\Gamma}\\
&-\int_0^t d\boldsymbol{u}(y)\int_0^{t-y} d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D})t\big]\boldsymbol{\Gamma}\Big],
\end{aligned} \tag{3.92}$$

respectively. Taking the LST on both sides of (3.91), we obtain

$$\begin{aligned}
\boldsymbol{w}^*_{\mathrm{D},k}(s) &= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^\infty dx\int_x^\infty d\boldsymbol{u}(y)\boldsymbol{D}_k\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}\\
&= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^\infty d\boldsymbol{u}(y)\boldsymbol{D}_k\int_0^y \exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}dx\\
&= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^\infty d\boldsymbol{u}(y)\boldsymbol{D}_k\big[\boldsymbol{I}-\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})y\big]\big][s\boldsymbol{I}-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma},
\end{aligned}$$

from which and (3.39), (3.41) follows. Also, taking the LST on both sides of (3.92) yields

$$\begin{aligned}
\overline{\boldsymbol{w}}^*_{\mathrm{D},k}(s) &= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\int_0^\infty dx\Big[\int_0^\infty d\boldsymbol{u}(y)\int_0^\infty d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}\\
&\qquad -\int_0^x d\boldsymbol{u}(y)\int_0^{x-y} d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}\Big]\\
&= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\Big[\boldsymbol{\pi D}_k\int_0^\infty \exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}dx\\
&\qquad -\int_0^\infty d\boldsymbol{u}(y)\int_y^\infty dx\int_0^{x-y} d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})x\big]\boldsymbol{\Gamma}\Big]\\
&= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\Big[\boldsymbol{\pi D}_k[s\boldsymbol{I}-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma}\\
&\qquad -\int_0^\infty d\boldsymbol{u}(y)\int_0^\infty d\boldsymbol{D}_k(v)\int_v^\infty dx'\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})(y+x')\big]\boldsymbol{\Gamma}\Big]\\
&= \frac{1}{\boldsymbol{\pi D}_k\boldsymbol{e}}\Big[\boldsymbol{\pi D}_k-\int_0^\infty d\boldsymbol{u}(y)\int_0^\infty d\boldsymbol{D}_k(v)\exp\big[(\boldsymbol{C}+\boldsymbol{D}-s\boldsymbol{I})(y+v)\big]\Big]
\end{aligned}$$

$$\cdot[s\boldsymbol{I}-(\boldsymbol{C}+\boldsymbol{D})]^{-1}\boldsymbol{\Gamma},$$

where $x' = x - y$. (3.42) now follows from (3.40). □

3.K Proof of Lemma 3.9

By definition, we have for $m = 0,1,\ldots$ and $n = 1,2,\ldots,$

$$\begin{aligned}\frac{\partial^n}{\partial s^n}\left[\boldsymbol{u}^{(m)}(s,\theta)\right] &= \int_0^\infty (-x)^n \exp\left[-(s+\theta)x\right]\frac{(\theta x)^m}{m!}d\boldsymbol{u}(x)\\ &= (-1)^n\int_0^\infty \exp\left[-(s+\theta)x\right]\frac{(\theta x)^{n+m}}{(n+m)!}\cdot\frac{(n+m)!}{m!\theta^n}d\boldsymbol{u}(x).\end{aligned}$$

We thus obtain (3.50) by taking the limit $s \to 0+$. (3.51) can be derived in the same way, so that we omit the proof. □

3.L Proof of Lemma 3.10

Because Lemma 3.10 can be proved in the same way as the proof of Lemma 3 in [Tak01], we provide only the outline of the proof. By definition, $\boldsymbol{u}^{(m)}(\theta)$ $(m = 0,1,\ldots)$ satisfies

$$\boldsymbol{u}^*(\theta-\theta z) = \sum_{m=0}^{\infty}\boldsymbol{u}^{(m)}(\theta)z^m, \qquad |z| < 1.$$

On the other hand, it follows from (3.25) that

$$\boldsymbol{u}^*(\theta-\theta z)[(\theta-\theta z)+\boldsymbol{C}+\boldsymbol{D}^*(\theta-\theta z)] = (\theta-\theta z)(1-\nu)\boldsymbol{\kappa}-\boldsymbol{\pi}\boldsymbol{\Gamma}, \quad |z|<1.$$

Comparing the coefficient of z^m $(m = 0,1,\ldots)$ on both sides of this equation, we obtain

$$\begin{aligned}\boldsymbol{u}^{(0)}(\theta) &= \boldsymbol{u}^{(0)}(\theta)(\boldsymbol{A}_0+\boldsymbol{A}_1)+\boldsymbol{u}^{(1)}(\theta)\boldsymbol{A}_0+\sum_{i=0}^{\infty}\boldsymbol{u}^{(i)}(\theta)\boldsymbol{F},\\ \boldsymbol{u}^{(m)}(\theta) &= \sum_{i=0}^{m+1}\boldsymbol{u}^{(i)}(\theta)\boldsymbol{A}_{m-i+1},\end{aligned}$$

from which (3.55) follows. Note that $\boldsymbol{T}$ is stochastic because

$$\left[\mathrm{E}+\sum_{m=0}^{\infty}\boldsymbol{A}_m\right]\boldsymbol{e} = \theta^{-1}(\boldsymbol{C}+\boldsymbol{D}+\boldsymbol{\Gamma})\boldsymbol{e}+\boldsymbol{e} = \boldsymbol{e}.$$

Furthermore, (3.56) immediately follows from $\sum_{m=0}^{\infty}\boldsymbol{u}^{(m)}(\theta) = \boldsymbol{\pi}$. □

3.M Proof of Lemma 3.11

It follows from (3.59) that

$$\boldsymbol{u}^{(0)}(\theta) = (1-\hat{\nu})\hat{\boldsymbol{\kappa}},$$

where $\hat{\nu} = \sum_{m=1}^{\infty} \boldsymbol{u}^{(m)}(\theta)\boldsymbol{e}$. Substituting $s = \theta$ in (3.25), we have

$$\boldsymbol{u}^{(0)}(\theta)(\theta\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^{(0)}(\theta)) = \theta(1-\nu)\boldsymbol{\kappa} - \boldsymbol{\pi}\boldsymbol{\Gamma}.$$

Therefore we obtain

$$(1-\hat{\nu})\hat{\boldsymbol{\kappa}}\boldsymbol{A}_0 = (1-\nu)\boldsymbol{\kappa} - \boldsymbol{\pi}\mathrm{E},$$

or equivalently,

$$1-\hat{\nu} = \frac{1-\nu-\boldsymbol{\pi}\mathrm{E}\boldsymbol{e}}{\hat{\boldsymbol{\kappa}}\boldsymbol{A}_0\boldsymbol{e}},$$

from which Lemma 3.11 follows. □

3.N Proof of Theorem 3.5

Because the proof for $\overline{\boldsymbol{w}}_{\mathrm{D},k}^{(n)}$ is almost the same as that for $\boldsymbol{w}_{\mathrm{D},k}^{(n)}$, we provide the proof only for $\boldsymbol{w}_{\mathrm{D},k}^{(n)}$. It follows from Theorem 3.3 that

$$\boldsymbol{w}_{\mathrm{D},k}^{*}(s) = \boldsymbol{\beta}_{\mathrm{D},k}^{*}(s)\boldsymbol{\Gamma},$$

where

$$\boldsymbol{\beta}_{\mathrm{D},k}^{*}(s) = \left(\frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \boldsymbol{w}_{\mathrm{N},k}^{*}(s)\right)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})]^{-1}.$$

We thus have

$$\boldsymbol{\beta}_{\mathrm{D},k}^{*}(s)[s\boldsymbol{I} - (\boldsymbol{C}+\boldsymbol{D})] = \frac{\boldsymbol{\pi}\boldsymbol{D}_k}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} - \boldsymbol{w}_{\mathrm{N},k}^{*}(s). \tag{3.93}$$

We then define $\boldsymbol{\beta}_{\mathrm{D},k}^{(n)}$ ($k \in \mathscr{K}$, $n = 0,1,\ldots$) as

$$\boldsymbol{\beta}_{\mathrm{D},k}^{(0)} = \lim_{s\to 0+} \boldsymbol{\beta}_{\mathrm{D},k}^{*}(s), \qquad \boldsymbol{\beta}_{\mathrm{D},k}^{(n)} = \lim_{s\to 0+} \frac{d^n}{ds^n}\boldsymbol{\beta}_{\mathrm{D},k}^{*}(s), \qquad n = 1,2,\ldots.$$

The result for $\boldsymbol{w}_{\mathrm{D},k}^{(0)}$ is immediate. For $n = 1,2,\ldots$, differentiating both sides of (3.93), taking the limit $s \to +0$, and rearranging terms, we obtain the recursion for $\boldsymbol{\beta}_{\mathrm{D},k}^{(n)}$, which completes the proof. □

3.O Proof of Theorem 3.6

We prove Theorem 3.6 with an observation similar to the proof of Lemma 3.7. Let $\mathrm{E}[T_{k,j}(\boldsymbol{n})]$ ($k \in \mathcal{K}$, $j \in \mathcal{M}$, $\boldsymbol{n} \in \mathcal{N}$) denote the mean length of time in which a randomly chosen class k customer is served, the numbers of waiting customers of class i is equal to n_i for each $i \in \mathcal{K}$, and the underlying Markov chain is in state j. Note that all waiting customers arrived in the waiting time and the elapsed service time of the customer being served because of the FCFS service discipline. We thus obtain

$$\begin{aligned}
&\sum_{n \in \mathcal{N}} \mathrm{E}[T_{k,j}(\boldsymbol{n})] z_1^{n_1} z_2^{n_2} \cdots z_K^{n_K} \\
&\quad = \left[\int_0^\infty \frac{d\boldsymbol{u}(x)}{\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}} \int_0^\infty d\boldsymbol{D}_k(y) \int_0^y \exp\left[\left(\boldsymbol{C} + \sum_{l \in \mathcal{K}} z_l \boldsymbol{D}_l\right)(x+t)\right] dt\right]_j, \\
&\qquad\qquad k \in \mathcal{K}, j \in \mathcal{M}, \boldsymbol{z} \in \mathcal{Z}. \qquad (3.94)
\end{aligned}$$

On the other hand, by definition, the j-th ($j \in \mathcal{M}$) element of $\boldsymbol{x}_k(\boldsymbol{n})$ ($k \in \mathcal{K}$, $\boldsymbol{n} \in \mathcal{N}$) represents the time-average joint probability that a class k customer is being served, the numbers of waiting customers of class i is equal to n_i for each $i \in \mathcal{K}$, and the underlying Markov chain is in state j. Note also that $\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}$ ($k \in \mathcal{K}$) represents the arrival rate of class k customers. Owing to the relation $H = \lambda G$ between time and customer averages [HS80, Tak02], we have

$$[\boldsymbol{x}_k(\boldsymbol{n})]_j = \boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e} \cdot \mathrm{E}[T_{k,j}(\boldsymbol{n})], \qquad k \in \mathcal{K}, j \in \mathcal{M}, \boldsymbol{n} \in \mathcal{N},$$

from which and (3.94), (3.68) follows. □

3.P Proof of Lemma 3.12

By definition, $\tilde{\boldsymbol{D}}_k^{(m)}(\theta)$ ($k \in \mathcal{K}$, $m = 0, 1, \ldots$) satisfies

$$\begin{aligned}
\sum_{m=0}^{\infty} \tilde{\boldsymbol{D}}_k^{(m)}(\theta) z^m &= \frac{\boldsymbol{D}_k - \boldsymbol{D}_k^*(\theta - \theta z)}{\theta - \theta z} \\
&= \frac{1}{\theta} \sum_{m=0}^{\infty} \left[\boldsymbol{D}_k - \sum_{i=0}^{m} \boldsymbol{D}_k^{(i)}(\theta)\right] z^m, \qquad |z| < 1, \qquad (3.95)
\end{aligned}$$

where we used $\boldsymbol{D}_k^*(\theta - \theta z) = \sum_{m=0}^{\infty} \boldsymbol{D}_k^{(m)}(\theta) z^m$. (3.70) now follows from (3.95). □

4 M/G/1-Type Markov Processes with Reducible Generators for Busy Periods

4.1 Introduction

We consider a bivariate Markov process $\{(U(t),S(t)); t \geq 0\}$, where $U(t)$ and $S(t)$ are referred to as the level and the phase, respectively, at time t. $U(t)$ $(t \geq 0)$ takes values in $[0,\infty)$ and $S(t)$ $(t \geq 0)$ takes values in a finite set $\mathscr{M} = \{1,2,\dots,M\}$. $\{U(t); t \geq 0\}$ either decreases at rate one or has upward jump discontinuities, so that $\{U(t); t \geq 0\}$ is skip-free to the left. We assume that when $(U(t-),S(t-)) = (x,i)$ $(x > 0,\ i \in \mathscr{M})$, an upward jump (possibly with size zero) occurs at a rate $\sigma^{[i]}$ $(\sigma^{[i]} > 0)$ and the phase $S(t)$ becomes j $(j \in \mathscr{M})$ with probability $p^{[i,j]}$. On the other hand, when $(U(t-),S(t-)) = (0,i)$ $(i \in \mathscr{M})$, an upward jump occurs with probability one and the phase $S(t)$ becomes j $(j \in \mathscr{M})$ with probability $\overline{p}^{[i,j]}$. Note here that for $i \in \mathscr{M}$,

$$\sum_{j\in\mathscr{M}} p^{[i,j]} = 1, \qquad \sum_{j\in\mathscr{M}} \overline{p}^{[i,j]} = 1.$$

When $U(t) > 0$ (resp. $U(t) = 0$), the sizes of upward jumps with phase transitions from $S(t-) = i$ to $S(t) = j$ are i.i.d. according to a general distribution function $B^{[i,j]}(x)$ $(x \geq 0)$ (resp. $\overline{B}^{[i,j]}(x)$ $(x \geq 0)$). To avoid trivialities, we assume $B^{[i,i]}(0) = 0$ $(i \in \mathscr{M})$ and $\overline{B}^{[i,j]}(0) = 0$ $(i,j \in \mathscr{M})$.

We introduce $M \times M$ matrices $\boldsymbol{C}$, $\boldsymbol{D}(x)$ $(x \geq 0)$, and $\overline{\boldsymbol{B}}(x)$ $(x \geq 0)$ to deal with this Markov process.

$$\begin{aligned}
[\boldsymbol{C}]_{i,j} &= \begin{cases} -\sigma^{[i]}, & i = j, \\ \sigma^{[i]} p^{[i,j]} B^{[i,j]}(0), & i \neq j, \end{cases} \\
[\boldsymbol{D}(0)]_{i,j} &= 0, \qquad [\boldsymbol{D}(x)]_{i,j} = \sigma^{[i]} p^{[i,j]} B^{[i,j]}(x), \quad x > 0, \\
[\overline{\boldsymbol{B}}(x)]_{i,j} &= \overline{p}^{[i,j]} \overline{B}^{[i,j]}(x).
\end{aligned}$$

We define $\boldsymbol{D}^*(s)$ (Re(s) > 0) and $\overline{\boldsymbol{B}}^*(s)$ (Re(s) > 0) as the LSTs of $\boldsymbol{D}(x)$ and $\overline{\boldsymbol{B}}(x)$, respectively.

$$\boldsymbol{D}^*(s) = \int_0^\infty \exp[-sx]d\boldsymbol{D}(x), \quad \overline{\boldsymbol{B}}^*(s) = \int_0^\infty \exp[-sx]d\overline{\boldsymbol{B}}(x).$$

Further we define $M \times M$ matrices $\boldsymbol{D}$ and $\overline{\boldsymbol{B}}$ as

$$\boldsymbol{D} = \lim_{x\to\infty} \boldsymbol{D}(x) = \lim_{s\to 0+} \boldsymbol{D}^*(s), \quad \overline{\boldsymbol{B}} = \lim_{x\to\infty} \overline{\boldsymbol{B}}(x) = \lim_{s\to 0+} \overline{\boldsymbol{B}}^*(s).$$

By definition, $\boldsymbol{C}+\boldsymbol{D}$ represents the infinitesimal generator of a continuous-time Markov chain defined on finite state-space $\mathscr{M}$. Also, $\overline{\boldsymbol{B}}$ represents the transition probability matrix of a discrete-time Markov chain defined on finite state-space $\mathscr{M}$. Therefore, $\boldsymbol{C}+\boldsymbol{D}$ and $\overline{\boldsymbol{B}}$ satisfy

$$(\boldsymbol{C}+\boldsymbol{D})\boldsymbol{e} = \boldsymbol{0}, \qquad \overline{\boldsymbol{B}}\boldsymbol{e} = \boldsymbol{e},$$

respectively, where $\boldsymbol{e}$ denotes an $M \times 1$ vector whose elements are all equal to one.

As mentioned in Chapter 1, this Markov process was first introduced in [Tak96] as a continuous analog of Markov chains of the M/G/1 type [Neu89]. We thus refer to this Markov process as the M/G/1-type Markov process. In [Tak96], the M/G/1-type Markov process is regarded as a generalization of the workload process in the queueing model with customer arrivals of MAP, and the LST of the stationary distribution is derived under the assumption that $\boldsymbol{C}+\boldsymbol{D}$ is *irreducible*. This assumption is appropriate when we consider the stationary behavior of the ordinary MAP/G/1 queues, because it is equivalent to assume that the underlying Markov chain governing the arrival process is irreducible. However, the irreducibility of $\boldsymbol{C}+\boldsymbol{D}$ is *not* necessary for the irreducibility of $\{(U(t),S(t));t \geq 0\}$. This assumption is thus too strong and restricts its applicability to queueing models.

In this chapter, we assume that an $M \times M$ infinitesimal generator

$$\boldsymbol{C}+\boldsymbol{D}+\overline{\boldsymbol{B}}-\boldsymbol{I}$$

is irreducible, where $\boldsymbol{I}$ denotes an $M \times M$ unit matrix. It is readily verified that $\{(U(t),S(t));t \geq 0\}$ is irreducible if and only if $\boldsymbol{C}+\boldsymbol{D}+\overline{\boldsymbol{B}}-\boldsymbol{I}$ is irreducible. Therefore, even when $\boldsymbol{C}+\boldsymbol{D}$ is reducible, $\{(U(t),S(t));t \geq 0\}$ is irreducible if all states in $\mathscr{M}$ can be reached from each other with transitions governed by $\boldsymbol{C}+\boldsymbol{D}$ and $\overline{\boldsymbol{B}}$. With this extension, the M/G/1-type Markov process become applicable to a wider class of queueing models including our fundamental model of the first kind introduced in Chapter 1. Note that for discrete-time M/G/1 type Markov chains, analytical results for the case corresponding to reducible $\boldsymbol{C}+\boldsymbol{D}$ is found in [Neu89, Section 3.5]. To the best of our knowledge, however, a continuous analog of such results have not been reported in the literature.

The rest of this chapter is organized as follows. In Section 4.2, we explain the application of the M/G/1-type Markov process to the analysis of queueing models. We show through some examples that its applicability is extended allowing $\boldsymbol{C}+\boldsymbol{D}$ to be reducible. In Section 4.3 we briefly review known results for the M/G/1-type Markov process with irreducible $\boldsymbol{C}+\boldsymbol{D}$ [Tak96]. In Section 4.4, we first show that results in [Tak96] are not applicable directly to reducible $\boldsymbol{C}+\boldsymbol{D}$, and then derive a formula applicable to the reducible case. In addition, we provide a recursion to compute the moments of the stationary distribution, and consider an efficient computational procedure of a fundamental matrix for reducible $\boldsymbol{C}+\boldsymbol{D}$. Finally, we conclude this chapter in Section 4.5.

4.2 Applications of the M/G/1-type Markov process to queueing models

In this section, we shortly explain applications of the M/G/1-type Markov process to the analysis of queueing models. We first make an explanation about queueing models formulated to be the M/G/1-type Markov process with irreducible $\boldsymbol{C}+\boldsymbol{D}$. Next, we show that our fundamental model of the first kind is formulated as the M/G/1-type Markov process with reducible $\boldsymbol{C}+\boldsymbol{D}$. We also present some other examples of queueing models that are formulated as this extended version of M/G/1-type Markov process.

The M/G/1-type Markov process was first introduced as a continuous analog of the M/G/1 type Markov chain. As mentioned in Section 1.2.1, the embedded queue length process at the departure instants in the MAP/G/1 queue can be described by the M/G/1-type Markov chain. On the other hand, the censored workload process in the MAP/G/1 queue obtained by observing only busy periods can be described by the M/G/1-type Markov process. Specifically, the censored workload process in the MAP/G/1 queue characterized by a MAP $(\boldsymbol{C}_{\mathrm{MAP}},\boldsymbol{D}_{\mathrm{MAP}})$ and a service time distribution $H(x)$ $(x \geq 0)$ corresponds to the M/G/1-type Markov process with

$$\boldsymbol{C}=\boldsymbol{C}_{\mathrm{MAP}}, \quad \boldsymbol{D}(x)=H(x)\boldsymbol{D}_{\mathrm{MAP}}, \quad \overline{\boldsymbol{B}}(x)=H(x)(-\boldsymbol{C}_{\mathrm{MAP}})^{-1}\boldsymbol{D}_{\mathrm{MAP}}.$$

As mentioned in Section 1.2.2, analysis of the workload process is important when we consider the multi-class FCFS MAP/G/1 queue. Consider a multi-class MAP/G/1 queue characterized by $(\boldsymbol{C}_{\mathrm{MAP}},\boldsymbol{D}_{\mathrm{MAP},k}(x))$ $(k \in \mathscr{K},\ x \geq 0)$, where $\mathscr{K} = \{1,2,\ldots,K\}$ denotes the set of customer classes. For this model, the censored workload process obtained by observing only busy periods is described by the M/G/1-type Markov process with

$$\boldsymbol{C}=\boldsymbol{C}_{\mathrm{MAP}}, \quad \boldsymbol{D}(x)=\sum_{k\in\mathscr{K}}\boldsymbol{D}_{\mathrm{MAP},k}(x), \quad \overline{\boldsymbol{B}}(x)=(-\boldsymbol{C}_{\mathrm{MAP}})^{-1}\sum_{k\in\mathscr{K}}\boldsymbol{D}_{\mathrm{MAP},k}(x). \tag{4.1}$$

In the analysis of the ordinary single-class and multi-class MAP/G/1 queues, it is usually assumed that the underlying Markov chain is irreducible because the existence of transient states has no effect on performance measures of the queues in steady state. In accordance with this convention, the analytical results for the M/G/1-type Markov process reported in [Tak96] are derived under an assumption that $\boldsymbol{C}+\boldsymbol{D}$ is irreducible.

As shown in examples below, however, by allowing $\boldsymbol{C}+\boldsymbol{D}$ to be reducible, the M/G/1-type Markov process become applicable to a wider class of queueing models including our fundamental model of the first kind.

Example 4.1. *Consider a multi-class MAP/G/1 queue with a transient underlying Markov chain, whose state gets reset when the system becomes empty. The censored workload process obtained by observing only busy periods is formulated as an M/G/1-type Markov process with*

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{T}} & \boldsymbol{C}_{\mathrm{T,N}} \\ \boldsymbol{O} & \boldsymbol{C}_{\mathrm{N}} \end{pmatrix}, \quad \boldsymbol{D}(x) = \begin{pmatrix} \boldsymbol{D}_{\mathrm{T}}(x) & \boldsymbol{D}_{\mathrm{T,N}}(x) \\ \boldsymbol{O} & \boldsymbol{D}_{\mathrm{N}}(x) \end{pmatrix}, \quad \overline{\boldsymbol{B}}(x) = \begin{pmatrix} \overline{\boldsymbol{B}}_{\mathrm{T,T}}(x) & \boldsymbol{O} \\ \overline{\boldsymbol{B}}_{\mathrm{N,T}}(x) & \boldsymbol{O} \end{pmatrix},$$

where "T" and "N" represent "transient" and "normal", respectively.

Example 4.2. *Consider the multi-class MAP/G/1 queue with working vacations. This model can be regarded as a modified version of Example 4.1, where the processing rate is given by $\gamma > 0$ during the transient periods. By means of the change of time scale, the censored workload process of this model can be converted to an M/G/1-type Markov process with [Tak05]*

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{T}}/\gamma & \boldsymbol{C}_{\mathrm{T,N}}/\gamma \\ \boldsymbol{O} & \boldsymbol{C}_{\mathrm{N}} \end{pmatrix}, \quad \boldsymbol{D}(x) = \begin{pmatrix} \boldsymbol{D}_{\mathrm{T}}(x)/\gamma & \boldsymbol{D}_{\mathrm{T,N}}(x)/\gamma \\ \boldsymbol{O} & \boldsymbol{D}_{\mathrm{N}}(x) \end{pmatrix}, \quad \overline{\boldsymbol{B}}(x) = \begin{pmatrix} \overline{\boldsymbol{B}}_{\mathrm{T,T}}(x) & \boldsymbol{O} \\ \overline{\boldsymbol{B}}_{\mathrm{N,T}}(x) & \boldsymbol{O} \end{pmatrix}.$$

Therefore, our fundamental model of the first kind can be dealt with as a special case of this Markov process with reducible $\boldsymbol{C}+\boldsymbol{D}$. Note here that the multi-class M/G/1 queue with exponential working vacations analyzed in Chapter 2 is a very special case of the queuing model in Example 4.2. Note also that the multi-class MAP/G/1 queue with disasters analyzed in Chapter 3 corresponds to a censored process of the model in Example 4.1, obtained by observing the system only in transient periods.

In addition, there are also some other queueing models formulated as the M/G/1-type Markov process with reducible $\boldsymbol{C}+\boldsymbol{D}$, which are of independent interest.

Example 4.3. *Consider a MAP/G/1 queue with two types of busy periods $\{1,2\}$, where the customer arrival process is governed by $(\boldsymbol{C}^{(i)}_{\mathrm{MAP}}, \boldsymbol{D}^{(i)}_{\mathrm{MAP}}(x))$ during busy periods of type i ($i = 1,2$). Transitions of busy-period type occur only when the system* *is empty. The censored workload process obtained by observing only busy periods is formulated as an M/G/1-type Markov process with*

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{MAP}}^{(1)} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{C}_{\mathrm{MAP}}^{(2)} \end{pmatrix}, \quad \boldsymbol{D}(x) = \begin{pmatrix} \boldsymbol{D}_{\mathrm{MAP}}^{(1)}(x) & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{D}_{\mathrm{MAP}}^{(2)}(x) \end{pmatrix}, \quad \overline{\boldsymbol{B}}(x) = \begin{pmatrix} \overline{\boldsymbol{B}}^{(11)}(x) & \overline{\boldsymbol{B}}^{(12)}(x) \\ \overline{\boldsymbol{B}}^{(21)}(x) & \overline{\boldsymbol{B}}^{(22)}(x) \end{pmatrix}.$$

An example of a queueing system with two (or more) types of busy periods is a host machine in a distributed server system with dedicated task assignment policy [Bal03]. Each host is dedicated to either "short" or "long" jobs during a busy period so that variability of job sizes to be processed at each host becomes low. Furthermore, when a host becomes idle, its role may be changed to the other one, which improves the utilization of the system.

Example 4.4. *Consider a MAP/G/1 queue with multiple vacations and exhaustive service discipline [LMN90]. For queueing models with vacations, lengths of vacations are usually assumed to be i.i.d. random variables. Using the M/G/1-type Markov process with reducible $\boldsymbol{C}+\boldsymbol{D}$, we can describe a MAP/G/1 queue with semi-Markovian vacation times, where a sequence of vacation lengths forms a semi-Markov process. For example, consider a 2-state semi-Markov process $\{S_{\mathrm{V}}(t);\ t \geq 0\}$, where $S_{\mathrm{V}}(t)$ takes value in $\{1,2\}$. Let $V^{[i,j]}(x)$ ($x \geq 0,\ i,j = 1,2$) denote the joint probability that a state transition from state i to state j occurs when the sojourn time in state i is elapsed, and the sojourn time in state i is not greater than x. The workload process in a MAP/G/1 queue with vacations whose lengths are governed by this semi-Markov process is then represented by the M/G/1-type Markov process with*

$$\begin{aligned} \boldsymbol{C} &= \begin{pmatrix} \boldsymbol{C}_{\mathrm{MAP}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{C}_{\mathrm{MAP}} \end{pmatrix}, \quad \boldsymbol{D}(x) = \begin{pmatrix} \boldsymbol{D}_{\mathrm{MAP}}(x) & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{D}_{\mathrm{MAP}}(x) \end{pmatrix}, \\ \overline{\boldsymbol{B}}(x) &= \begin{pmatrix} V^{[1,1]}(x)\boldsymbol{I}_{\mathrm{MAP}} & V^{[1,2]}(x)\boldsymbol{I}_{\mathrm{MAP}} \\ V^{[2,1]}(x)\boldsymbol{I}_{\mathrm{MAP}} & V^{[2,2]}(x)\boldsymbol{I}_{\mathrm{MAP}} \end{pmatrix}, \end{aligned}$$

where $\boldsymbol{I}_{\mathrm{MAP}}$ denotes a unit matrix with the same size as $\boldsymbol{C}_{\mathrm{MAP}}$. Note that in this case, vacations can be regarded as service times of virtual customers who arrive immediately after the system becomes empty, so that this M/G/1-type Markov process represents the original workload process in the exhaustive-service MAP/G/1 vacation queue with semi-Markovian vacation times.

In Section 4.4, we develop analytical methods for the M/G/1-type Markov processes with reducible $\boldsymbol{C}+\boldsymbol{D}$. The results in Section 4.4 enable us to obtain performance measures in varieties of queueing models including those described in the examples above.

4.3 Known results for irreducible $\boldsymbol{C}+\boldsymbol{D}$ [Tak96]

In this section, we review known results in [Tak96], assuming that $\boldsymbol{C}+\boldsymbol{D}$ is irreducible. Owing to this assumption, $\boldsymbol{C}+\boldsymbol{D}$ has its invariant probability vector $\boldsymbol{\pi}$, which is uniquely determined by

$$\boldsymbol{\pi}(\boldsymbol{C}+\boldsymbol{D})=\boldsymbol{0}, \quad \boldsymbol{\pi}\boldsymbol{e}=1.$$

Let $\boldsymbol{\beta}$ and $\overline{\boldsymbol{\beta}}$ denote $M \times 1$ vectors given by

$$\boldsymbol{\beta}=\int_0^\infty x d\boldsymbol{D}(x)\boldsymbol{e}, \qquad \overline{\boldsymbol{\beta}}=\int_0^\infty x d\overline{\boldsymbol{B}}(x)\boldsymbol{e}. \tag{4.2}$$

Throughout this section, we assume that

$$\overline{\boldsymbol{\beta}}<\infty, \qquad \boldsymbol{\pi}\boldsymbol{\beta}<1,$$

which ensures the irreducible Markov process $\{(U(t),S(t));\ t \ge 0\}$ being positive recurrent [Tak96, Theorem 1]. Let $\boldsymbol{u}(x)$ $(x>0)$ denote a $1 \times M$ vector whose j-th $(j \in \mathscr{M})$ element represents the joint probability that the level is not greater than x and the phase is equal to j in steady state and we define $\boldsymbol{u}^*(s)$ $(\mathrm{Re}(s)>0)$ as the LST of $\boldsymbol{u}(x)$.

$$\begin{aligned}[\boldsymbol{u}(x)]_j &= \lim_{t\to\infty}\Pr(U(t)\le x, S(t)=j), \quad j \in \mathscr{M},\\ \boldsymbol{u}^*(s) &= \int_0^\infty \exp[-sx]d\boldsymbol{u}(x).\end{aligned}$$

We can derive the following lemma from the balance equation for steady state.

Lemma 4.1. *(Theorem 2 in [Tak96])* $\boldsymbol{u}^*(s)$ *(*$\mathrm{Re}(s)>0$*) satisfies*

$$\boldsymbol{u}^*(s)[s\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(s)]=\acute{\boldsymbol{u}}(0)[\boldsymbol{I}-\overline{\boldsymbol{B}}^*(s)], \qquad \mathrm{Re}(s)>0, \tag{4.3}$$

where $\acute{\boldsymbol{u}}(0)$ *denotes the right derivative of* $\boldsymbol{u}(x)$ *at* $x=0$.

$$\acute{\boldsymbol{u}}(0)=\lim_{x\to 0+}\frac{\boldsymbol{u}(x)-\boldsymbol{u}(0)}{x}.$$

Let c denote the reciprocal of the mean recurrence time of the set of states $\{(0,i);\ i \in \mathscr{M}\}$. Further let $\boldsymbol{\eta}^{\mathrm{E}}$ denote the stationary probability vector of the phase just before the level becomes 0. $\acute{\boldsymbol{u}}(0)$ is then given by

$$\acute{\boldsymbol{u}}(0)=c\boldsymbol{\eta}^{\mathrm{E}}. \tag{4.4}$$

In order to determine c and $\boldsymbol{\eta}^{\mathrm{E}}$, we consider the first passage time to level 0. Let T^{E} denote the first passage time to level 0 after time 0.

$$T^{\mathrm{E}} = \begin{cases} 0, & U(0) = 0, \\ \inf\{t; U(t) = 0, t > 0\}, & \text{otherwise.} \end{cases}$$

We define $\boldsymbol{P}(t \mid x)$ ($t \geq 0$, $x \geq 0$) as an $M \times M$ matrix whose (i,j)th element ($i,j \in \mathscr{M}$) represents the joint probability that the first passage time is not greater than t and the phase is equal to j at the end of the first passage time, given that the level is equal to x and the phase is equal to i at time 0.

$$[\boldsymbol{P}(t \mid x)]_{i,j} = \Pr(T^{\mathrm{E}} \leq t, S(T^{\mathrm{E}}-) = j \mid U(0) = x, S(0) = i).$$

Let $\boldsymbol{P}^*(s \mid x)$ ($\mathrm{Re}(s) > 0$, $x \geq 0$) denote the LST of $\boldsymbol{P}(t \mid x)$ with respect to t.

$$\boldsymbol{P}^*(s \mid x) = \int_{t=0}^{\infty} \exp[-st] d\boldsymbol{P}(t \mid x).$$

Using

$$\boldsymbol{P}^*(s \mid x+y) = \boldsymbol{P}^*(s \mid x)\boldsymbol{P}^*(s \mid y), \qquad x \geq 0, y \geq 0,$$

[TH94] shows that $\boldsymbol{P}^*(s \mid x)$ ($x \geq 0$) is given by

$$\boldsymbol{P}^*(s \mid x) = \exp[\boldsymbol{Q}^*(s)x], \tag{4.5}$$

where $\boldsymbol{Q}^*(s)$ ($\mathrm{Re}(s) > 0$) denotes an $M \times M$ matrix that satisfies

$$\boldsymbol{Q}^*(s) = -s\boldsymbol{I} + \boldsymbol{C} + \int_0^{\infty} d\boldsymbol{D}(y) \exp[\boldsymbol{Q}^*(s)y]. \tag{4.6}$$

Let $\boldsymbol{P}(x)$ ($x \geq 0$) denote an $M \times M$ transition probability matrix whose (i,j)th element ($i,j \in \mathscr{M}$) is given by

$$[\boldsymbol{P}(x)]_{i,j} = \Pr(S(T^{\mathrm{E}}-) = j \mid U(0) = x, S(0) = i).$$

By definition, we have

$$\boldsymbol{P}(x) = \lim_{s \to 0+} \boldsymbol{P}^*(s \mid x) = \exp[\boldsymbol{Q}x], \quad x \geq 0, \tag{4.7}$$

where

$$\boldsymbol{Q} = \lim_{s \to 0+} \boldsymbol{Q}^*(s).$$

Because of (4.6), $\boldsymbol{Q}$ satisfies

$$\boldsymbol{Q} = \boldsymbol{C} + \int_0^{\infty} d\boldsymbol{D}(y) \exp[\boldsymbol{Q}y]. \tag{4.8}$$

Remark 4.1. *As shown in [TH94], $\boldsymbol{Q}$ is given by the limit $\lim_{n\to\infty}\boldsymbol{Q}^{(n)}$ of an elementwise increasing sequence of matrices $\{\boldsymbol{Q}^{(n)}\}_{n=0,1,\ldots}$ given by the following recursion.*

$$\boldsymbol{Q}^{(0)} = \boldsymbol{C}, \qquad \boldsymbol{Q}^{(n)} = \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(y)\exp\left[\boldsymbol{Q}^{(n-1)}y\right], \quad n = 1,2,\ldots. \tag{4.9}$$

Because the integral on the right-side of this equation can be computed with uniformization [Tij94, Page 154], we can numerically obtain $\boldsymbol{Q} = \lim_{n\to\infty}\boldsymbol{Q}^{(n)}$ with an adequate stopping criterion. More specifically, for a given allowable error $\epsilon > 0$, we may stop the iteration at n^* satisfying $\max_{i\in\mathcal{M}}\left|[\boldsymbol{Q}^{(n^*)}\boldsymbol{e}]_i\right| < \epsilon$.

$\boldsymbol{Q}$ is known to be an infinitesimal generator of a Markov chain on $\mathcal{M}$, and it is irreducible if $\boldsymbol{C}+\boldsymbol{D}$ is irreducible [Tak02, TH94]. Therefore, because of the assumption of the irreducible $\boldsymbol{C}+\boldsymbol{D}$, $\boldsymbol{Q}$ has its invariant probability vector $\boldsymbol{\kappa}$, which is uniquely determined by

$$\boldsymbol{\kappa}\boldsymbol{Q} = \boldsymbol{0}, \quad \boldsymbol{\kappa}\boldsymbol{e} = 1. \tag{4.10}$$

We define $\boldsymbol{f}(x)$ ($x \ge 0$) as an $M \times 1$ vector whose i-th ($i \in \mathcal{M}$) element represents the mean first passage time to level 0, given that the level is equal to x and the phase is equal to i at time 0.

$$[\boldsymbol{f}(x)]_i = \mathrm{E}[T^{\mathrm{E}} \mid U(0) = x, S(0) = i].$$

Noting (4.5) and (4.6), we obtain $\boldsymbol{f}(x)$ through a straightforward calculation.

$$\begin{aligned}
\boldsymbol{f}(x) &= (-1)\cdot \lim_{s\to 0+}\frac{\partial}{\partial s}\boldsymbol{P}^*(s \mid x)\boldsymbol{e} \\
&= \left(\sum_{n=1}^{\infty}\frac{x^n\boldsymbol{Q}^{n-1}}{n!}\right)\left((-1)\cdot\lim_{s\to 0+}\frac{\partial}{\partial s}\boldsymbol{Q}^*(s)\boldsymbol{e}\right) &\quad (4.11)\\
&= [x\boldsymbol{e}\boldsymbol{\kappa} - \exp[\boldsymbol{Q}x] + \boldsymbol{I}][(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}]^{-1}\boldsymbol{e}, &\quad (4.12)
\end{aligned}$$

because

$$\begin{aligned}
\left(\sum_{n=1}^{\infty}\frac{x^n\boldsymbol{Q}^{n-1}}{n!}\right) &= [x\boldsymbol{e}\boldsymbol{\kappa} - \exp[\boldsymbol{Q}x] + \boldsymbol{I}](\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q})^{-1}, &\quad (4.13)\\
(-1)\cdot\lim_{s\to 0+}\frac{\partial}{\partial s}\boldsymbol{Q}^*(s)\boldsymbol{e} &= (\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q})[(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}]^{-1}\boldsymbol{e}. &\quad (4.14)
\end{aligned}$$

It is known that both of $\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q}$ and $(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}$ are non-singular when $\boldsymbol{C}+\boldsymbol{D}$ is irreducible.

c and $\boldsymbol{\eta}^{\mathrm{E}}$ on the right-hand side of (4.4) is then given by the following lemma.

Lemma 4.2. *(Theorem 3 in [Tak96]) $\boldsymbol{\eta}^{\mathrm{E}}$ is uniquely determined by*

$$\boldsymbol{\eta}^{\mathrm{E}}\int_0^\infty d\overline{\boldsymbol{B}}(x)\exp[\boldsymbol{Q}x] = \boldsymbol{\eta}^{\mathrm{E}}, \quad \boldsymbol{\eta}^{\mathrm{E}}\boldsymbol{e} = 1, \tag{4.15}$$

and c is given by

$$c = \frac{1}{\boldsymbol{\eta}^{\mathrm{E}} \int_0^\infty d\overline{\boldsymbol{B}}(x)\boldsymbol{f}(x)} = \frac{1}{\boldsymbol{\eta}^{\mathrm{E}}(\overline{\boldsymbol{\beta}}\boldsymbol{\kappa} + \overline{\boldsymbol{B}} - \boldsymbol{I})[(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}]^{-1}\boldsymbol{e}}. \tag{4.16}$$

Before closing this section, we derive an alternative formula for $\boldsymbol{u}^*(s)$, which is similar to that given in [Tak02]. We define $M \times M$ matrices $\boldsymbol{R}^*(s)$ (Re$(s) > 0$) and $\overline{\boldsymbol{R}}^*(s)$ (Re$(s) > 0$) as

$$\begin{aligned} \boldsymbol{R}^*(s) &= \int_0^\infty \exp[-sx]dx \int_x^\infty d\boldsymbol{D}(y)\exp[\boldsymbol{Q}(y-x)], \\ \overline{\boldsymbol{R}}^*(s) &= \int_0^\infty \exp[-sx]dx \int_x^\infty d\overline{\boldsymbol{B}}(y)\exp\big[\boldsymbol{Q}(y-x)\big]. \end{aligned}$$

By definition, $\boldsymbol{R}^*(s)$ and $\overline{\boldsymbol{R}}^*(s)$ satisfy

$$\begin{aligned} [\boldsymbol{I} - \boldsymbol{R}^*(s)](s\boldsymbol{I} + \boldsymbol{Q}) &= s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s), \ \mathrm{Re}(s) > 0, & (4.17) \\ \overline{\boldsymbol{R}}^*(s)(s\boldsymbol{I} + \boldsymbol{Q}) &= \int_0^\infty d\overline{\boldsymbol{B}}(y)\exp[\boldsymbol{Q}y] - \overline{\boldsymbol{B}}^*(s), \ \mathrm{Re}(s) > 0. & (4.18) \end{aligned}$$

It follows from (4.4), (4.15), and (4.18) that

$$\acute{\boldsymbol{u}}(0)\overline{\boldsymbol{R}}^*(s)(s\boldsymbol{I} + \boldsymbol{Q}) = \acute{\boldsymbol{u}}(0)[\boldsymbol{I} - \overline{\boldsymbol{B}}^*(s)], \qquad \mathrm{Re}(s) > 0.$$

With (4.17), (4.3) is then rewritten to be

$$\boldsymbol{u}^*(s)[\boldsymbol{I} - \boldsymbol{R}^*(s)](s\boldsymbol{I} + \boldsymbol{Q}) = \acute{\boldsymbol{u}}(0)\overline{\boldsymbol{R}}^*(s)(s\boldsymbol{I} + \boldsymbol{Q}), \qquad \mathrm{Re}(s) > 0. \tag{4.19}$$

In the same way as in Section 3.4, it can be shown that (4.19) implies

$$\boldsymbol{u}^*(s)[\boldsymbol{I} - \boldsymbol{R}^*(s)] = \acute{\boldsymbol{u}}(0)\overline{\boldsymbol{R}}^*(s), \qquad \mathrm{Re}(s) > 0.$$

We thus obtain the following theorem.

Theorem 4.1. $\boldsymbol{u}^*(s)$ *is given by*

$$\boldsymbol{u}^*(s) = \acute{\boldsymbol{u}}(0)\overline{\boldsymbol{R}}^*(s)[\boldsymbol{I} - \boldsymbol{R}^*(s)]^{-1}, \qquad \mathrm{Re}(s) > 0. \tag{4.20}$$

Remark 4.2. *[Tak02] shows that* $\boldsymbol{I} - \boldsymbol{R}^*(s)$ *(*Re$(s) > 0$*) is non-singular when* $\boldsymbol{C} + \boldsymbol{D}$ *is irreducible.*

4.4 Results for reducible $\boldsymbol{C}+\boldsymbol{D}$

In this section, we generalize the results in Section 4.3 to the case of reducible $\boldsymbol{C}+\boldsymbol{D}$. More specifically, we assume that the infinitesimal generator $\boldsymbol{C}+\boldsymbol{D}$ is reducible and it has H closed irreducible classes of states. We define $\mathscr{H}=\{1,2,\dots,H\}$ as the set of such irreducible classes. $\boldsymbol{C}$ and $\boldsymbol{D}$ are then written in the following form.

$$\boldsymbol{C}=\begin{pmatrix}\boldsymbol{C}_{\mathrm{T}} & \boldsymbol{C}_{\mathrm{T},1} & \boldsymbol{C}_{\mathrm{T},2} & \cdots & \boldsymbol{C}_{\mathrm{T},H}\\ \boldsymbol{O} & \boldsymbol{C}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{C}_2 & \cdots & \boldsymbol{O}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{C}_H\end{pmatrix},\qquad \boldsymbol{D}=\begin{pmatrix}\boldsymbol{D}_{\mathrm{T}} & \boldsymbol{D}_{\mathrm{T},1} & \boldsymbol{D}_{\mathrm{T},2} & \cdots & \boldsymbol{D}_{\mathrm{T},H}\\ \boldsymbol{O} & \boldsymbol{D}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{D}_2 & \cdots & \boldsymbol{O}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{D}_H\end{pmatrix}, \tag{4.21}$$

where $\boldsymbol{C}_{\mathrm{T}}$ denotes an $M_{\mathrm{T}}\times M_{\mathrm{T}}$ defective infinitesimal generator, $\boldsymbol{C}_h$ $(h\in H)$ denotes an $M_h\times M_h$ defective infinitesimal generator, and $\boldsymbol{C}_{\mathrm{T},h}$ $(h\in H)$ denotes an $M_{\mathrm{T}}\times M_h$ transition rate matrix. Also, $\boldsymbol{D}_{\mathrm{T}}$ denotes an $M_{\mathrm{T}}\times M_{\mathrm{T}}$ transition rate matrix, $\boldsymbol{D}_h$ $(h\in H)$ denotes an $M_h\times M_h$ transition rate matrix, and $\boldsymbol{D}_{\mathrm{T},h}$ $(h\in H)$ denotes an $M_{\mathrm{T}}\times M_h$ transition rate matrix. Because $\boldsymbol{C}_h+\boldsymbol{D}_h$ $(h\in H)$ represents an irreducible infinitesimal generator, it has its invariant probability vector $\boldsymbol{\pi}_h$, which is uniquely determined by

$$\boldsymbol{\pi}_h(\boldsymbol{C}_h+\boldsymbol{D}_h)=\boldsymbol{0},\quad \boldsymbol{\pi}_h\boldsymbol{e}_h=1,$$

where $\boldsymbol{e}_h$ $(h\in\mathscr{H})$ denotes an $M_h\times 1$ vector whose elements are all equal to one.

Throughout this chapter, for any $M\times M$ block upper-triangular matrix similar to $\boldsymbol{C}$ and $\boldsymbol{D}$ in (4.21), we denote the $(0,0)$th block by the subscript "T", the $(0,h)$th block $(h\in\mathscr{H})$ by the subscript "T,h", and the (h,h)th block $(h\in\mathscr{H})$ by the subscript "h". We define $M_{\mathrm{T}}\times 1$ vector $\boldsymbol{\beta}_{\mathrm{T}}$ and $M_h\times 1$ vector $\boldsymbol{\beta}_h$ $(h\in\mathscr{H})$ as

$$\boldsymbol{\beta}_h=\int_0^\infty xd\boldsymbol{D}_h(x)\boldsymbol{e}_h,\quad \boldsymbol{\beta}_{\mathrm{T}}=\int_0^\infty xd\boldsymbol{D}_{\mathrm{T}}(x)\boldsymbol{e}_{\mathrm{T}}+\sum_{h\in\mathscr{H}}\int_0^\infty xd\boldsymbol{D}_{\mathrm{T},h}(x)\boldsymbol{e}_h,$$

respectively, where $\boldsymbol{e}_{\mathrm{T}}$ denotes an $M_{\mathrm{T}}\times 1$ vector whose elements are all equal to one (cf. (4.2)). We assume that an $M\times M$ infinitesimal generator $\boldsymbol{C}+\boldsymbol{D}+\overline{\boldsymbol{B}}-\boldsymbol{I}$ is irreducible, which is a necessary and sufficient condition for $\{(U(t),S(t));t\ge 0\}$ to be irreducible as noted in Section 4.1. We also assume that

$$\overline{\boldsymbol{\beta}}<\infty,\quad \boldsymbol{\beta}_{\mathrm{T}}<\infty,\quad \boldsymbol{\pi}_h\boldsymbol{\beta}_h<1,\ h\in\mathscr{H}. \tag{4.22}$$

With Theorem 1 in [Tak96], it is easy to see that $\{(U(t),S(t));t\ge 0\}$ is positive recurrent if and only if (4.22) holds.

As mentioned in Section 4.3, the assumption of the irreducible $\boldsymbol{C}+\boldsymbol{D}$ is a sufficient condition for the followings to hold:

(i) The matrix $\boldsymbol{Q}$ is irreducible, so that it has the invariant probability vector $\boldsymbol{\kappa}$ which is uniquely determined by (4.10).

(ii) Both $\boldsymbol{e}\boldsymbol{\kappa}-\boldsymbol{Q}$ and $(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa}-\boldsymbol{C}-\boldsymbol{D}$ are non-singular, and therefore $\boldsymbol{f}(x)$ $(x \ge 0)$ is given by (4.12).

(iii) $\boldsymbol{I}-\boldsymbol{R}^*(s)$ on the right-hand side of (4.20) is non-singular for $\mathrm{Re}(s)>0$.

Note that these are the only things in the discussion of Section 4.3, which are related to the irreducibility of $\boldsymbol{C}+\boldsymbol{D}$.

We can prove that (iii) still holds for reducible $\boldsymbol{C}+\boldsymbol{D}$. We provide an outline of its proof in Appendix 4.A. As shown below, on the other hand, neither of (i) and (ii) above is valid when $\boldsymbol{C}+\boldsymbol{D}$ is reducible with more than one irreducible classes of states (i.e., $H \ge 2$).

Noting that $\boldsymbol{Q}$ is given by the limit of the sequence of matrices $\{\boldsymbol{Q}^{(n)}\}_{n=0,1,\dots}$ defined as (4.9), it is easy to see that $\boldsymbol{Q}$ takes the form

$$\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{Q}_{\mathrm{T}} & \boldsymbol{Q}_{\mathrm{T},1} & \boldsymbol{Q}_{\mathrm{T},2} & \cdots & \boldsymbol{Q}_{\mathrm{T},H}\\ \boldsymbol{O} & \boldsymbol{Q}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{Q}_2 & \cdots & \boldsymbol{O}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{Q}_H\end{pmatrix}, \tag{4.23}$$

where $\boldsymbol{Q}_{\mathrm{T}}$ denotes a defective infinitesimal generator, $\boldsymbol{Q}_{\mathrm{T},h}$ $(h \in \mathscr{H})$ denotes a transition rate matrix, and $\boldsymbol{Q}_h$ $(h \in \mathscr{H})$ denotes an irreducible infinitesimal generator. $\boldsymbol{Q}$ is thus no longer irreducible. Furthermore, when $H \ge 2$, there are infinitely many invariant probability vectors of $\boldsymbol{Q}$, which are given by linear combinations of the invariant probability vectors of $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, ..., and $\boldsymbol{Q}_H$. The following lemma shows that $\boldsymbol{e}\boldsymbol{\kappa}-\boldsymbol{Q}$ and $(\boldsymbol{e}-\boldsymbol{\beta})\boldsymbol{\kappa}-\boldsymbol{C}-\boldsymbol{D}$ are no longer non-singular for any invariant probability vector $\boldsymbol{\kappa}$ of $\boldsymbol{Q}$ if $H \ge 2$.

Lemma 4.3. *Consider an $M \times M$ reducible infinitesimal generator $\boldsymbol{Y}$ with H closed irreducible classes of states.*

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}_{\mathrm{T}} & \boldsymbol{Y}_{\mathrm{T},1} & \boldsymbol{Y}_{\mathrm{T},2} & \cdots & \boldsymbol{Y}_{\mathrm{T},H}\\ \boldsymbol{O} & \boldsymbol{Y}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{Y}_2 & \cdots & \boldsymbol{O}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{Y}_H\end{pmatrix}.$$

Let $\boldsymbol{\gamma}_h$ $(h \in \mathscr{H})$ denote the invariant probability vector of $\boldsymbol{Y}_h$.

$$\boldsymbol{\gamma}_h\boldsymbol{Y}_h=\boldsymbol{0}, \qquad \boldsymbol{\gamma}_h\boldsymbol{e}_h=1, \quad h \in \mathscr{H}.$$

If $H \ge 2$, $\boldsymbol{v}\boldsymbol{\alpha} - \boldsymbol{Y}$ is singular for any $1 \times M$ real vector $\boldsymbol{\alpha}$ and any $M \times 1$ real vector $\boldsymbol{v}$ satisfying

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_{\mathrm{T}} \\ \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_H \end{pmatrix}, \qquad \boldsymbol{\gamma}_h \boldsymbol{v}_h \neq 0 \quad \text{for some } h \in \mathscr{H}, \tag{4.24}$$

where $\boldsymbol{v}_{\mathrm{T}}$ and $\boldsymbol{v}_h$ ($h \in \mathscr{H}$) denote $M_{\mathrm{T}} \times 1$ and $M_h \times 1$ vectors, respectively.

We prove Lemma 4.3 in Appendix 4.B.

When $H \ge 2$, we can verify that $\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q}$ (resp. $(\boldsymbol{e} - \boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}$) is singular for any invariant probability vector $\boldsymbol{\kappa}$ of $\boldsymbol{Q}$, by letting $\boldsymbol{\alpha} = \boldsymbol{\kappa}$, $\boldsymbol{v} = \boldsymbol{e}$ (resp. $\boldsymbol{v} = \boldsymbol{e} - \boldsymbol{\beta}$), and $\boldsymbol{Y} = \boldsymbol{Q}$ (resp. $\boldsymbol{Y} = \boldsymbol{C} + \boldsymbol{D}$) in Lemma 4.3. The formulas (4.12) and (4.16) in Section 4.3 is thus not applicable to reducible $\boldsymbol{C} + \boldsymbol{D}$ with more than one irreducible classes of states.

Remark 4.3. *If $\boldsymbol{C} + \boldsymbol{D}$ has transient states and only one irreducible class of states, i.e., $H = 1$, $\boldsymbol{Q}$ has the unique invariant probability vector $\boldsymbol{\kappa}$ even though it is reducible. In this case, we can prove that both of $\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q}$ and $(\boldsymbol{e} - \boldsymbol{\beta})\boldsymbol{\kappa} - \boldsymbol{C} - \boldsymbol{D}$ are non-singular.*

Remark 4.4. *Although analytical results for the $M/G/1$-type Markov chain corresponding to the case of reducible $\boldsymbol{C} + \boldsymbol{D}$ is obtained in [Neu89, Section 3.5], it considers only the case of $H = 1$ with transient states. As shown for the continuous version, however, the case of $H \ge 2$ is essentially different from that of $H = 1$.*

The rest of this section consists of three subsections. In Section 4.4.1, we consider the LST of the stationary distribution $\boldsymbol{u}^*(s)$ ($\mathrm{Re}(s) > 0$), and derive a formula applicable to reducible $\boldsymbol{C} + \boldsymbol{D}$. In Section 4.4.2, we provide an efficient computational procedure of reducible $\boldsymbol{Q}$ with the block structure (4.23). Finally, in Section 4.4.3, we consider the moments of the stationary distribution. We show that some modification from the irreducible case is necessary to obtain the moments.

4.4.1 LST of stationary distribution

In this subsection, we derive a formula for the LST of the stationary distribution $\boldsymbol{u}^*(s)$ ($\mathrm{Re}(s) > 0$) applicable to reducible $\boldsymbol{C} + \boldsymbol{D}$. Note that (4.3) and (4.20) are still valid for reducible $\boldsymbol{C} + \boldsymbol{D}$. The difference from the irreducible case is that $\acute{\boldsymbol{u}}(0)$ cannot be obtained from Lemma 4.2 because (4.12) and (4.16) does not hold for reducible $\boldsymbol{C} + \boldsymbol{D}$ with $H \ge 2$ as shown above.

Therefore, we first derive a formula for the mean first passage time $\boldsymbol{f}(x)$ ($x \ge 0$) applicable to reducible $\boldsymbol{C} + \boldsymbol{D}$ with the general structure (4.21). Let $\boldsymbol{\kappa}_h$ ($h \in$

$\mathscr{H}$) denote the invariant probability vector of $\boldsymbol{Q}_h$ (see (4.23)), which is uniquely determined by

$$\boldsymbol{\kappa}_h\boldsymbol{Q}_h = \boldsymbol{0}, \quad \boldsymbol{\kappa}_h\boldsymbol{e}_h = 1. \tag{4.25}$$

We then define $M \times M$ matrix $\check{\boldsymbol{Q}}$ as

$$\check{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{O} & \check{\boldsymbol{Q}}_{\mathrm{T},1} & \check{\boldsymbol{Q}}_{\mathrm{T},2} & \cdots & \check{\boldsymbol{Q}}_{\mathrm{T},H} \\ \boldsymbol{O} & \boldsymbol{e}_1\boldsymbol{\kappa}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{e}_2\boldsymbol{\kappa}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{e}_H\boldsymbol{\kappa}_H \end{pmatrix}, \tag{4.26}$$

where

$$\check{\boldsymbol{Q}}_{\mathrm{T},h} = (-\boldsymbol{Q}_{\mathrm{T}})^{-1}\boldsymbol{Q}_{\mathrm{T},h}\boldsymbol{e}_h\boldsymbol{\kappa}_h, \quad h \in \mathscr{H}.$$

Lemma 4.4. *$\boldsymbol{f}(x)$ ($x \geq 0$) is given by*

$$\boldsymbol{f}(x) = [\boldsymbol{I} - \exp[\boldsymbol{Q}x] + x\check{\boldsymbol{Q}}](\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D})^{-1}\boldsymbol{e}, \tag{4.27}$$

where $\boldsymbol{\Delta}$ is defined as

$$\begin{aligned} \boldsymbol{\Delta} &= \check{\boldsymbol{Q}} - \int_0^\infty x d\boldsymbol{D}(x)\check{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{\Delta}_{\mathrm{T},1} & \boldsymbol{\Delta}_{\mathrm{T},2} & \cdots & \boldsymbol{\Delta}_{\mathrm{T},H} \\ \boldsymbol{O} & (\boldsymbol{e}_1 - \boldsymbol{\beta}_1)\boldsymbol{\kappa}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & (\boldsymbol{e}_2 - \boldsymbol{\beta}_2)\boldsymbol{\kappa}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \ddots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & (\boldsymbol{e}_H - \boldsymbol{\beta}_H)\boldsymbol{\kappa}_H \end{pmatrix}, \\ \boldsymbol{\Delta}_{\mathrm{T},h} &= \check{\boldsymbol{Q}}_{\mathrm{T},h} - \int_0^\infty x d\boldsymbol{D}_{\mathrm{T}}(x)\check{\boldsymbol{Q}}_{\mathrm{T},h} - \int_0^\infty x d\boldsymbol{D}_{\mathrm{T},h}(x)\boldsymbol{e}_h\boldsymbol{\kappa}_h, \quad h \in \mathscr{H}. \end{aligned}$$

Remark 4.5. $(\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D})^{-1}$ *is given by*

$$(\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D})^{-1} = \begin{pmatrix} [-(\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}})]^{-1} & \boldsymbol{J}_{\mathrm{T},1} & \boldsymbol{J}_{\mathrm{T},2} & \cdots & \boldsymbol{J}_{\mathrm{T},H} \\ \boldsymbol{O} & \widehat{\boldsymbol{\Delta}}_1^{-1} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \widehat{\boldsymbol{\Delta}}_2^{-1} & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \widehat{\boldsymbol{\Delta}}_H^{-1} \end{pmatrix}, \tag{4.28}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_h &= (\boldsymbol{e}_h - \boldsymbol{\beta}_h)\boldsymbol{\kappa}_h - \boldsymbol{C}_h - \boldsymbol{D}_h, \quad h \in \mathscr{H}, \\ \boldsymbol{J}_{\mathrm{T},h} &= (-1)\cdot[-(\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}})]^{-1}(\boldsymbol{\Delta}_{\mathrm{T},h} - \boldsymbol{C}_{\mathrm{T},h} - \boldsymbol{D}_{\mathrm{T},h})\widehat{\boldsymbol{\Delta}}_h^{-1}, \quad h \in \mathscr{H}. \end{aligned}$$

Proof. Because we can prove Lemma 4.4 in almost the same way as the irreducible case in [TH94], we provide only an outline of the proof. By definition of $\check{\boldsymbol{Q}}$, it follows that

$$\boldsymbol{Q}\check{\boldsymbol{Q}} = \boldsymbol{O},$$

from which we obtain a similar result to (4.13).

$$\sum_{n=1}^{\infty} \frac{x^n}{n!} \boldsymbol{Q}^{n-1} = \left[\boldsymbol{I} - \exp[\boldsymbol{Q}x] + x\check{\boldsymbol{Q}}\right](\check{\boldsymbol{Q}} - \boldsymbol{Q})^{-1}.$$

Note here that the block upper-triangular matrix $\check{\boldsymbol{Q}} - \boldsymbol{Q}$ is non-singular because its diagonal block matrices $-\boldsymbol{Q}_{\mathrm{T}}$ and $\boldsymbol{e}_h \boldsymbol{\kappa}_h - \boldsymbol{Q}_h$ ($h \in \mathscr{H}$) are non-singular. Noting that $\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D}$ is non-singular by the same reasoning as the non-singularity of $\check{\boldsymbol{Q}} - \boldsymbol{Q}$, we also obtain

$$(-1) \cdot \lim_{s \to 0+} \frac{d}{ds} \boldsymbol{Q}^*(s)\boldsymbol{e} = (\check{\boldsymbol{Q}} - \boldsymbol{Q})(\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D})^{-1}\boldsymbol{e},$$

which corresponds to (4.14). We then obtain (4.27) from (4.11). □

We then obtain $\boldsymbol{u}^*(s)$ ($\mathrm{Re}(s) > 0$) for reducible $\boldsymbol{C}+\boldsymbol{D}$ using (4.3), (4.15), and (4.20).

Theorem 4.2. $\boldsymbol{u}^*(s)$ ($\mathrm{Re}(s) > 0$) *satisfies*

$$\boldsymbol{u}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = c\boldsymbol{\eta}^{\mathrm{E}}[\boldsymbol{I} - \overline{\boldsymbol{B}}^*(s)], \qquad \mathrm{Re}(s) > 0, \tag{4.29}$$

and it is given by

$$\boldsymbol{u}^*(s) = c\boldsymbol{\eta}^{\mathrm{E}}\overline{\boldsymbol{R}}^*(s)[\boldsymbol{I} - \boldsymbol{R}^*(s)]^{-1}, \qquad \mathrm{Re}(s) > 0,$$

where $\boldsymbol{\eta}^{\mathrm{E}}$ *denotes a* $1 \times M$ *probability vector which is uniquely determined by*

$$\boldsymbol{\eta}^{\mathrm{E}} = \boldsymbol{\eta}^{\mathrm{E}} \int_0^{\infty} d\overline{\boldsymbol{B}}(x) \exp[\boldsymbol{Q}x], \quad \boldsymbol{\eta}^{\mathrm{E}}\boldsymbol{e} = 1, \tag{4.30}$$

and c is given by

$$\begin{aligned} c &= \frac{1}{\boldsymbol{\eta}^{\mathrm{E}}(\overline{\boldsymbol{\Delta}} + \overline{\boldsymbol{B}} - \boldsymbol{I})(\boldsymbol{\Delta} - \boldsymbol{C} - \boldsymbol{D})^{-1}\boldsymbol{e}}, \\ \overline{\boldsymbol{\Delta}} &= \int_0^{\infty} x d\overline{\boldsymbol{B}}(x)\check{\boldsymbol{Q}}. \end{aligned} \tag{4.31}$$

Remark 4.6. *Let* $\boldsymbol{\Phi}$ *denote an* $M \times M$ *matrix given by*

$$\boldsymbol{\Phi} = \int_0^{\infty} d\overline{\boldsymbol{B}}(x) \exp[\boldsymbol{Q}x].$$

Since $\{(U(t), S(t)); t \geq 0\}$ *is irreducible and positive recurrent,* $\boldsymbol{\Phi}$ *represents an irreducible transition probability matrix. Therefore,* $\boldsymbol{\Phi}$ *has its invariant probability vector, so that* $\boldsymbol{\eta}^{\mathrm{E}}$ *is uniquely determined by (4.30)*

Remark 4.7. *When we apply Theorem 4.2 to the case that $\boldsymbol{C}+\boldsymbol{D}$ has no transient states, it is necessary to rewrite (4.26) and (4.28) as*

$$\check{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{e}_1\boldsymbol{\kappa}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{e}_2\boldsymbol{\kappa}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{e}_H\boldsymbol{\kappa}_H \end{pmatrix}, \quad (\boldsymbol{\Delta}-\boldsymbol{C}-\boldsymbol{D})^{-1} = \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_1^{-1} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \widehat{\boldsymbol{\Delta}}_2^{-1} & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \widehat{\boldsymbol{\Delta}}_H^{-1} \end{pmatrix}.$$

4.4.2 Computation of reducible $\boldsymbol{Q}$

In this subsection, we consider the computation of $\boldsymbol{Q}$ for reducible $\boldsymbol{C}+\boldsymbol{D}$. As mentioned in Remark 4.1, $\boldsymbol{Q}$ can be computed based on the recursion (4.9). However, a straightforward implementation of the computational procedure given in Remark 4.1 is not efficient for reducible $\boldsymbol{C}+\boldsymbol{D}$ because $\boldsymbol{Q}^{(n)}$ is a sparse block matrix and the number of iterations required is determined by the most slowly convergent sequence among non-zero blocks. Therefore, we can avoid unnecessary calculations by computing $\boldsymbol{Q}$ blockwise as follows.

It is readily to see from (4.9) that $\boldsymbol{Q}_h$ ($h \in \mathscr{H}$) is given by the limit $\lim_{n\to\infty}\boldsymbol{Q}_h^{(n)}$ of the elementwise increasing sequence of matrices $\boldsymbol{Q}_h^{(n)}$ ($n = 0,1,\ldots$) defined as

$$\boldsymbol{Q}_h^{(0)} = \boldsymbol{C}_h, \qquad \boldsymbol{Q}_h^{(n)} = \boldsymbol{C}_h + \int_0^\infty d\boldsymbol{D}_h(y)\exp[\boldsymbol{Q}_h^{(n-1)}y], \quad n = 1,2,\ldots. \tag{4.32}$$

Because $\boldsymbol{Q}_h$ ($h \in \mathscr{H}$) represents an infinitesimal generator and $\boldsymbol{Q}_h\boldsymbol{e}_h = \boldsymbol{0}$ holds, it can be computed individually with an adequate stopping criterion in the same way as the computation of $\boldsymbol{Q}$ with (4.9).

Similarly, $\boldsymbol{Q}_{\mathrm{T}}$ is given by the limit $\lim_{n\to\infty}\boldsymbol{Q}_{\mathrm{T}}^{(n)}$ of the elementwise increasing sequence of matrices $\boldsymbol{Q}_{\mathrm{T}}^{(n)}$ ($n = 0,1,\ldots$) defined as

$$\boldsymbol{Q}_{\mathrm{T}}^{(0)} = \boldsymbol{C}_{\mathrm{T}}, \qquad \boldsymbol{Q}_{\mathrm{T}}^{(n)} = \boldsymbol{C}_{\mathrm{T}} + \int_0^\infty d\boldsymbol{D}_{\mathrm{T}}(y)\exp[\boldsymbol{Q}_{\mathrm{T}}^{(n-1)}y], \quad n = 1,2,\ldots.$$

However, because $\boldsymbol{Q}_{\mathrm{T}}$ represents a defective infinitesimal generator, the stopping criterion of $\boldsymbol{Q}_{\mathrm{T}}^{(n)}$ is not clear. We thus need to compute $\boldsymbol{Q}_{\mathrm{T}}$ along with $\boldsymbol{Q}_{\mathrm{T},h}$ ($h \in \mathscr{H}$). Let $\widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n)}$ ($n = 0,1,\ldots$) denote a sequence of matrices defined as

$$\begin{aligned} \widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(0)} &= \boldsymbol{C}_{\mathrm{T},h}, \\ \widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n)} &= \boldsymbol{C}_{\mathrm{T},h} + \int_0^\infty d\boldsymbol{D}_{\mathrm{T},h}(y)\exp[\boldsymbol{Q}_h y] \\ &\quad + \int_0^\infty d\boldsymbol{D}_{\mathrm{T}}(y)\int_0^y \exp[\boldsymbol{Q}_{\mathrm{T}}^{(n-1)}t]\widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n-1)}\exp[\boldsymbol{Q}_h(y-t)]dt, \quad n = 1,2,\ldots. \end{aligned} \tag{4.33}$$

Note that $\boldsymbol{Q}_{\mathrm{T}}^{(n)}$ has the same structure as $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$ in the multi-class MAP/G/1 queue with disasters discussed in Chapter 3 (see (3.15)). According to the probabilistic interpretation of $\boldsymbol{Q}_{\mathrm{N}}^{(n)}$, it can be verified that

$$\lim_{n\to\infty} \widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n)} = \boldsymbol{Q}_{\mathrm{T},h}, \qquad h \in \mathscr{H}.$$

Therefore, we first compute $\boldsymbol{Q}_h$ ($h \in \mathscr{H}$) with (4.32), and then we compute $\boldsymbol{Q}_{\mathrm{T}}$ and $\boldsymbol{Q}_{\mathrm{T},h}$ with an adequate stopping criterion. More specifically, for a given allowable error $\epsilon > 0$, we may stop the iteration at n^* satisfying

$$\max_{i\in\mathscr{M}_{\mathrm{T}}} \left| [\boldsymbol{Q}_{\mathrm{T}}^{(n^*)}\boldsymbol{e}_{\mathrm{T}} + \sum_{h\in\mathscr{H}} \widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n^*)}\boldsymbol{e}_h]_i \right| < \epsilon.$$

Remark 4.8. *The second integral on the right-hand side of (4.33) can be computed with uniformization as follows. Let θ denote the maximum absolute value of the diagonal elements of the matrix $\boldsymbol{C}$. We then have*

$$\int_0^\infty d\boldsymbol{D}_{\mathrm{T}}(y) \int_0^y \exp[\boldsymbol{Q}_{\mathrm{T}}^{(n)}t]\widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n)} \exp[\boldsymbol{Q}_h(y-t)]dt$$
$$= \sum_{m=0}^{\infty} \boldsymbol{D}_{\mathrm{T}}^{(m+1)}(\theta) \sum_{j=0}^{m} [\boldsymbol{I}_{\mathrm{T}} + \theta^{-1}\boldsymbol{Q}_{\mathrm{T}}^{(n)}]^{m-j}\theta^{-1}\widehat{\boldsymbol{Q}}_{\mathrm{T},h}^{(n)}[\boldsymbol{I}_h + \theta^{-1}\boldsymbol{Q}_h]^j, \quad (4.34)$$

where

$$\boldsymbol{D}_{\mathrm{T}}^{(m)}(\theta) = \int_0^\infty \exp[-\theta x]\frac{(\theta x)^m}{m!} d\boldsymbol{D}_{\mathrm{T}}(x).$$

The derivation of (4.34) is given in Appendix 4.C.

Remark 4.9. *If $\boldsymbol{C} + \boldsymbol{D}$ has no transient states, $\boldsymbol{Q}$ is given by*

$$\begin{pmatrix} \boldsymbol{Q}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{Q}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{Q}_H \end{pmatrix},$$

so that we only need to compute $\boldsymbol{Q}_h$ ($h \in \mathscr{H}$) based on (4.32).

4.4.3 Moments of the stationary distribution

In this subsection, we derive a recursive formula for the moments of the stationary distribution. For this purpose, we introduce some notations. We first rewrite (4.29) to be

$$\boldsymbol{u}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = -\boldsymbol{\eta}^*(s) \qquad \mathrm{Re}(s) > 0, \quad (4.35)$$

where $\boldsymbol{\eta}^*(s)$ is given by

$$\boldsymbol{\eta}^*(s) = c\boldsymbol{\eta}^{\mathrm{E}}[\overline{\boldsymbol{B}}^*(s) - \boldsymbol{I}].$$

Let $\boldsymbol{u}^{(m)}$ $(m=0,1,\ldots)$ and $\boldsymbol{\eta}^{(m)}$ $(m=0,1,\ldots)$ denote $1\times M$ vectors given by

$$\begin{aligned}
\boldsymbol{u}^{(0)} &= \lim_{s\to 0+}\boldsymbol{u}^*(s), \qquad \boldsymbol{u}^{(m)} = \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}\big[\boldsymbol{u}^*(s)\big], \quad m=1,2,\ldots,\\
\boldsymbol{\eta}^{(0)} &= \lim_{s\to 0+}\boldsymbol{\eta}^*(s), \qquad \boldsymbol{\eta}^{(m)} = \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}\big[\boldsymbol{\eta}^*(s)\big], \quad m=1,2,\ldots.
\end{aligned}$$

Note that the j-th $(j\in\mathscr{M})$ element of $\boldsymbol{u}^{(0)}$ represents the stationary probability that the phase is equal to j.

We develop a recursion to compute $\boldsymbol{u}^{(m)}$ $(m=0,1,\ldots)$ utilizing the fact that $\boldsymbol{C}$ and $\boldsymbol{D}^*(s)$ are sparse block matrices. We thus partition $\boldsymbol{u}^*(s)$, $\boldsymbol{\eta}^*(s)$, $\boldsymbol{u}^{(m)}$ $(m=0,1,\ldots)$, and $\boldsymbol{\eta}^{(m)}$ $(m=0,1,\ldots)$ as follows.

$$\begin{aligned}
\boldsymbol{u}^*(s) &= (\boldsymbol{u}_{\mathrm{T}}^*(s),\boldsymbol{u}_1^*(s),\boldsymbol{u}_2^*(s),\ldots,\boldsymbol{u}_H^*(s)), \qquad \boldsymbol{\eta}^*(s) = (\boldsymbol{\eta}_{\mathrm{T}}^*(s),\boldsymbol{\eta}_1^*(s),\boldsymbol{\eta}_2^*(s),\ldots,\boldsymbol{\eta}_H^*(s)),\\
\boldsymbol{u}^{(m)} &= (\boldsymbol{u}_{\mathrm{T}}^{(m)},\boldsymbol{u}_1^{(m)},\boldsymbol{u}_2^{(m)},\ldots,\boldsymbol{u}_H^{(m)}), \qquad \boldsymbol{\eta}^{(m)} = (\boldsymbol{\eta}_{\mathrm{T}}^{(m)},\boldsymbol{\eta}_1^{(m)},\boldsymbol{\eta}_2^{(m)},\ldots,\boldsymbol{\eta}_H^{(m)}),
\end{aligned}$$

where $\boldsymbol{u}_{\mathrm{T}}^*(s)$, $\boldsymbol{\eta}_{\mathrm{T}}^*(s)$, $\boldsymbol{u}_{\mathrm{T}}^{(m)}$, and $\boldsymbol{\eta}_{\mathrm{T}}^{(m)}$ denote $1\times M_{\mathrm{T}}$ vectors and $\boldsymbol{u}_h^*(s)$, $\boldsymbol{\eta}_h^*(s)$, $\boldsymbol{u}_h^{(m)}$, and $\boldsymbol{\eta}_h^{(m)}$ $(h\in\mathscr{H})$ denote $1\times M_h$ vectors.

Note that (4.35) is equivalent to

$$\boldsymbol{u}_{\mathrm{T}}^*(s)[s\boldsymbol{I}_{\mathrm{T}}+\boldsymbol{C}_{\mathrm{T}}+\boldsymbol{D}_{\mathrm{T}}^*(s)] = -\boldsymbol{\eta}_{\mathrm{T}}^*(s), \tag{4.36}$$

$$\boldsymbol{u}_h^*(s)[s\boldsymbol{I}_h+\boldsymbol{C}_h+\boldsymbol{D}_h^*(s)] = -\boldsymbol{\phi}_h^*(s), \quad h\in\mathscr{H}, \tag{4.37}$$

where

$$\boldsymbol{\phi}_h^*(s) = \boldsymbol{\eta}_h^*(s)+\boldsymbol{u}_{\mathrm{T}}^*(s)[\boldsymbol{C}_{\mathrm{T},h}+\boldsymbol{D}_{\mathrm{T},h}^*(s)], \quad h\in\mathscr{H}.$$

We define $\boldsymbol{\phi}_h^{(m)}$ $(h\in\mathscr{H},\ m=0,1,\ldots)$ as

$$\begin{aligned}
\boldsymbol{\phi}_h^{(0)} &= \lim_{s\to 0+}\boldsymbol{\phi}_h^*(s) = \boldsymbol{\eta}_h^{(0)}+\boldsymbol{u}_{\mathrm{T}}^{(0)}(\boldsymbol{C}_{\mathrm{T},h}+\boldsymbol{D}_{\mathrm{T},h}),\\
\boldsymbol{\phi}_h^{(m)} &= \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}[\boldsymbol{\phi}_h^*(s)] = \boldsymbol{\eta}_h^{(m)}+\boldsymbol{u}_{\mathrm{T}}^{(m)}\boldsymbol{C}_{\mathrm{T},h}+\sum_{l=0}^{m}\boldsymbol{u}_{\mathrm{T}}^{(l)}\boldsymbol{D}_{\mathrm{T},h}^{(m-l)}, \quad m=1,2,\ldots,
\end{aligned}$$

where for $h\in\mathscr{H}$,

$$\boldsymbol{D}_{\mathrm{T},h}^{(0)} = \boldsymbol{D}_{\mathrm{T},h}, \qquad \boldsymbol{D}_{\mathrm{T},h}^{(m)} = \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}\big[\boldsymbol{D}_{\mathrm{T},h}^*(s)\big], \quad m=1,2,\ldots.$$

We also define $\boldsymbol{D}_{\mathrm{T}}^{(m)}$ $(m=0,1,\ldots)$ and $\boldsymbol{D}_h^{(m)}$ $(h\in\mathscr{H},\ m=0,1,\ldots)$ as

$$\begin{aligned}
\boldsymbol{D}_{\mathrm{T}}^{(0)} &= \boldsymbol{D}_{\mathrm{T}}, \qquad \boldsymbol{D}_{\mathrm{T}}^{(m)} = \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}\big[\boldsymbol{D}_{\mathrm{T}}^*(s)\big], \quad m=1,2,\ldots,\\
\boldsymbol{D}_h^{(0)} &= \boldsymbol{D}_h, \qquad \boldsymbol{D}_h^{(m)} = \lim_{s\to 0+}\frac{(-1)^m}{m!}\cdot\frac{d^m}{ds^m}\big[\boldsymbol{D}_h^*(s)\big], \quad m=1,2,\ldots.
\end{aligned}$$

Theorem 4.3. *$\boldsymbol{u}^{(m)} = (\boldsymbol{u}_{\mathrm{T}}^{(m)}, \boldsymbol{u}_1^{(m)}, \boldsymbol{u}_2^{(m)}, \ldots, \boldsymbol{u}_H^{(m)})$ $(m = 0, 1, \ldots)$ is given recursively by*

$$\begin{aligned}
\boldsymbol{u}_{\mathrm{T}}^{(0)} &= \boldsymbol{\eta}_{\mathrm{T}}^{(0)}[-(\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}})]^{-1}, \\
\boldsymbol{u}_{\mathrm{T}}^{(m)} &= \left[\boldsymbol{\eta}_{\mathrm{T}}^{(m)} - \boldsymbol{u}_{\mathrm{T}}^{(m-1)} + \sum_{l=0}^{m-1} \boldsymbol{u}_{\mathrm{T}}^{(l)} \boldsymbol{D}_{\mathrm{T}}^{(m-l)}\right][-(\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}})]^{-1}, \quad m = 1, 2, \ldots,
\end{aligned}$$

and for $h \in \mathscr{H}$,

$$\begin{aligned}
\boldsymbol{u}_h^{(0)} \boldsymbol{e}_h &= \frac{1}{1 - \boldsymbol{\pi}_h \boldsymbol{\beta}_h} \left[\boldsymbol{\phi}_h^{(1)} \boldsymbol{e}_h + \boldsymbol{\phi}_h^{(0)} (\boldsymbol{e}_h \boldsymbol{\pi}_h - \boldsymbol{C}_h - \boldsymbol{D}_h)^{-1} \boldsymbol{\beta}_h\right], && (4.38) \\
\boldsymbol{u}_h^{(0)} &= \boldsymbol{u}_h^{(0)} \boldsymbol{e}_h \boldsymbol{\pi}_h + \boldsymbol{\phi}_h^{(0)} (\boldsymbol{e}_h \boldsymbol{\pi}_h - \boldsymbol{C}_h - \boldsymbol{D}_h)^{-1}, && (4.39) \\
\boldsymbol{\psi}_h^{(m)} &= \left(\sum_{l=0}^{m-1} \boldsymbol{u}_h^{(l)} \boldsymbol{D}_h^{(m-l)} - \boldsymbol{u}_h^{(m-1)} + \boldsymbol{\phi}_h^{(m)}\right) (\boldsymbol{e}_h \boldsymbol{\pi}_h - \boldsymbol{C}_h - \boldsymbol{D}_h)^{-1}, \\
&\qquad m = 1, 2, \ldots, && (4.40) \\
\boldsymbol{u}_h^{(m)} \boldsymbol{e}_h &= \frac{1}{1 - \boldsymbol{\pi}_h \boldsymbol{\beta}_h} \left[\sum_{l=0}^{m-1} \boldsymbol{u}_h^{(l)} \boldsymbol{D}_h^{(m+1-l)} \boldsymbol{e}_h + \boldsymbol{\phi}_h^{(m+1)} \boldsymbol{e}_h + \boldsymbol{\psi}_h^{(m)} \boldsymbol{\beta}_h\right], \\
&\qquad m = 1, 2, \ldots, && (4.41) \\
\boldsymbol{u}_h^{(m)} &= \boldsymbol{u}_h^{(m)} \boldsymbol{e}_h \boldsymbol{\pi}_h + \boldsymbol{\psi}_h^{(m)}, \quad m = 1, 2, \ldots. && (4.42)
\end{aligned}$$

Proof. We first consider $\boldsymbol{u}_{\mathrm{T}}^{(m)}$ $(m = 0, 1, \ldots)$. It follows from (4.36) that

$$\begin{aligned}
&\boldsymbol{u}_T^{(0)} (\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}}) = -\boldsymbol{\eta}_T^{(0)}, \\
&\boldsymbol{u}_T^{(m)} (\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}}) - \boldsymbol{u}_T^{(m-1)} + \sum_{l=0}^{m-1} \boldsymbol{u}_T^{(l)} \boldsymbol{D}_{\mathrm{T}}^{(m-l)} = -\boldsymbol{\eta}_T^{(m)}, \quad m = 1, 2, \ldots.
\end{aligned}$$

Since $\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}}$ is a defective infinitesimal generator, it is non-singular. We thus obtain the recursion for $\boldsymbol{u}_{\mathrm{T}}^{(m)}$ $(m = 0, 1, \ldots)$ from the above equations.

Next we consider $\boldsymbol{u}_h^{(m)}$ $(h \in \mathscr{H},\ m = 0, 1, \ldots)$ based on (4.37). Since $\boldsymbol{C}_h + \boldsymbol{D}_h$ $(h \in \mathscr{H})$ is an irreducible infinitesimal generator, the recursion for $\boldsymbol{u}_h^{(m)}$ $(h \in \mathscr{H},$ $m = 0, 1, \ldots)$ can be obtained by standard manipulations in matrix-analytic methods (e.g., [TH94]), and therefore we omit the proof. □

Remark 4.10. *By definition of $\boldsymbol{\phi}_h$ $(h \in \mathscr{H})$, we need to compute $\boldsymbol{u}_{\mathrm{T}}^{(m+1)}$ before computing $\boldsymbol{u}_h^{(m)}$ by (4.38)–(4.42). When $\boldsymbol{C} + \boldsymbol{D}$ has no transient states, on the other hand, $\boldsymbol{u}_h^{(m)}$ $(h \in \mathscr{H})$ can be immediately obtained from (4.38)–(4.42) noting*

$$\boldsymbol{\phi}_h^{(m)} = \boldsymbol{\eta}_h^{(m)}, \qquad m = 0, 1, \ldots.$$

4.5 Conclusion

We extended the matrix analytic methods for the continuous-time bivariate Markov process $\{(U(t), S(t)); t \geq 0\}$ introduced in [Tak96] to the case of reducible $\boldsymbol{C}+\boldsymbol{D}$. We first proved Lemma 4.3, which implies that some known results for the boundary vector $\acute{\boldsymbol{u}}(0)$ of the stationary distribution is not valid for reducible $\boldsymbol{C}+\boldsymbol{D}$, when there exist more than one irreducible classes of states. We then derived a formula for the LST of the stationary distribution applicable to the reducible $\boldsymbol{C}+\boldsymbol{D}$ in Section 4.4.1. Furthermore, we provided an efficient computational procedure of the fundamental matrix $\boldsymbol{Q}$ and the moments of the stationary distribution.

Recall that the Markov process considered in this chapter corresponds to the (censored) workload processes in MAP/G/1 queues with various features including our fundamental model of the first kind (see examples in Section 4.2). Based on the results in this chapter, we can obtain performance measures of the corresponding queueing model such as the waiting time distribution and the queue length distribution in a straightforward manner by following the discussion in [Tak96] and [Tak01] for an ordinary multi-class MAP/G/1 queue.

Appendices

4.A Non-singularity of $\boldsymbol{I}-\boldsymbol{R}^*(s)$ for reducible $\boldsymbol{C}+\boldsymbol{D}$

In this appendix we provide a brief proof that $\boldsymbol{I}-\boldsymbol{R}^*(s)$ $(\mathrm{Re}(s)>0)$ is non-singular for reducible $\boldsymbol{C}+\boldsymbol{D}$, as is the case of irreducible $\boldsymbol{C}+\boldsymbol{D}$. Note first that $\boldsymbol{R}^*(s)$ takes the form

$$\boldsymbol{R}^*(s) = \begin{pmatrix} \boldsymbol{R}^*_{\mathrm{T}}(s) & \boldsymbol{R}^*_{\mathrm{T},1}(s) & \boldsymbol{R}^*_{\mathrm{T},2}(s) & \cdots & \boldsymbol{R}^*_{\mathrm{T},H}(s) \\ \boldsymbol{O} & \boldsymbol{R}^*_1(s) & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{R}^*_2(s) & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{R}^*_H(s) \end{pmatrix},$$

and its diagonal block matrices are given by

$$\begin{aligned} \boldsymbol{R}^*_{\mathrm{T}}(s) &= \int_0^\infty \exp[-sx] \int_x^\infty d\boldsymbol{D}_{\mathrm{T}}(y) \exp[\boldsymbol{Q}_{\mathrm{T}}(y-x)], \\ \boldsymbol{R}^*_h(s) &= \int_0^\infty \exp[-sx] \int_x^\infty d\boldsymbol{D}_h(y) \exp[\boldsymbol{Q}_h(y-x)], \qquad h \in \mathscr{H}. \end{aligned}$$

Since $\boldsymbol{C}_h + \boldsymbol{D}_h$ $(h \in \mathscr{H})$ is irreducible, we can verify that $\boldsymbol{I}_h - \boldsymbol{R}^*_h(s)$ $(\mathrm{Re}(s)>0)$ is non-singular in the same way as in [Tak02]. Furthermore, noting that $\boldsymbol{C}_{\mathrm{T}} + \boldsymbol{D}_{\mathrm{T}}$ denotes a defective infinitesimal generator, we can prove that $\boldsymbol{I}_{\mathrm{T}} - \boldsymbol{R}^*_{\mathrm{T}}(s)$ $(\mathrm{Re}(s)>0)$ is non-singular in the same way as in Appendix 3.I. Therefore, $\boldsymbol{I}-\boldsymbol{R}^*(s)$ $(\mathrm{Re}(s)>0)$ is non-singular because its diagonal block matrices are all non-singular. □

4.B Proof of Lemma 4.3

We first consider the case that $\boldsymbol{Y}$ has two irreducible classes of states and no transient states, i.e.,

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{Y}_2 \end{pmatrix},$$

where $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ denote irreducible infinitesimal generators. (4.24) is then rewritten to be

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}, \qquad \boldsymbol{\gamma}_h \boldsymbol{v}_h \neq 0 \quad \text{for some } h \in \{1,2\}.$$

Without loss of generality, we assume that $\boldsymbol{\gamma}_2\boldsymbol{v}_2 \neq \boldsymbol{0}$. We then define a $1 \times M$ vector $\boldsymbol{y}$ as

$$\boldsymbol{y} = \left(\boldsymbol{\gamma}_1, \frac{-\boldsymbol{\gamma}_1\boldsymbol{v}_1}{\boldsymbol{\gamma}_2\boldsymbol{v}_2} \cdot \boldsymbol{\gamma}_2\right)$$

It then follows that

$$\boldsymbol{y}\boldsymbol{v} = \boldsymbol{\gamma}_1\boldsymbol{v}_1 - \frac{\boldsymbol{\gamma}_1\boldsymbol{v}_1}{\boldsymbol{\gamma}_2\boldsymbol{v}_2} \cdot \boldsymbol{\gamma}_2\boldsymbol{v}_2 = 0.$$

Therefore we have

$$\boldsymbol{y}(\boldsymbol{v}\boldsymbol{\alpha} - \boldsymbol{Y}) = \boldsymbol{0}.$$

Since $\boldsymbol{y} \neq \boldsymbol{0}$, $\boldsymbol{v}\boldsymbol{\alpha} - \boldsymbol{Y}$ is singular.

In the exactly same way, we can easily verify that $\boldsymbol{v}\boldsymbol{\alpha} - \boldsymbol{Y}$ is still singular for the general case that $\boldsymbol{Y}$ has transient states and more than two irreducible classes. □

4.C Derivation of (4.34)

With uniformization at rate θ, we have

$$\begin{aligned}
&\int_0^\infty d\boldsymbol{D}_\mathrm{T}(y) \int_0^y \exp[\boldsymbol{Q}_\mathrm{T}^{(n)} t] \boldsymbol{Q}_{\mathrm{T},h}^{(n)} \exp[\boldsymbol{Q}_h(y-t)] dt \\
&\quad = \int_0^\infty d\boldsymbol{D}_\mathrm{T}(y) \int_0^y \exp[-\theta t] \exp[\theta(\boldsymbol{I}_\mathrm{T} + \theta^{-1}\boldsymbol{Q}_\mathrm{T}^{(n)})t] \boldsymbol{Q}_{\mathrm{T},h}^{(n)} \\
&\qquad\qquad \cdot \exp[-\theta(y-t)] \exp[\theta(\boldsymbol{I}_h + \theta^{-1}\boldsymbol{Q}_h)(y-t)] dt \\
&\quad = \int_0^\infty \exp[-\theta y] d\boldsymbol{D}_\mathrm{T}(y) \int_0^y \sum_{i=0}^\infty \frac{\theta^i t^i}{i!} [\boldsymbol{I}_\mathrm{T} + \theta^{-1}\boldsymbol{Q}_\mathrm{T}^{(n)}]^i \boldsymbol{Q}_{\mathrm{T},h}^{(n)} \\
&\qquad\qquad \cdot \sum_{j=0}^\infty \frac{\theta^j (y-t)^j}{j!} [\boldsymbol{I}_h + \theta^{-1}\boldsymbol{Q}_h]^j dt \\
&\quad = \sum_{m=0}^\infty \sum_{j=0}^m \int_0^\infty \exp[-\theta y] d\boldsymbol{D}_\mathrm{T}(y) \int_0^y \frac{\theta^{m-j} t^{m-j}}{(m-j)!} \cdot \frac{\theta^j (y-t)^j}{j!} dt \\
&\qquad\qquad \cdot [\boldsymbol{I}_\mathrm{T} + \theta^{-1}\boldsymbol{Q}_\mathrm{T}^{(n)}]^{m-j} \boldsymbol{Q}_{\mathrm{T},h}^{(n)} [\boldsymbol{I}_h + \theta^{-1}\boldsymbol{Q}_h]^j,
\end{aligned}$$

where $m = i + j$. Furthermore, calculating the integral with respect to t using

$$\int_0^y t^{m-j}(y-t)^j dt = \frac{j!(m-j)!}{(m+1)!} \cdot y^{m+1},$$

we obtain

$$\int_0^\infty d\boldsymbol{D}_{\mathrm{T}}(y) \int_0^y \exp[\boldsymbol{Q}_{\mathrm{T}}^{(n)} t] \boldsymbol{Q}_{\mathrm{T},h}^{(n)} \exp\big[\boldsymbol{Q}_h (y-t)\big] dt$$
$$= \sum_{m=0}^{\infty} \sum_{j=0}^{m} \int_0^\infty \exp[-\theta y] \frac{(\theta y)^{m+1}}{(m+1)!} d\boldsymbol{D}_{\mathrm{T}}(y) \cdot [\boldsymbol{I}_{\mathrm{T}} + \theta^{-1} \boldsymbol{Q}_{\mathrm{T}}^{(n)}]^{m-j} \theta^{-1} \boldsymbol{Q}_{\mathrm{T},h}^{(n)} [\boldsymbol{I}_h + \theta^{-1} \boldsymbol{Q}_h]^j.$$

(4.34) now follows immediately.

5 Workload Distribution in the M/G/1+G Queue

5.1 Introduction

In this chapter, we consider a stationary M/G/1 queue with general impatient customers. We assume that customers arrive according to a Poisson process with rate λ ($\lambda > 0$) and service times of customers are i.i.d. according to a general distribution with the PDF $H(x)$ ($x \geq 0$). Customers are served on the FCFS basis, unless otherwise stated. Each customer has his/her own maximum allowable waiting time, which is referred to as the impatience time hereafter. If elapsed waiting times of customers reach their impatience times, they leave the system immediately without receiving their services. Note that once the service of a customer starts, this customer remains in the system until his/her service completion, even if the impatience time expires. We assume that impatience times of customers are i.i.d. according to a general distribution with the PDF $G(x)$ ($x \geq 0$). Customers may have no waiting time limit and impatience times of such customers are defined as infinity. Therefore $G(x)$ ($x \geq 0$) may be defective, i.e.,

$$\lim_{x \to \infty} G(x) = 1 - g_\infty,$$

where g_∞ ($0 \leq g_\infty < 1$) denotes the probability that a randomly chosen customer has no waiting time limit. Usually, this model is denoted by M/G/1+G, where the last symbol represents the impatience time distribution. Let $\mathrm{E}[H]$ denote the mean service time, and let ρ denote the traffic intensity.

$$\rho = \lambda \mathrm{E}[H]. \tag{5.1}$$

Throughout this dissertation (i.e., in Chapters 5,6, and 7), we assume

$$\rho g_\infty < 1, \tag{5.2}$$

which ensures the system being stable [BBH84]. In addition, we assume $H(0) = 0$ and $G(0) = 0$ for simplicity. These assumptions can be made without loss of generality because customers with zero service times or zero impatience times do not contribute to the workload in system.

Let $v(x)$ $(x > 0)$ denote the p.d.f. of the workload in system, which is defined as the workload that the server will process eventually. In other words, the workload in this model is regarded as the waiting time of an arriving customer without waiting time limit. Also let π_0 denote the stationary probability that the system is empty. The following results are known for the stationary workload distribution in the M/G/1+G queue [BBH84, Kov61]. With the level-crossing argument [BP77, Coh77], it is readily shown that the p.d.f. $v(x)$ $(x > 0)$ of the workload satisfies [BBH84, Kov61]

$$v(x) = \lambda\pi_0\overline{H}(x) + \lambda\int_{0+}^{x} v(y)\overline{G}(y)\overline{H}(x-y)dy, \tag{5.3}$$

where $\overline{H}(x)$ and $\overline{G}(x)$ $(x \geq 0)$ denote complementary PDFs of service times and impatience times, respectively.

$$\overline{H}(x) = 1 - H(x), \qquad \overline{G}(x) = 1 - G(x).$$

(5.3) is a Volterra integral equation of the second kind, whose formal series solution is given by

$$v(x) = \pi_0\sum_{n=0}^{\infty}\lambda^n\phi_n(x), \qquad x > 0, \tag{5.4}$$

where $\{\phi_n(x), x > 0\}_{n=0,1,\ldots}$ is a sequence of functions determined recursively by

$$\phi_0(x) = \lambda\overline{H}(x), \tag{5.5}$$

$$\phi_n(x) = \int_{0+}^{x}\phi_{n-1}(y)\overline{G}(y)\overline{H}(x-y)dy, \quad n = 1,2,\ldots. \tag{5.6}$$

Furthermore, using

$$1 - \pi_0 = \int_{0+}^{\infty} v(x)dx = \pi_0\sum_{n=0}^{\infty}\lambda^n\int_{0+}^{\infty}\phi_n(x)dx,$$

we obtain

$$\pi_0 = \left(1 + \sum_{n=0}^{\infty}\lambda^n\int_{0+}^{\infty}\phi_n(x)dx\right)^{-1}. \tag{5.7}$$

In this chapter, we present a new perspective on the formal series solution (5.4) of $v(x)$. Our analysis is based on an observation that the workload process in the M/G/1+G queue is identical sample path-wise to an LCFS-PR M/G/1 queue with workload-dependent loss. We derive the joint p.d.f. of residual service times of customers in the LCFS-PR system. With this result, $v(x)$ $(x > 0)$ is interpreted as the p.d.f. of a random sum of dependent random variables, which clarifies the probabilistic meanings of the formal series solution (5.4). As a by-product, our formulation provides a unified understanding of special cases of the M/G/1+G queue, where the expression $v(x)$ $(x > 0)$ is simplified dramatically owing to specific distributions of service times/impatience times.

The rest of this chapter is organized as follows. In Section 5.2, we consider an LCFS-PR M/G/1 queue with workload-dependent loss, whose workload process is equivalent to that of the M/G/1+G queue. In Section 5.3, we discuss the connection between results in Section 5.2 and the formal series solution (5.4) of $v(x)$ in the M/G/1+G queue. In Section 5.4, we discuss some special cases where $v(x)$ $(x > 0)$ takes much simpler form than the general M/G/1+G queue. Finally, we conclude this chapter in Section 5.5.

5.2 LCFS-PR M/G/1 queue with workload-dependent loss

In this section, we consider a stationary LCFS-PR M/G/1 queue with workload-dependent loss. Suppose an arriving customer finds x $(x \geq 0)$ amount of workload on arrival. This arriving customer is admitted to the system with probability $\overline{G}(x)$, and he/she is lost with probability $1-\overline{G}(x)$. Customers admitted to the system are served under the LCFS-PR service discipline, so that customers start their services immediately on arrival if they are admitted to the system. We assume that customers arrive according to a Poisson process with rate λ, and service times are assumed to be i.i.d. according to the PDF $H(x)$ $(x \geq 0)$ with mean $\mathrm{E}[H]$. We assume that (5.2) holds.

Because the LCFS-PR service discipline is work-conserving, the workload process in the above-mentioned model is identical to that in the second fundamental model described in Chapter 1, which again has the identical workload process as the FCFS M/G/1+G queue, as mentioned in Section 1.3.2.

Let L denote the number of customers in the stationary LCFS-PR system. Given $L = n$ $(n = 1,2,\ldots)$, let X_i $(i = 1,2,\ldots,n)$ denote the residual service time of the i-th oldest customer. We then define $F(n;\, x_1,x_2,\ldots,x_n)$ $(n = 1,2,\ldots,\, x_i \geq 0$ $(i = 1,2,\ldots,n))$ as

$$F(n;\, x_1,x_2,\ldots,x_n) = \Pr(L = n, X_1 \leq x_1, X_2 \leq x_2,\ldots, X_n \leq x_n).$$

We further define $f(n;\, x_1,x_2,\ldots,x_n)$ $(n = 1,2,\ldots,\, x_i > 0$ $(i = 1,2,\ldots,n))$ as

$$f(n;\, x_1,x_2,\ldots,x_n) = \frac{\partial^n F(n;\, x_1,x_2,\ldots,x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

Let p_n $(n = 0, 1, \ldots)$ denote $p_n = \Pr(L = n)$. Note that

$$p_0 = \pi_0. \tag{5.8}$$

We then define $f(x_1, x_2, \ldots, x_n \mid n)$ $(n = 1, 2, \ldots,\ x_i > 0\ (i = 1, 2, \ldots, n))$ as the conditional joint p.d.f. of residual service times given $L = n$.

$$f(x_1, x_2, \ldots, x_n \mid n) = \frac{f(n;\ x_1, x_2, \ldots, x_n)}{p_n}, \quad n = 1, 2, \ldots. \tag{5.9}$$

Let $\tilde{h}(x)$ $(x \geq 0)$ denote the p.d.f. of the equilibrium random variable for service times.

$$\tilde{h}(x) = \frac{\overline{H}(x)}{\mathrm{E}[H]}, \quad x \geq 0. \tag{5.10}$$

We define $\{\tilde{H}_n\}_{n=1,2,\ldots}$ as a sequence of i.i.d. random variables with p.d.f. $\tilde{h}(x)$.

Theorem 5.1. *In the stationary LCFS-PR system described above,*

(i) π_0 *is given by*

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} c_n\right)^{-1}, \tag{5.11}$$

where c_n *(*$n = 1, 2, \ldots$*) is defined as*

$$c_n = \begin{cases} \rho, & n = 1, \\ \rho^n \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}\left(\sum_{j=1}^{i} \tilde{H}_j\right)\right], & n = 2, 3, \ldots. \end{cases} \tag{5.12}$$

Furthermore $f(n;\ x_1, x_2, \ldots, x_n)$ *(*$n = 1, 2, \ldots,\ x_i > 0$ *(*$i = 1, 2, \ldots, n$*)) is given in terms of* π_0*:*

$$f(n;\ x_1, x_2, \ldots, x_n) = \pi_0 \rho^n \tilde{h}(x_1) \prod_{i=2}^{n} \overline{G}\left(\sum_{j=1}^{i-1} x_j\right) \tilde{h}(x_i). \tag{5.13}$$

(ii) p_n *(*$n = 1, 2, \ldots$*) is given by*

$$p_n = \pi_0 c_n, \quad n = 1, 2, \ldots. \tag{5.14}$$

(iii) $f(x_1, x_2, \ldots, x_n \mid n)$ *(*$n = 1, 2, \ldots,\ x_i > 0$ *(*$i = 1, 2, \ldots, n$*)) is given by*

$$f(x_1, x_2, \ldots, x_n \mid n) = \tilde{h}(x_1) \prod_{i=2}^{n} \overline{G}\left(\sum_{j=1}^{i-1} x_j\right) \tilde{h}(x_i) \bigg/ \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}\left(\sum_{j=1}^{i} \tilde{H}_i\right)\right], \tag{5.15}$$

which is independent of the arrival rate λ.

Remark 5.1. *When all customers have no waiting time limit, i.e.,* $\overline{G}(x) = 1$ *(*$x \geq 0$*), (5.13) is reduced to*

$$f(n;\ x_1, x_2, \dots, x_n) = (1-\rho)\rho^n \cdot \prod_{i=1}^{n} \tilde{h}(x_i),$$

which agrees with the classical result for the ordinary LCFS-PR $M/G/1$ *queue [Kel79, Theorem 3.10].*

Proof. We first prove (5.13) by induction. Let $\mathscr{A}_0$ denote a set of admitted customers who arrive when the system is empty. Note that every customer in $\mathscr{A}_0$ is served only when $L = 1$ because of the LCFS-PR service discipline. Let $T_1(x_1)$ ($x_1 \geq 0$) denote the mean length of time in which customers in $\mathscr{A}_0$ receive their service and $X_1 \leq x_1$.

$$T_1(x_1) = \int_0^{x_1} \overline{H}(t)dt. \tag{5.16}$$

Note that $T_1(x_1)$ is a customer-average quantity. Owing to the relation between time- and customer-averages [HS80], we obtain

$$F(1;\ x_1) = \lambda\pi_0 T_1(x_1).$$

Note here that the arrival rate of customers in $\mathscr{A}_0$ is given by $\lambda\pi_0$. Substituting (5.1), (5.10), and (5.16) into the above equation and taking the derivative with respect to x_1, we have (5.13) for $n = 1$.

We then assume that (5.13) holds for some $n = m$ ($m = 1, 2, \dots$). Let $\mathscr{A}_m$ denote a set of admitted customers who arrive when $L = m$. Note that every customer in $\mathscr{A}_m$ is served only when $L = m+1$. We define $P_{\text{admit}}(m)$ ($m = 1, 2, \dots$) as

$$P_{\text{admit}}(m) = \int_{x_1=0+}^{\infty} \int_{x_2=0+}^{\infty} \cdots \int_{x_m=0+}^{\infty} f(x_1, x_2, \dots, x_m \mid m)\overline{G}(\sum_{j=1}^{m} x_j)dx_m dx_{m-1} \cdots dx_1. \tag{5.17}$$

Owing to conditional PASTA [DR88], $P_{\text{admit}}(m)$ ($m = 1, 2, \dots$) is interpreted as the conditional probability that a randomly chosen customer is admitted to the system given that he/she arrives when $L = m$.

Let $\hat{f}(x_1, x_2, \dots, x_m \mid m)$ denote the joint p.d.f. of (X_1, X_2, ..., X_m) seen by customers in $\mathscr{A}_m$ on arrival. It then follows from conditional PASTA [DR88] and (5.9) that

$$\begin{aligned} \hat{f}(x_1, x_2, \dots, x_m \mid m) &= \frac{1}{P_{\text{admit}}(m)} \cdot f(x_1, x_2, \dots, x_m \mid m)\overline{G}(\sum_{j=1}^{m} x_j) \\ &= \frac{1}{p_m P_{\text{admit}}(m)} \cdot f(m;\ x_1, x_2, \dots, x_m)\overline{G}(\sum_{j=1}^{m} x_j). \end{aligned} \tag{5.18}$$

We define $T_{m+1}(x_1,x_2,\dots,x_{m+1})$ ($x_i \ge 0$ ($i = 1,2,\dots,m+1$)) as the mean length of time in which customers in $\mathscr{A}_m$ receive their service, $X_1 \le x_1$, $X_2 \le x_2$, ..., and $X_{m+1} \le x_{m+1}$.

$$\begin{aligned} &T_{m+1}(x_1,x_2,\dots,x_{m+1}) \\ &= \int_{y_1=0+}^{x_1}\int_{y_2=0+}^{x_2}\cdots\int_{y_m=0+}^{x_m}\hat{f}(y_1,y_2,\dots,y_m \mid m)dy_m dy_{m-1}\cdots dy_1 \int_{t=0}^{x_{m+1}}\overline{H}(t)dt. \quad (5.19)\end{aligned}$$

Owing to the relation between time- and customer-averages [HS80], we obtain

$$F(m+1;\ x_1,x_2,\dots,x_{m+1}) = \lambda p_m P_{\text{admit}}(m)T_{m+1}(x_1,x_2,\dots,x_{m+1}),$$

because the arrival rate of customers in $\mathscr{A}_m$ is given by $\lambda p_m P_{\text{admit}}(m)$ owing to PASTA. Substituting (5.1), (5.10), (5.18), and (5.19) into the above equation, taking the partial derivative with respect to $x_1,x_2,\dots,x_{m+1}$, we have

$$f(m+1;\ x_1,x_2,\dots,x_{m+1}) = f(m;\ x_1,x_2,\dots,x_m)\cdot\rho\overline{G}(\sum_{j=1}^{m}x_j)\tilde{h}(x_{m+1}). \quad (5.20)$$

Therefore, if (5.13) holds for some $n = m$, we can verify that it also holds for $n = m+1$. We thus obtain (5.13) for $n = 1,2,\dots$. Furthermore, by definition, we have

$$\begin{aligned} &1-\pi_0 \\ &= \sum_{n=1}^{\infty}\int_{x_1=0+}^{\infty}\int_{x_2=0+}^{\infty}\cdots\int_{x_n=0+}^{\infty} f(n;\ x_1,x_2,\dots,x_n)dx_n dx_{n-1}\cdots dx_1 \\ &= \pi_0\sum_{n=1}^{\infty}\int_{x_1=0+}^{\infty}\int_{x_2=0+}^{\infty}\cdots\int_{x_n=0+}^{\infty}\rho^n\tilde{h}(x_1)\left\{\prod_{i=2}^{n}\overline{G}(\sum_{j=1}^{i-1}x_j)\tilde{h}(x_i)\right\}dx_n dx_{n-1}\cdots dx_1 \\ &= \pi_0\sum_{n=1}^{\infty}c_n, \quad (5.21)\end{aligned}$$

from which (5.11) follows.

The remaining is to prove (ii) and (iii). Note that p_n ($n = 1,2,\dots$) is given by

$$p_n = \int_{x_1=0+}^{\infty}\int_{x_2=0+}^{\infty}\cdots\int_{x_n=0+}^{\infty} f(n;\ x_1,x_2,\dots,x_n)dx_n dx_{n-1}\cdots dx_1.$$

We then obtain (5.14) with the same manipulation as in (5.21). We further obtain (5.15) substituting (5.12), (5.13), and (5.14) into (5.9). □

Remark 5.2. *The recursion (5.20) is helpful in understanding the structure of the product-form solution (5.13) of the joint p.d.f. $f(n;\ x_1,x_2,\dots,x_n)$ ($n = 1,2,\dots$, $x_i > 0$ ($i = 1,2,\dots,n$)). Intuitively, $\tilde{h}(x_{m+1})$ on the right-hand side of (5.20) corresponds to the residual service time of the newest customer, and $\overline{G}(\sum_{j=1}^{m}x_j)$ corresponds to the probability of admittance at the arrival instant of this newest customer. Note that in the ordinary LCFS-PR $M/G/1$ queue, (5.20) is reduced to be (cf. Remark 5.1)*

$$f(m+1;\ x_1,x_2,\dots,x_{m+1}) = f(m;\ x_1,x_2,\dots,x_m)\cdot\rho\tilde{h}(x_{m+1}).$$

5.3 Stationary workload in the M/G/1+G queue

We now relate the result in Theorem 5.1 to the workload in the M/G/1+G queue described in Section 5.1. We define V as the workload in the stationary M/G/1+G queue. Recall that the total workload process in the LCFS-PR M/G/1 queue with workload-dependent loss is identical to the workload process in the M/G/1+G queue. Therefore the following corollary is immediate from Theorem 5.1.

Corollary 5.1. *Let L denote a non-negative, integer-valued random variable whose probability function p_n ($n = 0,1,\ldots$) is given by (5.8) and (5.14). The stationary workload V is then given by the sum of L dependent, non-negative random variables X_i's.*

$$V = X_1 + X_2 + \cdots + X_L. \tag{5.22}$$

Let V_n ($n = 1,2,\ldots$) denote a conditional workload given $L = n$ and let $v(x \mid n)$ ($x > 0$, $n = 1,2,\ldots$) denote the p.d.f. of V_n. We then have

$$v(x \mid 1) = f(x \mid 1), \tag{5.23}$$

$$v(x \mid n) = \int_{\mathscr{D}_+(x|n-1)} f(\boldsymbol{x}^{[n-1]}, x - \sum_{m=1}^{n-1} x_m \mid n) d\boldsymbol{x}^{[n-1]}, \quad n = 2,3,\ldots, \tag{5.24}$$

where $\boldsymbol{x}^{[n]}$ ($n = 1,2,\ldots$) denotes a $1 \times n$ vector given by

$$\boldsymbol{x}^{[n]} = (x_1 \; x_2 \; \cdots \; x_n),$$

and $\mathscr{D}_+(x \mid n)$ ($x > 0$, $n = 1,2,\ldots$) is defined as a subspace of the n-dimensional Euclidean space given by

$$\mathscr{D}_+(x \mid n) = \{\boldsymbol{x}^{[n]}; \boldsymbol{x}^{[n]} > \mathbf{0}, \; \sum_{m=1}^{n} x_m < x\}.$$

Note that by definition,

$$\Pr(V_n = 0) = 0, \qquad \Pr(V_n \le x) = \int_{0+}^{x} v(y \mid n) dy \quad (x > 0), \quad n = 1,2,\ldots,$$

and

$$\int_{0+}^{\infty} v(x \mid n) = 1, \qquad n = 1,2,\ldots. \tag{5.25}$$

The p.d.f. $v(x)$ ($x > 0$) of the workload is then given by

$$v(x) = \sum_{n=1}^{\infty} p_n v(x \mid n), \tag{5.26}$$

where from Theorem 5.1, (5.23) and (5.24),

$$p_1 v(x \mid 1) = \pi_0 \rho \tilde{h}(x), \tag{5.27}$$

$$p_n v(x \mid n) = \pi_0 \rho^n \int_{\mathscr{D}_+(x|n-1)} \tilde{h}(x_1) \left\{ \prod_{i=2}^{n-1} \overline{G}(\sum_{j=1}^{i-1} x_j) \tilde{h}(x_i) \right\}$$
$$\cdot \overline{G}(\sum_{m=1}^{n-1} x_m) \tilde{h}(x - \sum_{m=1}^{n-1} x_m) d\boldsymbol{x}^{[n-1]}, \quad n = 2, 3, \ldots. \tag{5.28}$$

With a straightforward calculation based on these equations, we can verify that

$$p_n v(x \mid n) = \rho \int_{0+}^{x} p_{n-1} v(y \mid n-1) \overline{G}(y) \tilde{h}(x-y) dy, \qquad x > 0, n = 2, 3, \ldots, \tag{5.29}$$

from which the following corollary follows immediately.

Corollary 5.2. *$\phi_n(x)$ in (5.6) and $p_n v(x \mid n)$ on the right hand side of (5.26) are related by*

$$\pi_0 \lambda^{n-1} \phi_{n-1}(x) = p_n v(x \mid n), \quad n = 1, 2, \ldots. \tag{5.30}$$

Therefore the $(n-1)$st ($n = 1, 2, \ldots$) term in the formal series solution (5.4) represents the p.d.f. of V in (5.22) when $L = n$. It can also be shown that

$$c_n = \lambda^{n-1} \int_{0+}^{\infty} \phi_{n-1}(x) dx,$$

so that (5.11) is equivalent to (5.7).

Before closing this section, we take a look at some additional results, which will be used in Chapters 6 and 7. Recall that $P_{\mathrm{admit}}(n)$ ($n = 1, 2, \ldots$) is defined as (5.17) and it denotes the probability that a randomly chosen customer is admitted to the system given that the customer finds $L = n$ on arrival. By definition, (5.17) is rewritten to be

$$P_{\mathrm{admit}}(n) = \int_{0+}^{\infty} v(x \mid n) \overline{G}(x) dx \tag{5.31}$$
$$= \mathrm{E}[\overline{G}(V_n)], \quad n = 1, 2, \ldots. \tag{5.32}$$

Taking the integral over $x \in (0, \infty)$ on both sides of (5.27) and (5.29), and using (5.14), (5.25), and (5.31), we can verify that c_n ($n = 1, 2, \ldots$) satisfies the following recursion.

$$c_1 = \rho, \qquad c_n = c_{n-1} \cdot \rho P_{\mathrm{admit}}(n-1), \quad n = 2, 3, \ldots. \tag{5.33}$$

Remark 5.3. *Using (5.12) and (5.33), we obtain an explicit formula for $P_{\mathrm{admit}}(n)$ ($n = 1, 2, \ldots$).*

$$P_{\mathrm{admit}}(1) = \overline{G}(\tilde{H}_1), \qquad P_{\mathrm{admit}}(n) = \mathrm{E}\left[\prod_{i=1}^{n} \overline{G}(\sum_{j=1}^{i} \tilde{H}_j) \right] \Big/ \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}(\sum_{j=1}^{i} \tilde{H}_j) \right],$$
$$n = 2, 3, \ldots. \tag{5.34}$$

Note that $P_{\mathrm{admit}}(n)$ ($n = 1, 2, \ldots$) is independent of the arrival rate λ.

It follows from (5.14), (5.27), (5.29), and (5.33) that $v(x \mid n)$ $(n = 1,2,\ldots)$ satisfies

$$\begin{aligned} v(x \mid 1) &= \tilde{h}(x), \quad x > 0, && (5.35) \\ v(x \mid n) &= \frac{1}{P_{\text{admit}}(n-1)} \int_{0+}^{x} v(y \mid n-1)\overline{G}(y)\tilde{h}(x-y)dy, \quad x > 0, n = 2,3,\ldots. && (5.36) \end{aligned}$$

V_n is then characterized by recursively determined random variables as follows.

Corollary 5.3. *Let $\hat{V}_n$ denote the workload in system seen by a randomly chosen admitted customer who finds $L = n$ on arrival.*

$$\hat{V}_n = [V_n \mid V_n \le G_n], \quad n = 1,2,\ldots, \tag{5.37}$$

where $\{G_n\}_{n=1,2,\ldots}$ denotes a sequence of i.i.d. random variables distributed according to the impatience time distribution. V_n then satisfies

$$\begin{aligned} V_1 &= \tilde{H}_1, && (5.38) \\ V_{n+1} &= \hat{V}_n + \tilde{H}_{n+1}, \quad n = 1,2,\ldots. && (5.39) \end{aligned}$$

Proof. Note that the p.d.f. $\hat{v}(x \mid n)$ $(x > 0,\ n = 1,2,\ldots)$ of $\hat{V}_n$ is given by

$$\hat{v}(x \mid n) = \frac{v(x \mid n)\overline{G}(x)}{P_{\text{admit}}(n)}. \tag{5.40}$$

Therefore, Corollary 5.3 immediately follows from (5.35) and (5.36). □

5.4 Special cases

In this section, we discuss three special cases: Exponential impatience times (M/G/1+M), constant impatience times (M/G/1+D), and exponential service times (M/M/1+G). Note that these three cases include most of known results in the literature, where the p.d.f. $v(x)$ of the stationary workload becomes much simpler than the general case. Using the explicit formulas (5.11) for π_0 and (5.26) for the workload $v(x)$, we can clarify the reason why the stationary workload distributions in those queues take much simpler forms than the general case.

Recall that π_0 is given in terms of c_n in (5.12) and $v(x)$ $(x > 0)$ is given by the sum of $p_n v(x \mid n)$ $(n = 1,2,\ldots)$ in (5.27) and (5.28). To simplify those, we have to calculate multiple integrals involving $\overline{G}(x_1 + x_2 + \cdots + x_i)$ $(i = 1,2,\ldots,n-1)$. In what follows, we demonstrate how those integrals are calculated by assuming specific distributions of service times/impatience times.

5.4.1 Exponential impatience times (M/G/1+M)

Consider the case that impatience times are i.i.d. according to an exponential distribution with parameter $\theta > 0$. We then have

$$\overline{G}(x) = \exp[-\theta x], \quad x \geq 0.$$

Note that $\overline{G}(x)$ has a semi-group property.

$$\overline{G}(x_1 + x_2) = \overline{G}(x_1)\overline{G}(x_2), \quad x_1 \geq 0, x_2 \geq 0. \tag{5.41}$$

The exponential distribution is the only distribution satisfying (5.41). (5.12) for $n \geq 2$ is then rewritten to be

$$c_n = \rho^n \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}(\sum_{j=1}^{i} \tilde{H}_j)\right] = \rho^n \mathrm{E}\left[\prod_{i=1}^{n-1}\prod_{j=1}^{i} \overline{G}(\tilde{H}_j)\right] = \rho^n \prod_{i=1}^{n-1} \mathrm{E}\left[\left\{\overline{G}(H_i)\right\}^{n-i}\right]. \tag{5.42}$$

Note here that

$$\mathrm{E}\left[\left\{\overline{G}(H_i)\right\}^{n-i}\right] = \int_0^\infty \exp[-(n-i)\theta x_i]\tilde{h}(x_i)dx_i.$$

We thus define $\tilde{h}^*(s)$ ($\mathrm{Re}(s) > 0$) as the LST of the equilibrium random variable for service times.

$$\tilde{h}^*(s) = \int_0^\infty \exp[-sx]\tilde{h}(x)dx = \frac{1 - h^*(s)}{s\mathrm{E}[H]}, \quad \mathrm{Re}(s) > 0,$$

where $h^*(s)$ ($\mathrm{Re}(s) > 0$) denotes the LST of the service time distribution.

$$h^*(s) = \int_0^\infty \exp[-sx]dH(x), \quad \mathrm{Re}(s) > 0.$$

It then follows from (5.42) that

$$c_n = \rho^n \prod_{i=1}^{n-1} \tilde{h}^*(i\theta), \quad n = 1, 2, \ldots, \tag{5.43}$$

and therefore π_0 is given by

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} \rho^n \prod_{i=1}^{n-1} \tilde{h}^*(i\theta)\right)^{-1}. \tag{5.44}$$

Similarly, (5.28) is rewritten to be

$$p_n v(x \mid n) = \pi_0 \rho^n \int_{\mathscr{D}_+(x|n-1)} \left\{\prod_{i=1}^{n-1} \tilde{h}(x_i)\left[\overline{G}(x_i)\right]^{n-i}\right\} \tilde{h}(x - \sum_{m=1}^{n-1} x_m) d\boldsymbol{x}^{[n-1]}.$$

It then follows from (5.12), (5.14), and (5.43) that for $n = 1,2,\dots$

$$p_n v(x \mid n) = p_n \cdot \hat{h}_0 * \hat{h}_1 * \cdots * \hat{h}_{n-1}(x),$$

where $\hat{h}_m(x)$ $(x \geq 0,\ m = 0,1,\dots)$ is defined as the p.d.f. of residual service times twisted by $\exp[-m\theta x]$.

$$\hat{h}_m(x) = \frac{\tilde{h}(x)\Big[\overline{G}(x)\Big]^m}{\displaystyle\int_{0+}^{\infty} \tilde{h}(y)\Big[\overline{G}(y)\Big]^m dy} = \frac{\tilde{h}(x)\exp[-m\theta x]}{\tilde{h}^*(m\theta)},$$

and $*$ stands for the convolution operator. Therefore, from (5.26), we obtain

$$v(x) = \pi_0 \sum_{n=1}^{\infty} c_n \cdot \hat{h}_0 * \hat{h}_1 * \cdots * \hat{h}_{n-1}(x), \quad x > 0. \tag{5.45}$$

Note that (5.44) and (5.45) are consistent with Equations (43) and (44) in [Dal65].

Remark 5.4. *It is interesting to observe that in the $M/G/1$+M queue, the workload V is given by a random sum of independent random variables X_i $(i = 1,2,\dots)$, where X_i $(i = 1,2,\dots)$ has the p.d.f. $\hat{h}_{i-1}(x)$.*

5.4.2 Constant impatience times (M/G/1+D)

In this subsection, impatience times are assumed to be constant and equal to τ. By definition,

$$\overline{G}(x) = \begin{cases} 1, & x < \tau, \\ 0, & x \geq \tau. \end{cases}$$

Note here that $\overline{G}(x)$ $(x \geq 0)$ has the following property: For any $x_1, x_2 \geq 0$,

(i) if $\overline{G}(x_1) = 0$, then $\overline{G}(x_1 + x_2) = 0$, and

(ii) if $\overline{G}(x_1 + x_2) = 1$, then $\overline{G}(x_1) = 1$.

(5.12) is then rewritten to be

$$\begin{aligned} c_n &= \rho^n \int_{x_1=0+}^{\infty} \int_{x_2=0+}^{\infty} \cdots \int_{x_{n-1}=0+}^{\infty} \left\{ \prod_{i=1}^{n-1} \tilde{h}(x_i) \overline{G}(\sum_{j=1}^{i} x_j) \right\} dx_{n-1} dx_{n-2} \cdots dx_1 \\ &= \rho^n \int_{\mathscr{D}_+(\tau|n-1)} \left\{ \prod_{i=1}^{n-1} \tilde{h}(x_i) \right\} d\boldsymbol{x}^{[n-1]}. \end{aligned} \tag{5.46}$$

Let $\tilde{H}_i$ $(i = 1,2,\dots)$ denote a sequence of i.i.d. random variables whose p.d.f. is given by $\tilde{h}(x)$ $(x \geq 0)$. We then define $\tilde{H}^{(n)} = \sum_{i=1}^{n} \tilde{H}_i$, and denote its p.d.f. (resp. PDF) by $\tilde{h}^{(n)}(x)$ (resp. $\tilde{H}^{(n)}(x)$).

$$\tilde{h}^{(n)}(x) = \underbrace{\tilde{h} * \tilde{h} * \cdots * \tilde{h}}_{n}(x), \qquad \tilde{H}^{(n)}(x) = \int_0^x \tilde{h}^{(n)}(y) dy, \quad x \geq 0.$$

It then follows from (5.46) that

$$c_n = \rho^n \tilde{H}^{(n-1)}(\tau), \quad n = 2,3,\ldots.$$

Therefore, from (5.11), we obtain

$$\pi_0 = \left(1 + \rho \sum_{n=0}^{\infty} \rho^n \tilde{H}^{(n)}(\tau)\right)^{-1},$$

where $\tilde{H}^{(0)}(x) = 1$ for all $x \geq 0$.

Similarly, it follows from (5.28) that for $n = 2,3,\ldots$,

$$\begin{aligned} p_n v(x \mid n) &= \pi_0 \rho^n \int_{\mathscr{D}_+(x|n-1)} \tilde{h}(x_1) \left\{ \prod_{i=2}^{n-1} \overline{G}(\sum_{j=1}^{i-1} x_j) \tilde{h}(x_i) \right\} \overline{G}(\sum_{m=1}^{n-1} x_m) \tilde{h}(x - \sum_{m=1}^{n-1} x_m) d\boldsymbol{x}^{[n-1]} \\ &= \begin{cases} \pi_0 \rho^n \tilde{h}^{(n)}(x), & x \leq \tau, \\ \pi_0 \rho^n \displaystyle\int_{\mathscr{D}_+(\tau|n-1)} \left\{ \prod_{i=1}^{n-1} \tilde{h}(x_i) \right\} \tilde{h}(x - \sum_{i=1}^{n-1} x_i) d\boldsymbol{x}^{[n-1]}, & x > \tau. \end{cases} \end{aligned}$$

Therefore, using (5.26), we obtain

$$v(x) = \begin{cases} \pi_0 \displaystyle\sum_{n=1}^{\infty} \rho^n \tilde{h}^{(n)}(x), & x \leq \tau, \\ \pi_0 c_1 \tilde{h}(x) + \pi_0 \displaystyle\sum_{n=2}^{\infty} c_n \cdot \tilde{h}_{[\tau]}^{(n-1)} * \tilde{h}(x), & x > \tau, \end{cases} \tag{5.47}$$

where $\tilde{h}_{[\tau]}^{(n)}(x)$ ($n = 1,2,\ldots$, $x \geq 0$) is defined as the conditional p.d.f. of $\tilde{H}^{(n)}$ given $\tilde{H}^{(n)} \leq \tau$.

$$\tilde{h}_{[\tau]}^{(n)}(x) = \frac{d}{dx}\left[\Pr(\tilde{H}^{(n)} \leq x \mid \tilde{H}^{(n)} \leq \tau)\right] = \begin{cases} \tilde{h}^{(n)}(x)/\tilde{H}^{(n)}(\tau), & x \leq \tau, \\ 0, & x > \tau. \end{cases}$$

With a straightforward calculation, we can verify that (5.47) is equivalent to the result in [BKL01, Section 4].

Remark 5.5. *(5.47) implies that the conditional workload distribution given $V \leq \tau$ is identical to that in the ordinary $M/G/1$ queue. This observation is almost obvious because the censored workload process in the $M/G/1$+D queue is stochastically identical to that in the ordinary $M/G/1$ queue if we consider censored processes obtained by observing only periods with $V \leq \tau$.*

5.4.3 Exponential service times (M/M/1+G)

Finally, we consider the case that service times are i.i.d. according to an exponential distribution with parameter μ.

$$\overline{H}(x) = \exp[-\mu x], \quad x \ge 0.$$

Because of the memoryless property of the exponential distribution, residual service times are also exponentially distributed with parameter μ.

$$\tilde{h}(x) = \mu \exp[-\mu x], \quad x > 0.$$

Therefore, for any $x_1, x_2 \ge 0$, $\tilde{h}(x)$ satisfies

$$\tilde{h}(x_1)\tilde{h}(x_2) = \mu \cdot \tilde{h}(x_1 + x_2). \tag{5.48}$$

Using (5.48), we rewrite (5.28) to be

$$p_n v(x \mid n) = \pi_0 \rho^n \mu^{n-1} \tilde{h}(x) \int_{\mathscr{D}_+(x|n-1)} \left\{ \prod_{i=2}^{n} \overline{G}(\sum_{j=1}^{i-1} x_j) \right\} d\boldsymbol{x}^{[n-1]}. \tag{5.49}$$

The integral on the right-hand side of (5.49) can be interpreted probabilistically as follows. We define $\{\tilde{G}_i^\sharp\}_{i=1,2,\ldots}$ as a sequence of i.i.d. random variables whose PDF $\tilde{G}^\sharp(x) = \Pr(\tilde{G}_i^\sharp \le x)$ is given by

$$\tilde{G}^\sharp(x) = \begin{cases} \dfrac{1}{\mathrm{E}[G^\sharp]} \displaystyle\int_0^x \overline{G}(y) dy, & x < x^\sharp, \\ 1, & x \ge x^\sharp, \end{cases}$$

where $x^\sharp$ denotes a sufficiently large real number and $\mathrm{E}[G^\sharp]$ is given by

$$\mathrm{E}[G^\sharp] = \int_0^{x^\sharp} \overline{G}(y) dy + x^\sharp \overline{G}(x^\sharp).$$

Remark 5.6. *Note that $\tilde{G}_i^\sharp$ represents an equilibrium random variable for impatience times truncated at $x^\sharp$. If $\mathrm{E}[G] < \infty$, we simply set $x^\sharp = \infty$, so that $\mathrm{E}[G^\sharp] = \mathrm{E}[G]$ and $\{\tilde{G}_i^\sharp\}_{i=1,2,\ldots}$ is an i.i.d. sequence of equilibrium impatience times.*

We then rewrite (5.49) to be

$$p_n v(x \mid n) = \pi_0 \rho (\lambda \mathrm{E}[G^\sharp])^{n-1} \tilde{h}(x) \int_{\mathscr{D}_+(x|n-1)} \left\{ \prod_{i=2}^{n} \overline{G}(\sum_{j=1}^{i-1} x_j)/\mathrm{E}[G^\sharp] \right\} d\boldsymbol{x}^{[n-1]}.$$

Note here that for $x < x^\sharp$, the p.d.f. of $\tilde{G}_i^\sharp$ is given by $\overline{G}(x)/\mathrm{E}[G^\sharp]$. We then have for $x < x^\sharp$,

$$\int_{\mathscr{D}_+(x|n-1)} \left\{ \prod_{i=2}^{n} \overline{G}(\sum_{j=1}^{i-1} x_j)/\mathrm{E}[G^\sharp] \right\} d\boldsymbol{x}^{[n-1]} = \Pr(\tilde{G}_1^\sharp < \tilde{G}_2^\sharp < \cdots < \tilde{G}_{n-1}^\sharp < x) = \frac{[\tilde{G}^\sharp(x)]^{n-1}}{(n-1)!}, \tag{5.50}$$

where the second equality in (5.50) follows from that $\tilde{G}_i^\sharp$ $(i = 1,2,\ldots,n-1)$ are i.i.d. and there are equally likely $(n-1)!$ permutations of them. We thus have

$$p_n v(x \mid n) = \pi_0 \rho \tilde{h}(x) \cdot \frac{(\lambda \mathrm{E}[G^\sharp]\tilde{G}^\sharp(x))^{n-1}}{(n-1)!} = \pi_0 \rho \tilde{h}(x) \cdot \frac{1}{(n-1)!} \cdot \left(\lambda \int_0^x \overline{G}(y) dy \right)^{n-1}. \tag{5.51}$$

Note that (5.51) holds for all $x > 0$ by setting $x^\sharp$ appropriately. Therefore, from (5.26) and (5.51), we obtain

$$\begin{aligned} v(x) &= \sum_{n=1}^{\infty} p_n v(x \mid n) \\ &= \pi_0 \rho \tilde{h}(x) \exp\left[\lambda \int_0^x \overline{G}(y) dy \right] \\ &= \pi_0 \lambda \exp\left[-\mu x + \lambda \int_0^x \overline{G}(y) dy \right], \quad x > 0. \end{aligned}$$

This is identical to Equation (49) in [Sta79].

Furthermore, π_0 is determined by

$$\pi_0 + \int_{0+}^{\infty} v(x) dx = 1,$$

i.e.,

$$\pi_0 = \left(1 + \lambda \int_{0+}^{\infty} \exp\left[-\mu x + \lambda \int_0^x \overline{G}(y) dy \right] dx \right)^{-1}. \tag{5.52}$$

Alternatively, integrating both sides of (5.51) over $x \in (0,\infty)$ and noting (5.12), we have

$$c_n = \frac{\rho}{(n-1)!} \int_0^{\infty} \mu \exp[-\mu x] \left(\lambda \int_0^x \overline{G}(y) dy \right)^{n-1} dx, \quad n = 1,2,\ldots.$$

Although π_0 can be obtained from the above equation and (5.11), it is the same as taking the power series expansion of the exponential function on the right-hand side of (5.52) and calculating the integral term by term.

5.5 Conclusion

We considered the stationary workload distribution in the M/G/1+G queue. We analyzed the LCFS-PR M/G/1 queue with workload-dependent loss, whose workload process is identical to that in the M/G/1+G queue, and we derived the joint p.d.f. of the residual service times of customers in the system. This result reveals that the workload in the M/G/1+G queue is given by a random sum of dependent random variables. This observation enables us to clarify the reason why the workload distribution in the M/G/1+M, M/G/1+D, and M/M/1+G queues are simplified dramatically, as shown in Section 5.4.

Based on this characterization, we explore various properties of the stationary loss probability P_{loss} in the M/G/1+G queue in Chapter 6, and develop an efficient computational algorithm for P_{loss} in Chapter 7.

6 Analysis of the Loss Probability in the M/G/1+G Queue

6.1 Introduction

In this chapter, we analyze the stationary loss probability P_{loss} in the M/G/1+G queue described in Section 5.1. P_{loss} is defined as the probability that a randomly chosen customer leaves the system without receiving his/her service, and it is the most fundamental quantity of interest in queues with impatient customers. Recall that π_0 denotes the stationary probability that the system is empty. Owing to Little's law, the mean number of customers being served is given by $\lambda(1-P_{\text{loss}})\mathrm{E}[H] = 1 \times (1-\pi_0)$, and therefore we obtain [Dal65]

$$P_{\text{loss}} = \frac{\rho-(1-\pi_0)}{\rho}. \tag{6.1}$$

P_{loss} is thus given in terms of π_0.

With the results of [Kov61, BBH84] on the stationary virtual waiting time, P_{loss} in the M/G/1+G queue is given by (5.7) and (6.1). Note that (6.1) is equivalent to

$$P_{\text{loss}} = \int_{0+}^{\infty} v(x)G(x)dx.$$

To the best of our knowledge, however, any further results for P_{loss} in the M/G/1+G queue are not found in the literature. Because P_{loss} is given in terms of the sequence of recursively determined functions, it is not easy to evaluate the impacts of the arrival rate and service time/impatience time distributions on P_{loss} in this formulation.

The purpose of this chapter is to explore various properties of P_{loss} and related quantities in the M/G/1+G queue, based on the new perspective on the stationary workload shown in Chapter 5. In particular, we provide formal proofs of some intuition about impacts of the arrival rate and service time/impatience time distributions on P_{loss} in the M/G/1+G queue. Furthermore, using these results, we show a theoretically interesting result that P_{loss} in the M/D/1+D queue is smallest among all M/G/1+G queues with the same and finite arrival rate, mean service time, and mean impatience time.

The rest of this chapter is organized as follows. In Section 6.2, we introduce some known results for stochastic orders, and also derive some new results as preliminaries to the analysis. In Section 6.3, we derive some properties of the stationary workload and related quantities in the M/G/1+G using stochastic orders. In Section 6.4, we analyze the loss probability based on the results in Sections 6.3 and 6.4. Finally, we conclude this chapter in Section 6.5.

6.2 Stochastic orders

6.2.1 Some known results for stochastic orders

We start with the usual stochastic order, which is commonly used to compare the magnitude of random variables.

Definition 6.1 ([SS07, Eq. (1.A.1)])**.** *Let X and Y denote non-negative random variables with complementary PDFs $\overline{F}_X(x)$ and $\overline{F}_Y(x)$ $(x \ge 0)$, respectively. X is said to be smaller than or equal to Y in the usual stochastic order (denoted by $X \le_{\text{st}} Y$) if and only if*

$$\overline{F}_X(x) \le \overline{F}_Y(x) \quad \textit{for all } x \ge 0.$$

The usual stochastic order has the following basic properties.

Lemma 6.1 ([SS07, Eq. (1.A.7), Theorem 1.A.1])**.** *Let X and Y denote non-negative random variables. $X \le_{\text{st}} Y$ if and only if any one of the following conditions hold:*

(i) $\mathrm{E}[\phi(X)] \le \mathrm{E}[\phi(Y)]$ *holds for every non-decreasing function $\phi(x)$ $(x \ge 0)$ such that* $\mathrm{E}[\phi(X)]$ *and* $\mathrm{E}[\phi(Y)]$ *exist, or*

(ii) there exist two random variables $\hat{X}$ and $\hat{Y}$ defined on the same probability space, which satisfy $\hat{X} =_{\text{st}} X$, $\hat{Y} =_{\text{st}} Y$, and $\Pr(\hat{X} \le \hat{Y}) = 1$.

Remark 6.1. *By letting $\phi(x) = x$ $(x \ge 0)$ in Lemma 6.1 (i), it is readily verified that*

$$X \le_{\text{st}} Y \;\Rightarrow\; \mathrm{E}[X] \le \mathrm{E}[Y]. \tag{6.2}$$

Even when $X \le_{\text{st}} Y$ holds, it does not necessarily follow that $[X \mid X \le y] \le_{\text{st}} [Y \mid Y \le y]$ $(y \ge 0)$ for their conditional random variables. The reversed hazard rate order is a stronger relation than the usual stochastic order, which is conserved with respect to such a conditioning.

Definition 6.2 ([SS07, Eq.(1.B.41)]). *Let X and Y denote non-negative random variables with PDFs $F_X(x)$ and $F_Y(x)$ ($x \geq 0$), respectively. X is said to be smaller than or equal to Y in the reversed hazard rate order (denoted by $X \leq_{\mathrm{rh}} Y$) if and only if*

$$F_X(x)F_Y(y) \geq F_X(y)F_Y(x) \qquad \textit{for all } 0 \leq x \leq y. \tag{6.3}$$

Remark 6.2. *By letting $y \to \infty$ in (6.3), it is readily verified that [SS07, Theorem1.B.42]*

$$X \leq_{\mathrm{rh}} Y \;\Rightarrow\; X \leq_{\mathrm{st}} Y. \tag{6.4}$$

In addition, because the PDF $F_Y(x)$ of Y satisfies $F_Y(x) \leq F_Y(y)$ ($0 \leq x \leq y$), we have for any non-negative random variable Y,

$$0 \leq_{\mathrm{rh}} Y, \tag{6.5}$$

where 0 denotes a random variable which takes value 0 with probability one.

Remark 6.3 ([SS07, Eq.(1.B.43)]). *It is easy to see that (6.3) is equivalent to*

$$[X \mid X \leq y] \leq_{\mathrm{st}} [Y \mid Y \leq y] \quad \textit{for all } y \geq 0.$$

Lemma 6.2 ([BS06, Theorem 9 (b)]). *Let X and Y denote non-negative random variables with PDFs $F_X(x)$ and $F_Y(x)$ ($x \geq 0$), respectively. Further let $\hat{X}$ and $\hat{Y}$ denote non-negative random variables whose PDFs $F_{\hat{X}}(x)$ and $F_{\hat{Y}}(x)$ ($x \geq 0$) are given by*

$$F_{\hat{X}}(x) = \frac{1}{\mathrm{E}[\phi(X)]}\int_0^x \phi(w)dF_X(w), \qquad F_{\hat{Y}}(x) = \frac{1}{\mathrm{E}[\phi(Y)]}\int_0^x \phi(w)dF_Y(w),$$

where $\phi(x)$ ($x \geq 0$) denotes a non-increasing function for which $\mathrm{E}[\phi(X)]$ and $\mathrm{E}[\phi(Y)]$ exist. It then follows that

$$X \leq_{\mathrm{rh}} Y \;\Rightarrow\; \hat{X} \leq_{\mathrm{rh}} \hat{Y}.$$

Remark 6.4. *By letting*

$$\phi(x) = \begin{cases} 1, & 0 \leq x < t, \\ 0, & x \geq t, \end{cases}$$

in Lemma 6.2, we can prove (cf. Remark 6.2)

$$X \leq_{\mathrm{rh}} Y \;\Rightarrow\; [X \mid X \leq y] \leq_{\mathrm{rh}} [Y \mid Y \leq y] \quad \textit{for all } y \geq 0.$$

Lemma 6.3 ([SS07, Lemma 1.B.44]). *Let X and Y denote non-negative random variables. Also let Z denote a non-negative random variable independent of X and Y. If Z has a non-increasing reversed hazard rate,*

$$X \leq_{\mathrm{rh}} Y \;\Rightarrow\; X + Z \leq_{\mathrm{rh}} Y + Z.$$

Remark 6.5. *For the usual stochastic order, Lemma 6.3 holds under a weaker condition. Specifically, $X \le_{\mathrm{st}} Y \Rightarrow X + Z \le_{\mathrm{st}} Y + Z$ holds for any non-negative random variable Z independent of X and Y [SS07, Theorem 1.A.3 (b)].*

We next introduce stochastic orders which compare the variability of random variables.

Definition 6.3 ([SS07, Theorem 3.A.1]). *Let X and Y denote non-negative random variables with finite equal means* $\mathrm{E}[X] = \mathrm{E}[Y]$ *and complementary PDFs* $\overline{F}_X(x)$ *and* $\overline{F}_Y(x)$ $(x \ge 0)$*, respectively. X is said to be smaller than or equal to Y in the convex order (denoted by* $X \le_{\mathrm{cx}} Y$*) if and only if*

$$\int_x^{\infty} \overline{F}_X(w)dw \le \int_x^{\infty} \overline{F}_Y(w)dw \quad \textit{for all } x \ge 0, \tag{6.6}$$

or equivalently,

$$\tilde{X} \le_{\mathrm{st}} \tilde{Y}, \tag{6.7}$$

where $\tilde{X}$ and $\tilde{Y}$ denote equilibrium random variables for X and Y, respectively.

Lemma 6.4 ([SS07, Eq. (3.A.1)]). *Let X and Y denote non-negative random variables with finite equal means* $\mathrm{E}[X] = \mathrm{E}[Y]$. *$X \le_{\mathrm{cx}} Y$ holds if and only if* $\mathrm{E}[\phi(X)] \le \mathrm{E}[\phi(Y)]$ *holds for every convex functions $\phi(x)$ $(x \ge 0)$ such that* $\mathrm{E}[\phi(X)]$ *and* $\mathrm{E}[\phi(Y)]$ *exist.*

Remark 6.6. *It follows from Lemma 6.4 that* $X \le_{\mathrm{cx}} Y \Rightarrow \mathrm{Cv}[X] \le \mathrm{Cv}[Y]$*, where* $\mathrm{Cv}[Z]$ *denotes the coefficient of variation of a non-negative random variable Z.*

Lemma 6.5 ([SS07, Theorem 3.A.24]). *Among all non-negative random variables with the same finite mean, the deterministic random variable is smallest in the convex order, i.e., for any non-negative random variable Z with finite mean* $\mathrm{E}[Z]$*,*

$$\mathrm{E}[Z] \le_{\mathrm{cx}} Z.$$

The convex order is defined as (6.6) in terms of the integrals of complementary PDFs. The excess wealth order is defined in a similar way as follows, where the lower limits of corresponding integrals are determined based on quantiles of random variables.

Definition 6.4 ([SS07, Page 164]). *Let X and Y denote non-negative random variables with finite means. X is said to be smaller than or equal to Y in the excess wealth order (denoted by* $X \le_{\mathrm{ew}} Y$*) if and only if*

$$\int_{\overline{F}_X^{-1}(p)}^{\infty} \overline{F}_X(w)dw \le \int_{\overline{F}_Y^{-1}(p)}^{\infty} \overline{F}_Y(w)dw \quad \textit{for all } 0 < p < 1,$$

where $\overline{F}_X(x)$ *and* $\overline{F}_Y(x)$ $(x \ge 0)$ *denote the complementary PDFs of X and Y, respectively.*

Lemma 6.6 ([SS07, Eq. (3.C.2)]). *Let X and Y denote non-negative random variables with finite means and complementary PDFs $\overline{F}_X(x)$ and $\overline{F}_Y(x)$, respectively. $X \le_{\mathrm{ew}} Y$ holds if and only if*

$$\Psi_Y^{-1}(z) - \Psi_X^{-1}(z) \quad \textit{is non-increasing in } z > 0, \tag{6.8}$$

where $\Psi_X(x)$ and $\Psi_Y(x)$ ($x \ge 0$) are given by

$$\Psi_X(x) = \int_x^\infty \overline{F}_X(w)dw, \qquad \Psi_Y(x) = \int_x^\infty \overline{F}_Y(w)dw. \tag{6.9}$$

Lemma 6.7 ([SS07, Eq. (3.C.8)]). *If non-negative random variables X and Y have equal means* $\mathrm{E}[X] = \mathrm{E}[Y]$,

$$X \le_{\mathrm{ew}} Y \;\Rightarrow\; X \le_{\mathrm{cx}} Y.$$

By definition, we can readily verify that the excess wealth order is *location-independent*, i.e., $X \le_{\mathrm{ew}} Y \Rightarrow X + a \le_{\mathrm{ew}} Y + b$ for any $a, b \in [0, \infty)$. The dispersive order, defined as follows, is also a location-independent order which compares the variability of random variables.

Definition 6.5 ([SS07, Eq. (3.B.1)]). *Let X and Y denote two random variables with PDFs $F_X(x)$ and $F_Y(x)$, respectively. X is said to be smaller than or equal to Y in the dispersive order (denoted by $X \le_{\mathrm{disp}} Y$) if and only if*

$$F_X^{-1}(\beta) - F_X^{-1}(\alpha) \le F_Y^{-1}(\beta) - F_Y^{-1}(\alpha) \quad \textit{for all } 0 < \alpha \le \beta < 1. \tag{6.10}$$

Lemma 6.8. *[SS07, Eq. (3.B.8)] Let X and Y denote random variables with complementary PDFs $\overline{F}_X(x)$ and $\overline{F}_Y(x)$, respectively. $X \le_{\mathrm{disp}} Y$ holds if and only if*

$$\overline{F}_Y^{-1}(p) - \overline{F}_X^{-1}(p) \;\; \textit{is non-increasing in } p \in (0,1). \tag{6.11}$$

For a sequence of random variables $X_1, X_2, \ldots, X_n$ ($n = 1, 2, \ldots$), let $X_{i:n}$ ($i = 1, 2, \ldots, n$) denote its i-th order statistic, which is defined as the i-th smallest value of $\{X_1, X_2, \ldots, X_n\}$.

Lemma 6.9 ([Bar86, Lemma 3]). *Let $\{X_j\}_{j=1,2,\ldots}$ and $\{Y_j\}_{j=1,2,\ldots}$ denote i.i.d. random variables whose PDFs are given by $F_X(x)$ and $F_Y(x)$, respectively. For $n = 1, 2, \ldots$, and $i = 1, 2, \ldots n$, we define $X_{i:n}$ (resp. $Y_{i:n}$) as the i-th order statistic of $\{X_j\}_{j=1,2,\ldots,n}$ (resp. $\{Y_j\}_{j=1,2,\ldots,n}$), and we define $C_{i:n} = X_{i:n} - X_{i-1:n}$ (resp. $D_{i:n} = Y_{i:n} - Y_{i-1:n}$), where $X_{0:n} = \inf\{x; F_X(x) > 0\}$ (resp. $Y_{0:n} = \inf\{x; F_Y(x) > 0\}$). If $X_1 \le_{\mathrm{disp}} Y_1$, it follows that*

$$\phi(C_{1:n}, C_{2:n}, \ldots, C_{n:n}) \le_{\mathrm{st}} \phi(D_{1:n}, D_{2:n}, \ldots, D_{n:n}),$$

for every function $\phi : \mathrm{R}^n \to \mathrm{R}$ non-decreasing in each argument.

6.2.2 New results for the excess wealth and dispersive orders

In this subsection, we derive some new results for the excess wealth and dispersive orders, which are used in the next section to analyze the loss probability in the M/G/1+G queue. Although they seem to be basic results for these stochastic orders, we could not find them in the literature.

We first consider the relation between the excess wealth and dispersive orders.

Lemma 6.10. *Let X and Y denote non-negative random variables with finite equal means* $\mathrm{E}[X] = \mathrm{E}[Y]$, *and let $\tilde{X}$ and $\tilde{Y}$ denote the equilibrium random variables for X and Y, respectively. We then have $X \le_{\mathrm{ew}} Y$ if and only if $\tilde{X} \le_{\mathrm{disp}} \tilde{Y}$.*

Proof. We define $\overline{F}_{\tilde{X}}(x)$ and $\overline{F}_{\tilde{Y}}(x)$ $(x \ge 0)$ as the complementary PDFs of the equilibrium random variables $\tilde{X}$ and $\tilde{Y}$, respectively. By definition, it follows that

$$\overline{F}_{\tilde{X}}(x) = \frac{\Psi_X(x)}{\mathrm{E}[X]}, \qquad \overline{F}_{\tilde{Y}}(x) = \frac{\Psi_Y(x)}{\mathrm{E}[Y]}, \tag{6.12}$$

where $\Psi_X(x)$ and $\Psi_Y(x)$ are defined as (6.9). Because $\mathrm{E}[X] = \mathrm{E}[Y]$ is assumed, it is readily verified from (6.12) and Lemmas 6.6 and 6.8 that

$$\begin{aligned} X \le_{\mathrm{ew}} Y &\Leftrightarrow \Psi_Y^{-1}(z) - \Psi_X^{-1}(z) \quad \text{is non-increasing in } z > 0 \\ &\Leftrightarrow \overline{F}_{\tilde{Y}}^{-1}(p) - \overline{F}_{\tilde{X}}^{-1}(p) \quad \text{is non-increasing in } p \in (0,1) \\ &\Leftrightarrow \tilde{X} \le_{\mathrm{disp}} \tilde{Y}. \end{aligned}$$

□

With Lemma 6.10, it is shown that a similar result to Lemma 6.5 still holds for the excess wealth order, which is a stronger relation than the convex order (see Lemma 6.7).

Lemma 6.11. *Let $U(a,b)$ $(a < b)$ denote a uniform random variable with support $[a,b)$. For any non-negative random variable Z with finite mean* $\mathrm{E}[Z]$, *let $\tilde{Z}$ denote its equilibrium random variable. It then follows that*

$$U(0, \mathrm{E}[Z]) \le_{\mathrm{disp}} \tilde{Z}.$$

Proof. If X and Y in Definition 6.5 have their p.d.f.s $f_X(x)$ and $f_Y(y)$, respectively, (6.10) is equivalent to the following inequality [SS07, Eq. (3.B.11)].

$$f_X(F_X^{-1}(p)) \ge f_Y(F_Y^{-1}(p)) \quad \text{for all } 0 < p < 1. \tag{6.13}$$

Let $f_{U(0,\mathrm{E}[Z])}(x)$ and $F_{U(0,\mathrm{E}[Z])}^{-1}(p)$ denote the p.d.f. and the quantile function of the uniform random variable $U(0, \mathrm{E}[Z])$, respectively. We then have

$$\{F_{U(0,\mathrm{E}[Z])}^{-1}(p); 0 < p < 1\} \ = \ \{x; 0 < x < \mathrm{E}[Z]\},$$

$$f_{U(0,\mathrm{E}[Z])}(x) = \frac{1}{\mathrm{E}[Z]} \quad \text{for all } 0 < x < \mathrm{E}[Z].$$

Let $\overline{F}_Z(x)$ $(x \ge 0)$ denote the complementary PDF of Z, and let $f_{\tilde{Z}}(x) = \overline{F}_Z(x)/\mathrm{E}[Z]$ denote the p.d.f. of $\tilde{Z}$. Further let $F_{\tilde{Z}}^{-1}(p)$ $(0 < p < 1)$ denote the quantile function of $\tilde{Z}$. It is readily seen that for any p $(0 < p < 1)$,

$$\begin{aligned} f_{U(0,\mathrm{E}[Z])}(F_{U(0,\mathrm{E}[Z])}^{-1}(p)) = \frac{1}{\mathrm{E}[Z]} &\ge \frac{\overline{F}_Z(F_{\tilde{Z}}^{-1}(p))}{\mathrm{E}[Z]} \\ &= f_{\tilde{Z}}(F_{\tilde{Z}}^{-1}(p)). \end{aligned}$$

Lemma 6.11 now follows from (6.13) and the above equation. □

Theorem 6.1. *Among all non-negative random variables with the same finite mean, the deterministic random variable is smallest in the excess wealth order, i.e., for any non-negative random variable Z with finite mean* $\mathrm{E}[Z]$*,*

$$\mathrm{E}[Z] \le_{\mathrm{ew}} Z.$$

Proof. Theorem 6.1 follows immediately from Lemmas 6.10 and 6.11 because the equilibrium random variable for constant $\mathrm{E}[Z]$ is a uniform random variable $U(0,\mathrm{E}[Z])$. □

6.3 Properties of workload and related quantities

As shown in Section 5.3, the stationary workload V is characterized in terms of c_n $(n = 1,2,\ldots)$ given by (5.12) and the conditional workload V_n $(n = 1,2,\ldots)$ determined by (5.37), (5.38), and (5.39). In this section, we derive some results on these quantities, using stochastic orders presented in the previous section.

Lemma 6.12. $V_n \le_{\mathrm{rh}} V_{n+1}$ *holds for* $n = 1,2,\ldots$.

Proof. We prove Lemma 6.12 by induction. We define $\hat{V}_0$ as a random variable such that $\Pr(\hat{V}_0 = 0) = 1$. It follows from (5.38) and (5.39) that $V_1 \le_{\mathrm{rh}} V_2$ is equivalent to

$$\hat{V}_0 + \tilde{H}_1 \le_{\mathrm{rh}} \hat{V}_1 + \tilde{H}_2. \tag{6.14}$$

Note that $\hat{V}_0 \le_{\mathrm{rh}} \hat{V}_1$ follows from (6.5). On the other hand, because the p.d.f. $\tilde{h}(x)$ $(x \ge 0)$ of $\tilde{H}_n$ $(n = 1,2,\ldots)$ is given by (5.10) and it is a non-increasing function of x, its reversed hazard rate $\tilde{h}(x)/\Pr(\tilde{H}_n \le x)$ is also non-increasing in $x > 0$. Therefore, using Lemma 6.3, we can verify that (6.14) holds, so that $V_1 \le_{\mathrm{rh}} V_2$.

We then assume that $V_m \le_{\mathrm{rh}} V_{m+1}$ holds for some $m = 1,2,\ldots$. Recall that the p.d.f. $\hat{v}(x \mid m)$ of $\hat{V}_m$ is given by (5.40). Using (5.32), we rewrite (5.40) to be

$$\hat{v}(x \mid m) = \frac{1}{\mathrm{E}[\overline{G}(V_n)]} \cdot v(x \mid m)\overline{G}(x).$$

Because $\overline{G}(x)$ $(x \geq 0)$ is a non-increasing function, it then follows from Lemma 6.2 and the assumption $V_m \leq_{\text{rh}} V_{m+1}$ that

$$\hat{V}_m \leq_{\text{rh}} \hat{V}_{m+1}. \tag{6.15}$$

Therefore, in the same way as (6.14), it is shown that

$$\hat{V}_m + \tilde{H}_{m+1} \leq_{\text{rh}} \hat{V}_{m+1} + \tilde{H}_{m+2},$$

which implies $V_{m+1} \leq_{\text{rh}} V_{m+2}$. We thus proved $V_n \leq_{\text{rh}} V_{n+1}$ for $n = 1, 2, \ldots$. □

Let $\overline{G}^+(x)$ $(x \geq 0)$ denote the complementary PDF of a proper random variable $[G \mid G < \infty]$, where G denotes a generic random variable for impatience times. Noting $\lim_{x\to\infty} \overline{G}(x) = g_\infty$, we rewrite (5.31) to be

$$\begin{aligned} P_{\text{admit}}(n) &= \int_0^\infty v(x \mid n) \left[g_\infty + (1 - g_\infty) \frac{\overline{G}(x) - g_\infty}{1 - g_\infty} \right] dx \\ &= g_\infty + (1 - g_\infty) \int_0^\infty v(x \mid n) \overline{G}^+(x) dx. \end{aligned} \tag{6.16}$$

We then have

$$g_\infty \leq P_{\text{admit}}(n) \leq 1, \quad n = 1, 2, \ldots. \tag{6.17}$$

Lemma 6.13. *If* $g_\infty = \lim_{x\to\infty} \overline{G}(x) > 0$,

$$\lim_{n\to\infty} \Pr(V_n > x) = 1 \quad \textit{for every } x \geq 0.$$

Proof. For $n = 2$, we have from (5.35), (5.36), and (6.17),

$$\begin{aligned} v(x \mid 2) &= \frac{1}{P_{\text{admit}}(1)} \int_0^x \tilde{h}(y) \overline{G}(y) \tilde{h}(x - y) dy \\ &\leq \frac{1}{g_\infty} \int_0^x \frac{1}{\mathrm{E}[H]} \cdot 1 \cdot \frac{1}{\mathrm{E}[H]} dy \\ &= \frac{1}{\mathrm{E}[H]} \cdot \frac{x}{g_\infty \mathrm{E}[H]}. \end{aligned}$$

Suppose for some $n \geq 2$,

$$v(x \mid n) \leq \frac{1}{\mathrm{E}[H]} \cdot \frac{1}{(n-1)!} \left(\frac{x}{g_\infty \mathrm{E}[H]} \right)^{n-1}. \tag{6.18}$$

We then have from (5.36), (6.17), and (6.18),

$$v(x \mid n+1) \leq \frac{1}{g_\infty} \int_0^x \frac{1}{\mathrm{E}[H]} \cdot \frac{1}{(n-1)!} \left(\frac{y}{g_\infty \mathrm{E}[H]} \right)^{n-1} \frac{1}{\mathrm{E}[H]} dy$$

$$= \frac{1}{\mathrm{E}[H]} \cdot \frac{1}{n!} \left(\frac{x}{g_\infty \mathrm{E}[H]} \right)^n,$$

so that (6.18) holds for $n+1$, and therefore it holds for all $n = 2, 3, \ldots$. Note that

$$\lim_{n\to\infty} \frac{1}{n!} \left(\frac{x}{g_\infty \mathrm{E}[H]} \right)^n = 0 \quad \text{for every } x \ge 0.$$

For an arbitrarily fixed $x > 0$, we consider $v(y \mid n)$ for $y \in (0, x]$. It then follows from (6.18) that $v(y \mid n)$ uniformly converges to 0 in $(0, x]$ as $n \to \infty$. We then have

$$\lim_{n\to\infty} \Pr(V_n > x) = 1 - \lim_{n\to\infty} \int_0^x v(y \mid n) dy = 1 - \int_0^x \lim_{n\to\infty} v(y \mid n) dy = 1,$$

which completes the proof. □

Theorem 6.2. $\{P_{\mathrm{admit}}(n)\}_{n=1,2,\ldots}$ *has the following properties.*

(i) $\{P_{\mathrm{admit}}(n)\}_{n=1,2,\ldots}$ *is a non-increasing sequence, and*

(ii) $\lim_{n\to\infty} P_{\mathrm{admit}}(n) = g_\infty$.

Proof. Note that $-\overline{G}(x)$ $(x \ge 0)$ is a non-decreasing function. Because $V_n \le_{\mathrm{st}} V_{n+1}$ follows from (6.4) and Lemma 6.12, we have $\mathrm{E}[\overline{G}(V_{n+1})] \ge \mathrm{E}[\overline{G}(V_n)]$ from Lemma 6.1 (i). We then obtain Theorem 6.2 (i) from (5.32).

Next we consider (ii). Note that $\lim_{n\to\infty} P_{\mathrm{admit}}(n)$ exists because of (6.17) and Theorem 6.2 (i). We first prove (ii) for $g_\infty > 0$. In this case, we have from (6.16),

$$\begin{aligned} P_{\mathrm{admit}}(n) &= g_\infty + (1 - g_\infty) \int_0^\infty v(x \mid n) \overline{G}^+(x) dx \\ &= g_\infty + (1 - g_\infty) \int_0^\infty \Pr(V_n \le x) dG^+(x), \end{aligned}$$

where $G^+(x) = 1 - \overline{G}^+(x)$. Note that

$$\int_0^\infty \Pr(V_n \le x) dG^+(x) \le \int_0^\infty dG^+(x) = 1. \tag{6.19}$$

It then follows from the dominated convergence theorem and Lemma 6.13 that

$$\lim_{n\to\infty} \int_0^\infty \Pr(V_n \le x) dG^+(x) = \int_0^\infty \left(\lim_{n\to\infty} \Pr(V_n \le x) \right) dG^+(x) = 0.$$

Therefore the theorem holds for $g_\infty > 0$.

Next we consider the case of $g_\infty = 0$. Suppose

$$\lim_{n\to\infty} P_{\mathrm{admit}}(n) = a, \tag{6.20}$$

for some $a > 0$. Because of Theorem 6.2 (i), (6.20) implies $P_{\mathrm{admit}}(n) \ge a$ for all $n = 1,2,\ldots$. Using this observation, we can show $\lim_{n\to\infty} \Pr(V_n > x) = 1$ in the same way as in the proof of Lemma 6.13. Furthermore, noting (6.19), we have

$$\begin{aligned}\lim_{n\to\infty} P_{\mathrm{admit}}(n) &= \lim_{n\to\infty} \int_0^\infty v(x \mid n)\overline{G}(x)dx = \lim_{n\to\infty} \int_0^\infty \Pr(V_n \le x)dG(x) \\ &= \int_0^\infty \left(\lim_{n\to\infty} \Pr(V_n \le x)\right) dG(x) = 0,\end{aligned}$$

which contradicts (6.20), so that we conclude $a = 0$. □

Corollary 6.1. *In a stable $M/G/1$+G queue, there exists a unique natural number n^* such that*

$$\begin{aligned}\rho P_{\mathrm{admit}}(n) &\ge 1, \quad n = 1,2,\ldots,n^* - 1, \\ \rho P_{\mathrm{admit}}(n) &< 1, \quad n = n^*, n^* + 1,\ldots.\end{aligned}$$

Therefore $\{c_n\}_{n=1,2,\ldots}$ is non-increasing if $\rho P_{\mathrm{admit}}(1) = \rho \mathrm{E}[\overline{G}(\tilde{H}_1)] \le 1$, and otherwise it is unimodal, taking its maximum value at $n = n^*$.

Proof. Corollary 6.1 immediately follows from (5.2), (5.33), and Theorem 6.2. □

Remark 6.7. *As noted in Remark 5.3, $P_{\mathrm{admit}}(n)$ is independent of the arrival rate λ. Therefore even when the stability condition (5.2) is not fulfilled, we can define $P_{\mathrm{admit}}(n)$ $(n = 1,2,\ldots)$ and c_n $(n = 1,2,\ldots)$ by (5.34) and (5.12), respectively. It is easy to see from (5.33) and Theorem 6.2 that $\sum_{n=1}^\infty c_n$ in (5.11) converges if and only if (5.2) holds.*

6.4 Analysis of loss probability

We now consider the stationary loss probability P_{loss} in the M/G/1+G queue. Recall that P_{loss} is given in terms of π_0 by (6.1), and therefore P_{loss} is determined by $\{c_n\}_{n=1,2,\ldots}$ in (5.12).

Theorem 6.3. *In the stationary $M/G/1$+G queue, P_{loss} is bounded by*

$$\left(\frac{\rho - 1}{\rho}\right)^+ < P_{\mathrm{loss}} \le \frac{\rho}{1+\rho},$$

where $(x)^+ = \max(0,x)$.

Proof. The lower bound is obvious because Little's law implies $\lambda(1 - P_{\mathrm{loss}})\mathrm{E}[H] < 1$. On the other hand, from Theorem 5.1 (i), we have

$$\pi_0 = \left(1 + \rho + \sum_{n=2}^\infty c_n\right)^{-1} \le \frac{1}{1+\rho}, \tag{6.21}$$

because $c_n \ge 0$. The upper bound is obtained from (6.1) and (6.21). □

Remark 6.8. *One may think Theorem 6.3 is trivial. However, both upper and lower bounds are fairly strict in a sense that there exists a non-trivial M/G/1+G queue whose loss probability can be approximated well by either bound as we will see in Section 7.3. Note also that the upper bound is identical to the loss probability in the ordinary M/G/1/1 queue.*

Theorem 6.4. *Consider two stationary M/G/1+G queues. Let $G^{\langle k\rangle}$ and $\tilde{H}^{\langle k\rangle}$ ($k=1,2$) denote generic random variables for impatience times and equilibrium service times, respectively, in the k-th queue. Also, let $\lambda^{\langle k\rangle}$ and $P_{\text{loss}}^{\langle k\rangle}$ ($k=1,2$) denote the arrival rate and the loss probability, respectively, in the k-th queue.*

(i) Suppose the two queues have the same equilibrium service time distribution $\tilde{H}(x)$ and the same impatience time distribution $G(x)$. We then have

$$\lambda^{\langle 1\rangle} \le \lambda^{\langle 2\rangle} \ \Rightarrow\ P_{\text{loss}}^{\langle 1\rangle} \le P_{\text{loss}}^{\langle 2\rangle}.$$

(ii) Suppose the two queues have the same traffic intensity ρ and the same impatience time distribution $G(x)$. We then have

$$\tilde{H}^{\langle 1\rangle} \le_{\text{st}} \tilde{H}^{\langle 2\rangle} \ \Rightarrow\ P_{\text{loss}}^{\langle 1\rangle} \le P_{\text{loss}}^{\langle 2\rangle}. \tag{6.22}$$

(iii) Suppose the two queues have the same arrival rate λ and the same service time distribution $H(x)$. We then have

$$G^{\langle 1\rangle} \le_{\text{st}} G^{\langle 2\rangle} \ \Rightarrow\ P_{\text{loss}}^{\langle 1\rangle} \ge P_{\text{loss}}^{\langle 2\rangle}.$$

(iv) Suppose the two queues have the same arrival rate λ, the same service time distribution $H(x)$, and the same finite mean impatience time $\mathrm{E}[G] < \infty$. *We then have*

$$G^{\langle 1\rangle} \le_{\text{ew}} G^{\langle 2\rangle} \ \Rightarrow\ P_{\text{loss}}^{\langle 1\rangle} \le P_{\text{loss}}^{\langle 2\rangle}.$$

Proof. Let $c_n^{\langle k\rangle}$ ($k=1,2$, $n=1,2,\ldots$) denote c_n in the k-th queue. Also let ρ_k ($k=1,2$) denote the traffic intensity in the k-th queue. We first consider (i). It follows from (6.1) and Theorem 5.1 (i) that

$$\begin{aligned} P_{\text{loss}}^{\langle 2\rangle} - P_{\text{loss}}^{\langle 1\rangle} &= \frac{1}{\rho_1}\cdot\frac{\sum_{n=1}^{\infty}\rho_1^n d_n}{1+\sum_{n=1}^{\infty}\rho_1^n d_n} - \frac{1}{\rho_2}\cdot\frac{\sum_{n=1}^{\infty}\rho_2^n d_n}{1+\sum_{n=1}^{\infty}\rho_2^n d_n} \\ &= \frac{\rho_2-\rho_1}{\rho_1\rho_2\left(1+\sum_{n=1}^{\infty}\rho_1^n d_n\right)\left(1+\sum_{n=1}^{\infty}\rho_2^n d_n\right)} \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(d_m d_n - d_{m+n})\rho_1^m\rho_2^n, \end{aligned}$$

(6.23)

where

$$d_1 = 1, \qquad d_n = \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}(\sum_{j=1}^{i} \tilde{H}_j)\right], \quad n = 2,3,\ldots.$$

Note here that we use the following identity in the second equality of (6.23).

$$\rho_2 \sum_{n=1}^{\infty} \rho_1^n d_n - \rho_1 \sum_{n=1}^{\infty} \rho_2^n d_n = -(\rho_2 - \rho_1) \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} d_{m+n} \rho_1^m \rho_2^n.$$

Therefore it suffices to prove $d_m d_n - d_{m+n} \geq 0$ $(m,n = 1,2,\ldots)$. From (5.34), we have $P_{\mathrm{admit}}(n) = d_{n+1}/d_n$ $(n = 1,2,\ldots)$. It then follows from Theorem 6.2 (i) that

$$\frac{d_{m+n-1}}{d_{m+n-2}} \leq \frac{d_{m+n-2}}{d_{m+n-3}} \leq \frac{d_{m+n-3}}{d_{m+n-4}} \leq \cdots \leq \frac{d_4}{d_3} \leq \frac{d_3}{d_2} \leq \frac{d_2}{d_1}.$$

Because $d_n > 0$ $(n = 1,2,\ldots)$, we have for $1 \leq m \leq n$,

$$d_1 d_{m+n-1} \leq d_2 d_{m+n-2} \leq d_3 d_{m+n-3} \leq \cdots \leq d_m d_n.$$

Furthermore, noting that $\{d_n\}_{n=1,2,\ldots}$ is a non-increasing sequence, we have

$$d_{m+n} \leq d_{m+n-1} = d_1 d_{m+n-1}.$$

We thus have $d_m d_n - d_{m+n} \geq 0$ $(m,n = 1,2,3,\ldots)$, which proves (i).

It is easy to see from (6.1) and Theorem 5.1 (i) that when the two queues have the same traffic intensity,

$$c_n^{\langle 1\rangle} \leq (\geq)\, c_n^{\langle 2\rangle} \text{ for all } n = 1,2,\ldots \Rightarrow P_{\mathrm{loss}}^{\langle 1\rangle} \geq (\leq)\, P_{\mathrm{loss}}^{\langle 2\rangle}. \qquad (6.24)$$

Keeping this in mind, we next consider (ii). By assumption, the two queues have the same traffic intensity and therefore $c_1^{\langle 1\rangle} = c_1^{\langle 2\rangle}$. Let $\overline{G}^{\langle k\rangle}(x) = \Pr(G^{\langle k\rangle} > x)$ $(k = 1,2)$. If $\overline{G}^{\langle 1\rangle}(x) \leq \overline{G}^{\langle 2\rangle}(x)$ for all $x > 0$, $\overline{G}^{\langle 1\rangle}(\sum_{j=1}^{i} \tilde{H}_j) \leq \overline{G}^{\langle 2\rangle}(\sum_{j=1}^{i} \tilde{H}_j)$ $(i = 1,2,\ldots)$ on every sample path of $\{\tilde{H}_n\}_{n=1,2,\ldots}$. It then follows from (5.12) that $c_n^{\langle 1\rangle} \leq c_n^{\langle 2\rangle}$ for all $n = 1,2,\ldots$, which proves (ii).

In the case of (iii), we can define $\{\tilde{H}_n^{\langle 1\rangle}\}_{n=1,2,\ldots}$ and $\{\tilde{H}_n^{\langle 2\rangle}\}_{n=1,2,\ldots}$ on the same probability space, and therefore we assume $\Pr(\tilde{H}_n^{\langle 1\rangle} \leq \tilde{H}_n^{\langle 2\rangle}) = 1$ $(n = 1,2,\ldots)$ without loss of generality (see Lemma 6.1 (ii)). Because $\overline{G}(x)$ is non-increasing, $\overline{G}(\sum_{j=1}^{i} \tilde{H}_j^{\langle 1\rangle}) \geq \overline{G}(\sum_{j=1}^{i} \tilde{H}_j^{\langle 2\rangle})$ $(i = 1,2,\ldots)$ on every sample path of $\{(\tilde{H}_n^{\langle 1\rangle}, \tilde{H}_n^{\langle 2\rangle})\}_{n=1,2,\ldots}$. It then follows from (5.12) that $c_n^{\langle 1\rangle} \geq c_n^{\langle 2\rangle}$ for all $n = 1,2,\ldots$, which proves (iii).

Finally we consider (iv). To prove (iv), we need the following lemma, whose proof is provided in Appendix 6.A.

Lemma 6.14. *In the M/G/1+G queue where the impatience time distribution has finite mean* $\mathrm{E}[G] < \infty$*,* c_n *(*$n = 1, 2, \ldots$*) is given by* $c_1 = \rho$ *and for* $n = 2, 3, \ldots$*,*

$$c_n = \frac{\rho(\lambda \mathrm{E}[G])^{n-1}}{(n-1)!} \cdot \mathrm{E}\left[\prod_{i=1}^{n-1} \overline{H}(\tilde{G}_{i:n-1} - \tilde{G}_{i-1:n-1})\right], \tag{6.25}$$

where $\tilde{G}_{0:n} = 0$ *(*$n = 1, 2, \ldots$*), and* $\tilde{G}_{i:n}$ *denotes the i-th order statistic of i.i.d. random variables* $\tilde{G}_1, \tilde{G}_2, \ldots, \tilde{G}_n$ *distributed according to the equilibrium distribution of impatience times.*

Note that the assumption $\mathrm{E}[G] < \infty$ is necessary for the existence of the equilibrium distribution of impatience times. Let $\tilde{G}^{\langle k \rangle}$ ($k = 1, 2$) denote a generic random variable for equilibrium impatience times in the k-th queue. Noting that $-\prod_{i=1}^{n} \overline{H}(x_i)$ ($x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$) is non-decreasing in each argument, it follows from (6.2) and Lemmas 6.9 and 6.14 that

$$\tilde{G}^{\langle 1 \rangle} \leq_{\mathrm{disp}} \tilde{G}^{\langle 2 \rangle} \Rightarrow c_n^{\langle 1 \rangle} \geq c_n^{\langle 2 \rangle}.$$

On the other hand, we have from Lemma 6.10,

$$\tilde{G}^{\langle 1 \rangle} \leq_{\mathrm{disp}} \tilde{G}^{\langle 2 \rangle} \Leftrightarrow G^{\langle 1 \rangle} \leq_{\mathrm{ew}} G^{\langle 2 \rangle}.$$

We thus obtain

$$G^{\langle 1 \rangle} \leq_{\mathrm{ew}} G^{\langle 2 \rangle} \Rightarrow c_n^{\langle 1 \rangle} \geq c_n^{\langle 2 \rangle}.$$

(iv) now follows from (6.24). □

Remark 6.9. *As stated in [VVB06], if service times are exponentially distributed (M/M/1+G queue), we can show that* $G^{\langle 1 \rangle} \leq_{\mathrm{cx}} G^{\langle 2 \rangle} \Rightarrow P_{\mathrm{loss}}^{\langle 1 \rangle} \leq P_{\mathrm{loss}}^{\langle 2 \rangle}$*. Theorem 6.4 (iv) holds for general service times, while the condition* $G^{\langle 1 \rangle} \leq_{\mathrm{ew}} G^{\langle 2 \rangle}$ *is stronger than* $G^{\langle 1 \rangle} \leq_{\mathrm{cx}} G^{\langle 2 \rangle}$*.*

Let $H^{\langle k \rangle}$ ($k = 1, 2$) denote a generic random variable for service times in the k-th queue of Theorem 6.4. If $\mathrm{E}[H^{\langle 1 \rangle}] = \mathrm{E}[H^{\langle 2 \rangle}]$ in Theorem 6.4 (ii), (6.22) can be replaced by (see (6.7))

$$H^{\langle 1 \rangle} \leq_{\mathrm{cx}} H^{\langle 2 \rangle} \Rightarrow P_{\mathrm{loss}}^{\langle 1 \rangle} \leq P_{\mathrm{loss}}^{\langle 2 \rangle}.$$

The following corollary immediately follows from this result and Lemma 6.5.

Corollary 6.2. *Consider a stationary M/G/1+G queue with mean service time* $\mathrm{E}[H]$*. The loss probability* P_{loss} *in this M/G/1+G queue is bounded below by*

$$P_{\mathrm{loss}} \geq P_{\mathrm{loss}}^{(\mathrm{M/D/1+G})},$$

where $P_{\mathrm{loss}}^{(\mathrm{M/D/1+G})}$ *denotes the loss probability in the stationary M/D/1+G queue with the same arrival rate* λ*, constant service times equal to* $\mathrm{E}[H]$*, and the same impatience time distribution* $G(x)$*.*

Corollary 6.2 shows that the stationary M/D/1+G queue has the minimum loss probability among all stationary M/G/1+G queues with the same arrival rate, the same mean service time, and the same impatience time distribution. Similarly, using Theorem 6.1 and Theorem 6.4 (iv), we can readily verify that the M/G/1+D queue is the minimum-loss model with respect to impatience times.

Corollary 6.3. *Consider a stationary M / G / 1+G queue with finite mean impatience time* $\mathrm{E}[G] < \infty$*. The loss probability* P_{loss} *in this M / G / 1+G queue is bounded below by*

$$P_{\mathrm{loss}} \geq P_{\mathrm{loss}}^{(\mathrm{M/G/1+D})},$$

where $P_{\mathrm{loss}}^{(\mathrm{M/G/1+D})}$ *denotes the loss probability in the stationary M / G / 1+D queue with the same arrival rate* λ*, the same service time distribution* $H(x)$*, and constant impatience times equal to* $\mathrm{E}[G]$.

As a consequence of these results, it is shown that P_{loss} in the M/D/1+D queue is smallest among all M/G/1+G queues with the same and finite arrival rate, mean service time, and mean impatience time.

Theorem 6.5. *Consider a stationary M / G / 1+G queue with arrival rate* $\lambda < \infty$*, mean service time* $\mathrm{E}[H] < \infty$*, and mean impatience time* $\mathrm{E}[G] < \infty$*. We then have*

$$P_{\mathrm{loss}} \geq P_{\mathrm{loss}}^{(\mathrm{M/D/1+D})},$$

where $P_{\mathrm{loss}}^{(\mathrm{M/D/1+D})}$ *denotes the loss probability in the stationary M / D / 1+D queue with the same arrival rate* λ*, constant service times equal to* $\mathrm{E}[H]$*, and constant impatience times equal to* $\mathrm{E}[G]$.

Proof. Owing to Corollary 6.2, the loss probability P_{loss} in this M/G/1+G queue is bounded below by P_{loss} in the M/D/1+G queue with the same mean service time.

$$P_{\mathrm{loss}} \geq P_{\mathrm{loss}}^{(\mathrm{M/D/1+G})}.$$

Furthermore, applying Corollary 6.3 to this M/D/1+G queue yields

$$P_{\mathrm{loss}} \geq P_{\mathrm{loss}}^{(\mathrm{M/D/1+G})} \geq P_{\mathrm{loss}}^{(\mathrm{M/D/1+D})}.$$

□

6.5 Conclusion

We considered the loss probability P_{loss} in the stationary M/G/1+G queue. We introduced some known results on stochastic orders in Section 6.2.1, and derived new results on excess wealth and dispersive orders in Section 6.2.2. In Section 6.3, we explored properties of the stationary workload and related quantities using stochastic orders and the results in Chapter 5. In Section 6.4, we analyzed the stationary loss probability P_{loss} in the M/G/1+G queue and obtained bounds in Theorem 6.3 and stochastic ordering in Theorem 6.4. Furthermore, we proved that the M/D/1+D queue achieves the minimum loss probability among all M/G/1+G queues with the same and finite arrival rate, mean service time, and mean impatience time.

Appendix

6.A Proof of Lemma 6.14

Let $\tilde{g}(x)$ $(x \ge 0)$ denote the p.d.f. of the equilibrium distribution of impatience times. For $n = 2, 3, \ldots$, it follows that

$$\begin{aligned}
&\mathrm{E}\left[\prod_{i=1}^{n-1} \overline{G}(\sum_{j=1}^{i} \tilde{H}_j)\right] \\
&= \int_{x_1=0}^{\infty}\int_{x_2=0}^{\infty}\cdots\int_{x_{n-1}=0}^{\infty} \left\{\prod_{i=1}^{n-1}\overline{G}(\sum_{j=1}^{i} x_j)\right\}\left\{\prod_{m=1}^{n-1}\tilde{h}(x_m)\right\} dx_{n-1}dx_{n-2}\cdots dx_1 \\
&= \frac{(\mathrm{E}[G])^{n-1}}{(\mathrm{E}[H])^{n-1}(n-1)!}\int_{w_1=0}^{\infty}\int_{w_2=w_1}^{\infty}\cdots\int_{w_{n-1}=w_{n-2}}^{\infty}(n-1)!\left\{\prod_{i=1}^{n-1}\tilde{g}(w_i)\right\} \\
&\qquad \cdot \overline{H}(w_1-0)\left\{\prod_{m=2}^{n-1}\overline{H}(w_m-w_{m-1})\right\} dw_{n-1}dw_{n-2}\cdots dw_1.
\end{aligned}$$

Because $(n-1)!\prod_{i=1}^{n-1}\tilde{g}(w_i)$ represents the joint p.d.f. of $(\tilde{G}_{1:n-1}, \tilde{G}_{2:n-1}, \ldots, \tilde{G}_{n-1:n-1})$, we obtain

$$\mathrm{E}\left[\prod_{i=1}^{n-1}\overline{G}(\sum_{j=1}^{i}\tilde{H}_j)\right] = \frac{(\mathrm{E}[G])^{n-1}}{(\mathrm{E}[H])^{n-1}(n-1)!}\cdot \mathrm{E}\left[\prod_{i=1}^{n-1}\overline{H}(\tilde{G}_{i:n-1}-\tilde{G}_{i-1:n-1})\right], \quad n = 2, 3, \ldots.$$

(6.25) now follows from (5.12) and the above equation. □

7 Computation of the Loss Probability in the M/G/1+PH Queue

7.1 Introduction

In this chapter we develop a computational algorithm for the loss probability P_{loss} in the stationary M/G/1+PH queue, which is a special case of the M/G/1+G queue considered in Chapters 5 and 6. We assume that impatience times of customers are i.i.d. according to a phase-type distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$, i.e., its complementary PDF $\overline{G}(x)$ is given by

$$\overline{G}(x) = \boldsymbol{\alpha} \exp[\boldsymbol{T}x]\boldsymbol{e}, \quad x \geq 0, \tag{7.1}$$

where $\boldsymbol{\alpha}$ denotes a probability vector, $\boldsymbol{T}$ denotes a defective infinitesimal generator, and $\boldsymbol{e}$ denotes a column vector whose elements are all equal to one. Note that $G(0) = 0$ since $\boldsymbol{\alpha e} = 1$. To avoid trivialities, $\boldsymbol{T} + (-\boldsymbol{T})\boldsymbol{e\alpha}$ is assumed to be irreducible. Because $\lambda < \infty$ and $\lim_{x\to\infty} G(x) = 1$, the assumption $\mathrm{E}[H] < \infty$ ensures the stability of the system [BBH84].

Recall that P_{loss} is theoretically determined by (5.11), (5.12), and (6.1). However, it is not straightforward to compute P_{loss} in the M/G/1+PH queue based on it, because the assumption (7.1) of complementary phase-type PDF does not simplify c_n $(n = 1, 2, \ldots)$ substantially. For example, for $n = 4$,

$$\begin{aligned} c_4 &= \rho^4 \mathrm{E}\left[\overline{G}(\tilde{H}_1)\overline{G}(\tilde{H}_1 + \tilde{H}_2)\overline{G}(\tilde{H}_1 + \tilde{H}_2 + \tilde{H}_3)\right] \\ &= \rho^4 \mathrm{E}\left[\boldsymbol{\alpha} \exp[\boldsymbol{T}\tilde{H}_1]\boldsymbol{e\alpha} \exp[\boldsymbol{T}(\tilde{H}_1 + \tilde{H}_2)]\boldsymbol{e\alpha} \exp[\boldsymbol{T}(\tilde{H}_1 + \tilde{H}_2 + \tilde{H}_3]\boldsymbol{e}\right] \\ &= \rho^4 \mathrm{E}\left[\boldsymbol{\alpha} \exp[\boldsymbol{T}\tilde{H}_1]\boldsymbol{e\alpha} \exp[\boldsymbol{T}\tilde{H}_1] \exp[\boldsymbol{T}\tilde{H}_2]\boldsymbol{e\alpha}\right. \\ &\qquad \left. \cdot \exp[\boldsymbol{T}\tilde{H}_1] \exp[\boldsymbol{T}\tilde{H}_2] \exp[\boldsymbol{T}\tilde{H}_3]\boldsymbol{e}\right]. \end{aligned} \tag{7.2}$$

Therefore the independence of $\tilde{H}_1, \tilde{H}_2, \ldots$ cannot be utilized in this form, except the case that $\exp[\boldsymbol{T}x]$ and $\boldsymbol{e\alpha}$ commute, which is a necessary and sufficient condition for the impatience time distribution being exponential.

To utilize the independence of $\tilde{H}_1, \tilde{H}_2, \ldots$, we may rewrite (7.2) using Kronecker product, e.g., for $n = 4$,

$$\begin{aligned} c_4 &= \rho^4(\boldsymbol{\alpha} \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha})\mathrm{E}\left[\exp[\boldsymbol{T} \otimes \boldsymbol{T} \otimes \boldsymbol{T} \cdot \tilde{H}_1]\right]\mathrm{E}\left[\exp[\boldsymbol{I} \otimes \boldsymbol{T} \otimes \boldsymbol{T} \cdot \tilde{H}_2]\right] \\ &\qquad \cdot \mathrm{E}\left[\exp[\boldsymbol{I} \otimes \boldsymbol{I} \otimes \boldsymbol{T} \cdot \tilde{H}_3]\right](\boldsymbol{e} \otimes \boldsymbol{e} \otimes \boldsymbol{e}), \end{aligned}$$

where $\otimes$ stands for Kronecker product and $\boldsymbol{I}$ denotes a unit matrix. It is also hard to compute c_n for a large n in the above formulation because the size of matrix exponents increases exponentially with n.

To the best of our knowledge, the M/G/1+PH queue has been studied only in [BB13], which considers a generalized version of the M/G/1+G queue (i.e., the M/G/1 queue with Kovalenko's impatience mechanism [Kov61]). [BB13] derives the LSTs of the workload and the busy period for this generalized model, and as a special case, explicit formulas for those in the M/G/1+PH queue are obtained. The approach taken in [BB13] is to rewrite the phase-type complementary distribution (7.1) to be the following form.

$$\overline{G}(x) = \sum_{m=1}^{M} p_m(x)\exp[-\gamma_m x], \quad x \geq 0. \tag{7.3}$$

In [BB13], a specialized formula for π_0 in the M/G/1+PH queue is derived, which is essentially the same as the result obtained by substituting (7.3) into (5.11) and (5.12). Unfortunately, this formula is also not suitable for numerical computation because if we followed it, we would have to deal with the exponentially growing number of terms. We take a look at this problem using an example of hyper-exponential impatience times.

Example 7.1. *Consider an $M/G/1$+H_2 queue, where impatience times follow a hyper-exponential distribution of order two.*

$$\overline{G}(x) = p_1\exp[-\gamma_1 x] + p_2\exp[-\gamma_2 x], \quad x \geq 0, \tag{7.4}$$

where $p_1 + p_2 = 1$. Obviously, (7.4) is a special case of (7.3). Let $\tilde{h}^*(s)$ ($\mathrm{Re}(s) > 0$*) denote the LST of the equilibrium distribution of service times. It then follows from (5.12) that*

$$\begin{aligned} c_2 &= \rho^2\mathrm{E}\left[p_1\exp[-\gamma_1\tilde{H}_1] + p_2\exp[-\gamma_2\tilde{H}_1]\right] \\ &= \rho^2\left\{p_1\tilde{h}^*(\gamma_1) + p_2\tilde{h}^*(\gamma_2)\right\}, \\ c_3 &= \rho^3\mathrm{E}\left[\left\{p_1\exp[-\gamma_1\tilde{H}_1] + p_2\exp[-\gamma_2\tilde{H}_1]\right\}\right. \\ &\qquad \left. \cdot \left\{p_1\exp\left[-\gamma_1(\tilde{H}_1 + \tilde{H}_2)\right] + p_2\exp\left[-\gamma_2(\tilde{H}_1 + \tilde{H}_2)\right]\right\}\right] \\ &= \rho^3\left\{p_1^2\tilde{h}^*(2\gamma_1)\tilde{h}^*(\gamma_1) + p_1p_2\tilde{h}^*(\gamma_1 + \gamma_2)\tilde{h}^*(\gamma_2)\right. \\ &\qquad \left. + p_1p_2\tilde{h}^*(\gamma_1 + \gamma_2)\tilde{h}^*(\gamma_1) + p_2^2\tilde{h}^*(2\gamma_2)\tilde{h}^*(\gamma_2)\right\}, \end{aligned}$$

and we can readily verify that c_n ($n = 1, 2, \ldots$) is given by the sum of 2^{n-1} different terms. We thus have to compute the sum of at least 2^{n-1} terms in computing c_n. On *the other hand, in computing π_0 by (5.11), we have to compute c_n ($n = 2, 3, \dots, N_{\text{trunc}}$) for a sufficiently large N_{trunc}. This shows that the formulation based on (7.3) is not suitable for numerical computation.*

Remark 7.1. *The LST of the workload in the M/G/1+PH queue derived in [BB13] also consists of the exponentially growing number of terms, and therefore it is difficult to use it for numerical inversion.*

To overcome this difficulty, we take another approach to compute P_{loss}, based on the uniformization technique [Tij94, Page 154] and the probabilistic structure of the workload in the M/G/1+G queue shown in Chapters 5 and 6. Note that the uniformization yields a discretized workload $N_\zeta(V)$ ($\zeta > 0$), whose probability function is given by

$$\Pr(N_\zeta(V) = n) = \int_{0+}^{\infty} \frac{\exp[-\zeta x](\zeta x)^n}{n!} \cdot v(x)dx, \quad n = 0, 1, \dots,$$

where ζ is a parameter. In the standard M/G/1 queue, $N_\zeta(V)$ is given by the stationary distribution of a Markov chain of M/G/1-type, but it is not the case in the M/G/1+PH queue owing to the level-dependent nature of the workload process. In this chapter, using the results in Chapters 5 and 6, we show that $N_\zeta(V)$ in the M/G/1+PH queue has a special structure that can be utilized in computing P_{loss}. With this approach, we develop a computational algorithm for P_{loss} in the M/G/1+PH queue, which also outputs an upper bound of its numerical error. As we will see, our algorithm is readily applicable to the M/D/1+PH, M/PH/1+PH, and M/Pareto/1+PH queues.

The rest of this chapter is organized as follows. In Section 7.2, we develop a computational algorithm for P_{loss}. Next, we provide some numerical examples in Section 7.3. Finally, we conclude this chapter in Section 7.4.

7.2 Development of computational algorithm

In this section, we develop a computational algorithm for P_{loss} under the assumption that the impatience time distribution is of phase-type, i.e., its complementary PDF $\overline{G}(x)$ ($x \ge 0$) is given by (7.1). In view of (5.33), $\{P_{\text{admit}}(n)\}_{n=1,2,\dots}$ in (5.31) is a key quantity in computing P_{loss}.

7.2.1 Uniformization

In order to obtain $\{P_{\text{admit}}(n)\}_{n=1,2,\dots}$ numerically, we apply the uniformization technique [Tij94, Page 154] to rewrite (7.1), i.e.,

$$\overline{G}(x) \;=\; \boldsymbol{\alpha} \exp[-\theta x] \exp\left[\theta x(\boldsymbol{I} + \theta^{-1}\boldsymbol{T})\right] \boldsymbol{e}$$

$$= \sum_{m=0}^{\infty} \frac{\exp[-\theta x](\theta x)^m}{m!} \cdot \overline{g}_m, \quad x \geq 0, \tag{7.5}$$

where

$$\begin{aligned} \theta &= \max_i \left|[\boldsymbol{T}]_{i,i}\right|, \\ \overline{g}_m &= \boldsymbol{\alpha}[\boldsymbol{I} + \theta^{-1}\boldsymbol{T}]^m \boldsymbol{e}, \quad m = 0, 1, \ldots. \end{aligned}$$

Note that $\{\overline{g}_m\}_{m=0,1,\ldots}$ is a non-increasing sequence,

$$\overline{g}_0 = 1, \tag{7.6}$$

and

$$\sum_{m=0}^{\infty} \overline{g}_m = \theta \boldsymbol{\alpha}(-\boldsymbol{T})^{-1}\boldsymbol{e} = \theta \mathrm{E}[G], \tag{7.7}$$

where E$[G]$ denotes the mean impatience time. Let $\overline{\boldsymbol{g}}^{\star}$ denote an $\infty \times 1$ vector whose m-th ($m = 0, 1, \ldots$) element is given by $\overline{g}_m$.

$$\overline{\boldsymbol{g}}^{\star} = (\overline{g}_0 \ \ \overline{g}_1 \ \ \overline{g}_2 \ \ \cdots)^{\top},$$

where $\top$ stands for the transpose operator.

Remark 7.2. *(7.5) implies that the phase-type distribution with representation ($\boldsymbol{\alpha}$, $\boldsymbol{T}$) is equivalent to the Coxian distribution described in Figure 7.1, which has an infinite number of stages in general*[1]*, the identical mean sojourn time θ^{-1} at stages, and heterogeneous absorption probabilities $\{q_n\}_{n=0,1,\ldots}$ satisfying*

$$\overline{g}_n = \prod_{i=0}^{n-1}(1 - q_i), \quad n = 1, 2, \ldots.$$

As we will see, the structure of stages in series enables us to develop a numerically suitable formula for $\{P_{\mathrm{admit}}(n)\}_{n=1,2,\ldots}$.

For a non-negative random variable X, let $N_\zeta(X)$ ($\zeta > 0$) denote the number of Poisson arrivals with rate ζ in an interval of length X. Associated with V_n, $\hat{V}_n$, and $\tilde{H}_n$ ($n = 1, 2, \ldots$) in (5.38) and (5.39), we define $\boldsymbol{v}^{\star}(\zeta \mid n)$ ($\zeta > 0$, $n = 1, 2, \ldots$), $\hat{\boldsymbol{v}}^{\star}(\zeta \mid n)$ ($\zeta > 0$, $n = 1, 2, \ldots$), and $\tilde{\boldsymbol{h}}^{\star}(\zeta)$ ($\zeta > 0$) as $1 \times \infty$ vectors given by

$$\begin{aligned} \boldsymbol{v}^{\star}(\zeta \mid n) &= (v^{[0]}(\zeta \mid n) \ \ v^{[1]}(\zeta \mid n) \ \ \cdots), \\ \hat{\boldsymbol{v}}^{\star}(\zeta \mid n) &= (\hat{v}^{[0]}(\zeta \mid n) \ \ \hat{v}^{[1]}(\zeta \mid n) \ \ \cdots), \\ \tilde{\boldsymbol{h}}^{\star}(\zeta) &= (\tilde{h}^{[0]}(\zeta) \ \ \tilde{h}^{[1]}(\zeta) \ \ \cdots), \end{aligned}$$

[1] The number of stages is finite if diagonal elements of $\boldsymbol{T}$ are identical, e.g., an Erlang distribution.

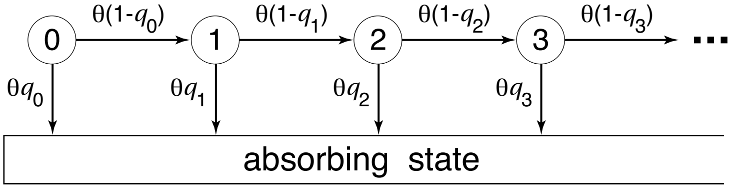

Figure 7.1: Infinite-stage Coxian distribution equivalent to the impatience time distribution.

respectively, where for $m = 0, 1, \ldots,$

$$\begin{aligned}
v^{[m]}(\zeta \mid n) &= \Pr(N_\zeta(V_n) = m) \\
&= \int_{0+}^{\infty} \frac{\exp[-\zeta x](\zeta x)^m}{m!} \cdot v(x \mid n) dx, \\
\hat{v}^{[m]}(\zeta \mid n) &= \Pr(N_\zeta(\hat{V}_n) = m) \\
&= \frac{1}{P_{\text{admit}}(n)} \int_{0+}^{\infty} \frac{\exp[-\zeta x](\zeta x)^m}{m!} \cdot v(x \mid n)\overline{G}(x) dx, \qquad (7.8) \\
\tilde{h}^{[m]}(\zeta) &= \Pr(N_\zeta(\tilde{H}_n) = m) \\
&= \int_{0+}^{\infty} \frac{\exp[-\zeta x](\zeta x)^m}{m!} \cdot \tilde{h}(x) dx.
\end{aligned}$$

Using (5.38) and (5.39), we can readily verify the following relations:

$$\begin{aligned}
N_\zeta(V_1) &= N_\zeta(\tilde{H}_1), \qquad (7.9) \\
N_\zeta(V_{n+1}) &= N_\zeta(\hat{V}_n) + N_\zeta(\tilde{H}_{n+1}), \quad n = 1, 2, \ldots. \qquad (7.10)
\end{aligned}$$

7.2.2 Main theorem for $\{P_{\text{admit}}(n)\}_{n=1,2,\ldots}$

We define $\tilde{\boldsymbol{H}}_n$ $(n = 1, 2, \ldots)$ and $\overline{\boldsymbol{B}}_n$ $(n = 1, 2, \ldots)$ as $\infty \times \infty$ stochastic and $\infty \times \infty$ substochastic matrices given by

$$\tilde{\boldsymbol{H}}_n = \begin{pmatrix} \tilde{h}^{[0]}(n\theta) & \tilde{h}^{[1]}(n\theta) & \tilde{h}^{[2]}(n\theta) & \cdots \\ 0 & \tilde{h}^{[0]}(n\theta) & \tilde{h}^{[1]}(n\theta) & \cdots \\ 0 & 0 & \tilde{h}^{[0]}(n\theta) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad (7.11)$$

$$\overline{\boldsymbol{B}}_n = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_n(1,0)\overline{g}_1 & b_n(1,1) & 0 & \cdots \\ b_n(2,0)\overline{g}_2 & b_n(2,1)\overline{g}_1 & b_n(2,2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad (7.12)$$

respectively, where $b_n(k,m)$ $(n,k = 1,2,\ldots,\ m = 0,1,\ldots,k)$ denotes the probability function of a binomial distribution given by

$$b_n(k,m) = \binom{k}{m}\left[\frac{n}{n+1}\right]^m \left[\frac{1}{n+1}\right]^{k-m}, \quad m = 0,1,\ldots,k. \tag{7.13}$$

$\{P_{\mathrm{admit}}(n)\}_{n=1,2,\ldots}$ in the M/G/1+PH queue is then given by the following theorem, which is the basis of our numerical algorithm.

Theorem 7.1. *Let $\{\boldsymbol{a}_n\}_{n=0,1,\ldots}$ denote a sequence of $\infty \times 1$ vectors given recursively by*

$$\boldsymbol{a}_0 = \overline{\boldsymbol{g}}^\star, \tag{7.14}$$

$$\boldsymbol{a}_n = \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{a}_{n-1}, \quad n = 1,2,\ldots. \tag{7.15}$$

We then have $[\boldsymbol{a}_0]_0 = 1$ and

$$[\boldsymbol{a}_n]_0 = \prod_{i=1}^{n} P_{\mathrm{admit}}(i), \quad n = 1,2,\ldots. \tag{7.16}$$

$P_{\mathrm{admit}}(n)$ is thus given by the ratio of the first elements of $\boldsymbol{a}_n$ and $\boldsymbol{a}_{n-1}$.

$$P_{\mathrm{admit}}(n) = \frac{[\boldsymbol{a}_n]_0}{[\boldsymbol{a}_{n-1}]_0}, \quad n = 1,2,\ldots. \tag{7.17}$$

Remark 7.3. *It follows from (5.33) and (7.16) that*

$$c_n = \rho^n [\boldsymbol{a}_{n-1}]_0, \quad n = 1,2,\ldots.$$

Proof. It follows from (5.32) and (7.5) that

$$\begin{aligned} P_{\mathrm{admit}}(n) &= \sum_{m=0}^{\infty} \int_{0+}^{\infty} \frac{\exp[-\theta x](\theta x)^m}{m!} \cdot v(x \mid n)dx \cdot \overline{g}_m \\ &= \boldsymbol{v}^\star(\theta \mid n)\overline{\boldsymbol{g}}^\star, \quad n = 1,2,\ldots. \end{aligned} \tag{7.18}$$

To proceed further, we need the following lemma whose proof is provided in Appendix 7.A.

Lemma 7.1. *The probability function of $N_\zeta(\hat{V}_n)$ $(\zeta > 0,\ n = 1,2,\ldots)$ is given by*

$$\Pr(N_\zeta(\hat{V}_n) = m) = \frac{1}{P_{\mathrm{admit}}(n)} \sum_{k=m}^{\infty} \Pr(N_{\zeta+\theta}(V_n) = k) b_{\zeta,\theta}(k,m) \cdot \overline{g}_{k-m}, \quad m = 0,1,\ldots, \tag{7.19}$$

where $b_{\zeta_1,\zeta_2}(k,m)$ $(\zeta_1,\zeta_2 > 0,\ k = 1,2,\ldots,\ m = 0,1,\ldots,k)$ denotes the probability function of a binomial distribution given by

$$b_{\zeta_1,\zeta_2}(k,m) = \binom{k}{m}\left[\frac{\zeta_1}{\zeta_1+\zeta_2}\right]^m \left[\frac{\zeta_2}{\zeta_1+\zeta_2}\right]^{k-m}, \quad m = 0,1,\ldots,k. \tag{7.20}$$

With straightforward calculations using (7.9), (7.10), (7.11), (7.12), and (7.19), we obtain

$$\begin{aligned} \boldsymbol{v}^{\star}(j\theta \mid 1) &= \tilde{\boldsymbol{h}}^{\star}(j\theta), \quad j = 1,2,\ldots, & (7.21)\\ \boldsymbol{v}^{\star}(j\theta \mid n) &= \hat{\boldsymbol{v}}^{\star}(j\theta \mid n-1)\tilde{\boldsymbol{H}}_j, \quad n = 2,3,\ldots,\ j = 1,2,\ldots, & (7.22)\end{aligned}$$

and for $n, j = 1,2,\ldots,$

$$\hat{\boldsymbol{v}}^{\star}(j\theta \mid n) = \frac{1}{P_{\mathrm{admit}}(n)} \cdot \boldsymbol{v}^{\star}((j+1)\theta \mid n)\overline{\boldsymbol{B}}_j. \tag{7.23}$$

Note here that from (7.13) and (7.20), we have

$$b_n(k,m) = b_{(n-1)\theta,\theta}(k,m), \qquad n = 1,2,\ldots,\ k = 1,2,\ldots,\ m = 0,1,\ldots,k.$$

It then follows from (7.21), (7.22), and (7.23) that $\boldsymbol{v}^{\star}(\theta \mid n)$ ($n = 2,3,\ldots$) is given by

$$\boldsymbol{v}^{\star}(\theta \mid n) = \frac{\tilde{\boldsymbol{h}}^{\star}(n\theta)\cdot\overline{\boldsymbol{B}}_{n-1}\tilde{\boldsymbol{H}}_{n-1}\cdot\overline{\boldsymbol{B}}_{n-2}\tilde{\boldsymbol{H}}_{n-2}\cdot\cdots\cdot\overline{\boldsymbol{B}}_1\tilde{\boldsymbol{H}}_1}{\prod_{i=1}^{n-1} P_{\mathrm{admit}}(i)}. \tag{7.24}$$

Using (7.18) and (7.24), we have

$$\begin{aligned}\prod_{i=1}^{n} P_{\mathrm{admit}}(i) &= \tilde{\boldsymbol{h}}^{\star}(n\theta)\cdot\overline{\boldsymbol{B}}_{n-1}\tilde{\boldsymbol{H}}_{n-1}\cdot\overline{\boldsymbol{B}}_{n-2}\tilde{\boldsymbol{H}}_{n-2}\cdot\cdots\cdot\overline{\boldsymbol{B}}_1\tilde{\boldsymbol{H}}_1\cdot\overline{\boldsymbol{g}}^{\star}\\ &= \tilde{\boldsymbol{h}}^{\star}(n\theta)\boldsymbol{a}_{n-1}, \quad n = 1,2,\ldots. & (7.25)\end{aligned}$$

On the other hand, it is readily seen from (7.11) and (7.12) that the first row of $\overline{\boldsymbol{B}}_n\tilde{\boldsymbol{H}}_n$ ($n = 1,2,\ldots$) is equal to $\tilde{\boldsymbol{h}}^{\star}(n\theta)$. Therefore, we have from (7.15) and (7.25),

$$[\boldsymbol{a}_n]_0 = [\overline{\boldsymbol{B}}_n\tilde{\boldsymbol{H}}_n\boldsymbol{a}_{n-1}]_0 = \tilde{\boldsymbol{h}}(n\theta)\boldsymbol{a}_{n-1} = \prod_{i=1}^{n} P_{\mathrm{admit}}(i), \quad n = 1,2,\ldots.$$

We then obtain Theorem 7.1 noting that (7.6) and (7.14) imply

$$[\boldsymbol{a}_0]_0 = [\overline{\boldsymbol{g}}^{\star}]_0 = 1.$$

□

7.2.3 Computational algorithm for P_{loss}

Based on Theorem 7.1, we develop a computational algorithm for $\boldsymbol{a}_n$ ($n = 0,1,\ldots$), from which P_{loss} is computed. Because $\boldsymbol{a}_n$ ($n = 0,1,\ldots$) in (7.14) and (7.15) has infinitely many elements, we have to truncate it, i.e., (i) we represent $\boldsymbol{a}_n$ in the form

$$\boldsymbol{a}_n = \begin{pmatrix} \boldsymbol{a}_n^{\text{finite}} \\ \boldsymbol{a}_n^{\text{error}} \end{pmatrix},$$

where $\boldsymbol{a}_n^{\text{finite}}$ denotes a finite subvector of $\boldsymbol{a}_n$, and (ii) compute an approximation $\boldsymbol{a}_n^{\text{comp}}$ to $\boldsymbol{a}_n^{\text{finite}}$ assuming $\boldsymbol{a}_n^{\text{error}} \simeq \boldsymbol{0}$. We thus approximate $\boldsymbol{a}_n$ by an ∞ × 1 vector $\boldsymbol{a}_n^{\overline{\text{comp}}}$ given by

$$\boldsymbol{a}_n^{\overline{\text{comp}}} = \begin{pmatrix} \boldsymbol{a}_n^{\text{comp}} \\ \boldsymbol{0} \end{pmatrix}. \tag{7.26}$$

Throughout this chapter, finite-size approximation vectors (resp. matrices) that we actually compute are denoted with superscript "comp," while infinite-size vectors (resp. matrices) implicitly represented by the finite-size approximation vectors (resp. matrices) are denoted with superscript "$\overline{\text{comp}}$."

To ensure that $\boldsymbol{a}_n$ is well approximated by the truncated vector of the form (7.26), $\boldsymbol{a}_n^{\text{error}}$ should be negligible when the size of $\boldsymbol{a}_n^{\text{finite}}$ is sufficiently large. The following lemma shows that it is indeed the case, whose proof is given in Appendix 7.B.

Lemma 7.2. *For each $n = 0,1,\ldots$, the sequence $\{[\boldsymbol{a}_n]_i\}_{i=0,1,\ldots}$ of the elements of $\boldsymbol{a}_n$ satisfies*

$$[\boldsymbol{a}_n]_i \geq [\boldsymbol{a}_n]_{i+1}, \quad i = 0,1,\ldots, \tag{7.27}$$

and

$$\lim_{i\to\infty} [\boldsymbol{a}_n]_i = 0. \tag{7.28}$$

In what follows, we show a specific procedure for computing $\boldsymbol{a}_n^{\text{comp}}$. For a certain $m_{\text{g}}^* \geq 1$, we first truncate the initial vector $\boldsymbol{a}_0 = \overline{\boldsymbol{g}}^{\star}$ at the m_{g}^*-th element and then simply compute the recursion (7.15) to obtain an approximation to $\boldsymbol{a}_n$ ($n = 1,2,\ldots$). Specifically, we define an ∞ × 1 vector $\overline{\boldsymbol{g}}^{\star\overline{\text{comp}}}$ as

$$\overline{\boldsymbol{g}}^{\star\overline{\text{comp}}} = \begin{pmatrix} \overline{\boldsymbol{g}}^{\star\text{comp}} \\ \boldsymbol{0} \end{pmatrix},$$

where $\overline{\boldsymbol{g}}^{\star\text{comp}}$ denotes an $m_{\text{g}}^* \times 1$ vector whose elements are given by

$$[\overline{\boldsymbol{g}}^{\star\text{comp}}]_m = \overline{g}_m = \boldsymbol{\alpha}[\boldsymbol{I} + \theta^{-1}\boldsymbol{T}]^m \boldsymbol{e}, \quad m = 0,1,\ldots,m_{\text{g}}^* - 1. \tag{7.29}$$

We assume that the truncation point m_{g}^* for $\overline{\boldsymbol{g}}^{\star}$ is determined such that the following equation holds for some $\epsilon_{\text{g}} > 0$.

$$\sum_{m=0}^{\infty} \left| [\overline{\boldsymbol{g}}^{\star} - \overline{\boldsymbol{g}}^{\star\overline{\text{comp}}}]_m \right| \leq \epsilon_{\text{g}}, \tag{7.30}$$

where from (7.7) and (7.29),

$$\begin{aligned}\sum_{m=0}^{\infty}\left|[\overline{\boldsymbol{g}}^{\star}-\overline{\boldsymbol{g}}^{\star\overline{\mathrm{comp}}}]_m\right| &= \sum_{m=0}^{\infty}\overline{g}_m-\sum_{m=0}^{m_{\mathrm{g}}^{*}-1}\overline{g}_m \\ &= \theta\boldsymbol{\alpha}(-\boldsymbol{T})^{-1}\boldsymbol{e}-\sum_{m=0}^{m_{\mathrm{g}}^{*}-1}[\overline{\boldsymbol{g}}^{\star\mathrm{comp}}]_m. \qquad (7.31)\end{aligned}$$

Note that the truncation approximation $\overline{\boldsymbol{g}}^{\star\overline{\mathrm{comp}}}$ is carried to $\overline{\boldsymbol{B}}_n$ ($n = 1,2,\ldots$) in (7.12), i.e., $\overline{\boldsymbol{B}}_n$ is approximated by $\overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}}$ as follows.

$$\overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_n(1,0)\overline{g}_1 & b_n(1,1) & 0 & \cdots \\ b_n(2,0)\overline{g}_2 & b_n(2,1)\overline{g}_1 & b_n(2,2) & \cdots \\ \vdots & \vdots & \vdots & \\ b_n(m_{\mathrm{g}}^{*}-1,0)\overline{g}_{m_{\mathrm{g}}^{*}-1} & b_n(m_{\mathrm{g}}^{*}-1,1)\overline{g}_{m_{\mathrm{g}}^{*}-2} & b_n(m_{\mathrm{g}}^{*}-1,2)\overline{g}_{m_{\mathrm{g}}^{*}-3} & \cdots \\ 0 & b_n(m_{\mathrm{g}}^{*},1)\overline{g}_{m_{\mathrm{g}}^{*}-1} & b_n(m_{\mathrm{g}}^{*},2)\overline{g}_{m_{\mathrm{g}}^{*}-2} & \cdots \\ 0 & 0 & b_n(m_{\mathrm{g}}^{*}+1,2)\overline{g}_{m_{\mathrm{g}}^{*}-1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

With those, we compute the recursion (7.14) and (7.15) to obtain an approximation $\boldsymbol{a}_n^{\overline{\mathrm{comp}}}$ to $\boldsymbol{a}_n$, i.e.,

$$\begin{aligned}\boldsymbol{a}_0^{\overline{\mathrm{comp}}} &= \overline{\boldsymbol{g}}^{\star\overline{\mathrm{comp}}}, \\ \boldsymbol{a}_n^{\overline{\mathrm{comp}}} &= \overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}}\tilde{\boldsymbol{H}}_n\boldsymbol{a}_{n-1}^{\overline{\mathrm{comp}}}, \quad n = 1,2,\ldots. \qquad (7.32)\end{aligned}$$

As illustrated in Figure 7.2, using the structures of $\tilde{\boldsymbol{H}}_n$ and $\overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}}$, we can show by induction that $\boldsymbol{a}_n^{\overline{\mathrm{comp}}}$ ($n = 1,2,\ldots$) takes the form as in (7.26) and that the size of $\boldsymbol{a}_n^{\mathrm{comp}}$ is equal to $(n+1)(m_{\mathrm{g}}^{*}-1)+1$. Furthermore, it is easy to verify that $\boldsymbol{a}_n^{\mathrm{comp}}$ ($n = 0,1,\ldots$) is computed recursively by

$$\begin{aligned}\boldsymbol{a}_0^{\mathrm{comp}} &= \overline{\boldsymbol{g}}^{\star\mathrm{comp}}, \\ \boldsymbol{a}_n^{\mathrm{comp}} &= \overline{\boldsymbol{B}}_n^{\mathrm{comp}}\tilde{\boldsymbol{H}}_n^{\mathrm{comp}}\boldsymbol{a}_{n-1}^{\mathrm{comp}}, \quad n = 1,2,\ldots, \qquad (7.33)\end{aligned}$$

where $\tilde{\boldsymbol{H}}_n^{\mathrm{comp}}$ ($n = 1,2,\ldots$) denotes an $\{n(m_{\mathrm{g}}^{*}-1)+1\}\times\{n(m_{\mathrm{g}}^{*}-1)+1\}$ matrix given by

$$\tilde{\boldsymbol{H}}_n^{\mathrm{comp}} = \begin{pmatrix} \tilde{h}^{[0]}(n\theta) & \tilde{h}^{[1]}(n\theta) & \tilde{h}^{[2]}(n\theta) & \cdots & \tilde{h}^{[n(m_{\mathrm{g}}^{*}-1)]}(n\theta) \\ 0 & \tilde{h}^{[0]}(n\theta) & \tilde{h}^{[1]}(n\theta) & \cdots & \tilde{h}^{[n(m_{\mathrm{g}}^{*}-1)-1]}(n\theta) \\ 0 & 0 & \tilde{h}^{[0]}(n\theta) & \cdots & \tilde{h}^{[n(m_{\mathrm{g}}^{*}-1)-2]}(n\theta) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{h}^{[0]}(n\theta) \end{pmatrix},$$

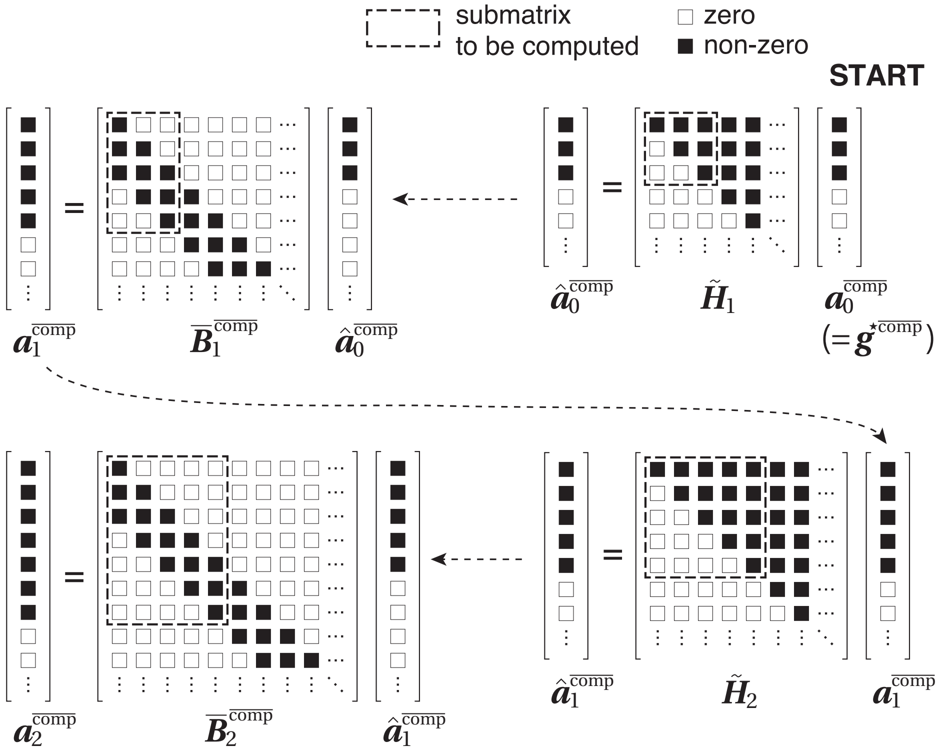
submatrix to be computed
□ zero
■ non-zero
START
=
$\boldsymbol{a}_1^{\overline{\mathrm{comp}}}$
$\overline{\boldsymbol{B}}_1^{\overline{\mathrm{comp}}}$
$\hat{\boldsymbol{a}}_0^{\overline{\mathrm{comp}}}$
$\tilde{\boldsymbol{H}}_1$
$\boldsymbol{a}_0^{\overline{\mathrm{comp}}}$
$(=\boldsymbol{g}^{\star\overline{\mathrm{comp}}})$
$\boldsymbol{a}_2^{\overline{\mathrm{comp}}}$
$\overline{\boldsymbol{B}}_2^{\overline{\mathrm{comp}}}$
$\hat{\boldsymbol{a}}_1^{\overline{\mathrm{comp}}}$
$\tilde{\boldsymbol{H}}_2$
$\boldsymbol{a}_1^{\overline{\mathrm{comp}}}$

Figure 7.2: Illustration of the recursion (7.32) in the case of $m_{\mathrm{g}}^* = 3$, where $\hat{\boldsymbol{a}}_n^{\overline{\mathrm{comp}}} = \boldsymbol{a}_n^{\overline{\mathrm{comp}}}\tilde{\boldsymbol{H}}_{n+1}$.

and $\overline{\boldsymbol{B}}_n^{\mathrm{comp}}$ $(n = 1,2,\ldots)$ denotes an $\{(n+1)(m_{\mathrm{g}}^*-1)+1\} \times \{n(m_{\mathrm{g}}^*-1)+1\}$ northwest-corner submatrix of $\overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}}$, i.e.,

$$\overline{\boldsymbol{B}}_n^{\overline{\mathrm{comp}}} = \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\mathrm{comp}} & \overline{\boldsymbol{B}}_n^{\mathrm{rest1}} \\ \boldsymbol{O} & \overline{\boldsymbol{B}}_n^{\mathrm{rest2}} \end{pmatrix}. \tag{7.34}$$

Note that $\overline{\boldsymbol{B}}_n^{\mathrm{rest1}}$ and $\overline{\boldsymbol{B}}_n^{\mathrm{rest2}}$ in (7.34) are matrices of $\{(n+1)(m_{\mathrm{g}}^*-1)+1\} \times \infty$ and $\infty \times \infty$, respectively (see Figure 7.3).

In this way, $\boldsymbol{a}_n^{\overline{\mathrm{comp}}}$ can be computed in a finite number of operations for any n. As mentioned above, the number of non-zero elements in $\boldsymbol{a}_n^{\overline{\mathrm{comp}}}$ (the size of $\boldsymbol{a}_n^{\mathrm{comp}}$) is equal to $(n+1)(m_{\mathrm{g}}^*-1)+1$ and it increases only linearly with n.

Finally, with (7.17), the approximation $P_{\mathrm{admit}}^{\mathrm{comp}}(n)$ to $P_{\mathrm{admit}}(n)$ is given by

$$P_{\mathrm{admit}}^{\mathrm{comp}}(n) = \frac{[\boldsymbol{a}_n^{\mathrm{comp}}]_0}{[\boldsymbol{a}_{n-1}^{\mathrm{comp}}]_0}, \quad n = 1,2,\ldots. \tag{7.35}$$

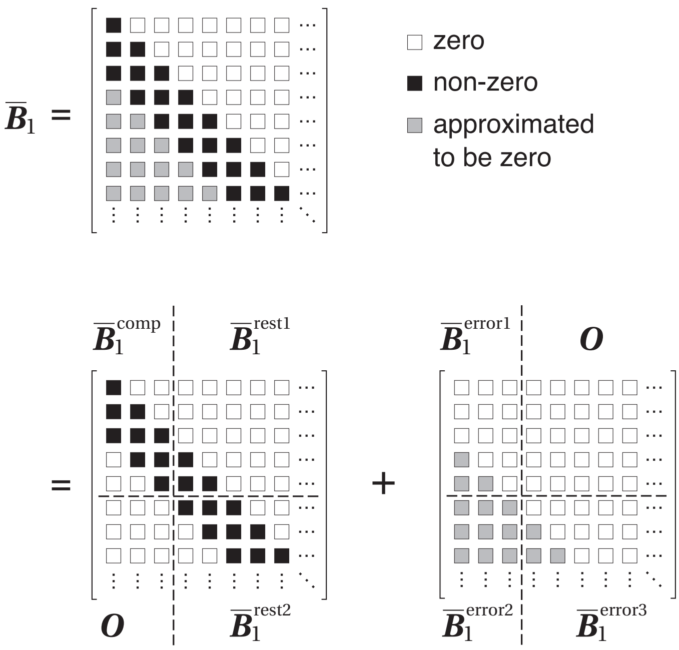
$\overline{B}_1$ =
zero
non-zero
approximated to be zero
$\overline{B}_1^{\text{comp}}$
$\overline{B}_1^{\text{rest1}}$
$\overline{B}_1^{\text{error1}}$
O
=
+
O
$\overline{B}_1^{\text{rest2}}$
$\overline{B}_1^{\text{error2}}$
$\overline{B}_1^{\text{error3}}$

Figure 7.3: Partitioning of $\overline{\boldsymbol{B}}_1$ for $m_{\text{g}}^* = 3$. It is readily verified that $\overline{\boldsymbol{B}}_2, \overline{\boldsymbol{B}}_3, \ldots$ can be partitioned in a similar way.

Remark 7.4. *In view of Remark 7.3, the approximation* c_n^{comp} *($n = 2, 3, \ldots$) to* c_n *is given by*

$$c_n^{\text{comp}} = \rho^n [\boldsymbol{a}_{n-1}^{\text{comp}}]_0, \tag{7.36}$$

which is numerically feasible if $\rho \le 1$.

Figure 7.4 summarizes the numerical algorithm for computing P_{loss} in the M/G/1+PH queue. Note that our algorithm also outputs the error bound ΔP_{loss} for $P_{\text{loss}}^{\text{comp}}$ using $\psi_{\text{c}}(n)$ in Step (c), which will be discussed in the next subsection. In Figure 7.4, we adopt a simple algorithm that determines the truncation point N_{trunc} for the infinite sum of c_n; for given $\epsilon_{\text{a}} > 0$, we stop computation if either $c_n^{\text{comp}} < \epsilon_{\text{a}}$ or $[\boldsymbol{a}_{n-1}^{\text{comp}}]_0 < \epsilon_{\text{a}}$ holds, and output $P_{\text{loss}}^{\text{comp}}$ and its error bound ΔP_{loss}. Note that in Step (b-1), we have to compute $\tilde{h}^{[i]}(n\theta)$ ($i = 0, 1, \ldots, n(m_{\text{g}}^* - 1)$). In appendix 7.D, we provide methods of computing $\tilde{h}^{[i]}(n\theta)$ for three types of service time distributions: (i) constant, (ii) phase-type distribution, and (iii) Pareto distribution.

7.2.4 Error bounds for P_{loss}

We consider error bounds for $P_{\text{loss}}^{\text{comp}}$ computed by the algorithm in Figure 7.4. We define $\boldsymbol{a}_n^{\overline{\text{error}}}$ ($n = 0, 1, \ldots$) as an $\infty \times 1$ vector given by

$$\boldsymbol{a}_n^{\overline{\text{error}}} = \boldsymbol{a}_n - \boldsymbol{a}_n^{\overline{\text{comp}}}.$$

```
Input:  $\lambda$, $H(x)$ $(x \ge 0)$, $(\boldsymbol{\alpha}, \boldsymbol{T})$, $\epsilon_{\mathrm{g}}$, and $\epsilon_{\mathrm{a}}$.
Output: $P_{\mathrm{loss}}^{\mathrm{comp}}$, $N_{\mathrm{trunc}}$, and $\Delta P_{\mathrm{loss}}$.
(a) Compute $\overline{\boldsymbol{g}}^{\star\mathrm{comp}}$ by (7.29), and determine $m_{\mathrm{g}}^*$ based on (7.30) and (7.31).
(b) Let $\boldsymbol{a}_0^{\mathrm{comp}} := \overline{\boldsymbol{g}}^{\star\mathrm{comp}}$, $c_1^{\mathrm{comp}} := \rho$, and $n := 1$.
    (b-1) Compute $\boldsymbol{a}_n^{\mathrm{comp}}$ by (7.33).
    (b-2) If $\rho \le 1$, compute $c_{n+1}^{\mathrm{comp}}$ by (7.36), and otherwise by $c_{n+1}^{\mathrm{comp}} := c_n^{\mathrm{comp}} \cdot \rho P_{\mathrm{admit}}^{\mathrm{comp}}(n)$, where $P_{\mathrm{admit}}^{\mathrm{comp}}(n)$ is given by (7.35).
    (b-3) If $\min(c_{n+1}^{\mathrm{comp}}, [\boldsymbol{a}_n^{\mathrm{comp}}]_0) < \epsilon_{\mathrm{a}}$, let $N_{\mathrm{trunc}} := n+1$ and go to (c). Otherwise $n := n+1$ and go to (b-1).
(c) Compute $\psi_{\mathrm{c}}(n)$ $(n = 1, 2, \dots, N_{\mathrm{trunc}})$ by (7.39).
(d) Compute $P_{\mathrm{loss}}^{\mathrm{comp}}$ and $\Delta P_{\mathrm{loss}}$ by (7.40) and (7.47), respectively.
```

Figure 7.4: Algorithm for computing the loss probability in the M/G/1+PH queue.

Note that

$$\boldsymbol{a}_n^{\overline{\mathrm{error}}} \ge \boldsymbol{0}. \tag{7.37}$$

We measure $\boldsymbol{a}_n^{\overline{\mathrm{error}}}$ by the L_∞ norm, where the L_∞ norm of an $\infty \times 1$ vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_\infty = \sup_{m \in \{0,1,\dots\}} \left|[\boldsymbol{x}]_m\right|.$$

Measuring the error by the L_∞ norm is reasonable because we eventually use only the first element $[\boldsymbol{a}_n]_0$ $(n = 0, 1, \dots)$ of $\boldsymbol{a}_n$ to compute c_n $(n = 1, 2, \dots)$ with (5.33) and (7.17), or with (7.36) alternatively. Note that $[\boldsymbol{a}_0^{\overline{\mathrm{error}}}]_0 = 0$ and for $n = 1, 2, \dots$,

$$\left|[\boldsymbol{a}_n^{\overline{\mathrm{error}}}]_0\right| \le \|\boldsymbol{a}_n^{\overline{\mathrm{error}}}\|_\infty. \tag{7.38}$$

Lemma 7.3. *The truncation error $\|\boldsymbol{a}_n^{\overline{\mathrm{error}}}\|_\infty$ $(n = 0, 1, \dots)$ is bounded above by*

$$\begin{aligned} \|\boldsymbol{a}_0^{\overline{\mathrm{error}}}\|_\infty &= \overline{g}_{m_{\mathrm{g}}^*} \le \epsilon_{\mathrm{g}}, \\ \|\boldsymbol{a}_n^{\overline{\mathrm{error}}}\|_\infty &\le \|\boldsymbol{a}_{n-1}^{\overline{\mathrm{error}}}\|_\infty + \epsilon_{\mathrm{g}} [\boldsymbol{a}_{n-1}^{\mathrm{comp}}]_0, \quad n = 1, 2, \dots. \end{aligned}$$

The proof of Lemma 7.3 is provided in Appendix 7.C. Lemma 7.3 implies that the computational error of $P_{\mathrm{admit}}^{\mathrm{comp}}(n)$ $(n = 1, 2, \dots)$ caused by the truncation of $\overline{\boldsymbol{g}}^\star$ can be small arbitrarily, by choosing sufficiently small ϵ_{g}. In what follows, we establish error bounds for $P_{\mathrm{loss}}^{\mathrm{comp}}$ using Lemma 7.3.

We define $\psi_{\mathrm{c}}(n)$ $(n=1,2,\ldots)$ as

$$\psi_{\mathrm{c}}(1)=\epsilon_{\mathrm{g}}+\overline{g}_{m_{\mathrm{g}}^*},\quad \psi_{\mathrm{c}}(n)=\psi_{\mathrm{c}}(n-1)+\epsilon_{\mathrm{g}}[\boldsymbol{a}_{n-1}^{\mathrm{comp}}]_0,\ n=2,3,\ldots. \tag{7.39}$$

It is readily seen from (7.37), (7.38), and Lemma 7.3 that $\psi_{\mathrm{c}}(n)$ $(n=1,2,\ldots)$ satisfies

$$0\le[\boldsymbol{a}_n]_0-[\boldsymbol{a}_n^{\mathrm{comp}}]_0\le\psi_{\mathrm{c}}(n),\quad n=1,2,\ldots.$$

Let π_0^{comp} and $P_{\mathrm{loss}}^{\mathrm{comp}}$ denote truncation approximations to π_0 and P_{loss}, respectively, which are given by

$$\pi_0^{\mathrm{comp}}=\left(1+\rho+\sum_{n=2}^{N_{\mathrm{trunc}}}c_n^{\mathrm{comp}}\right)^{-1},\qquad P_{\mathrm{loss}}^{\mathrm{comp}}=\frac{\rho-(1-\pi_0^{\mathrm{comp}})}{\rho}, \tag{7.40}$$

respectively, where N_{trunc} denotes the truncation point for the infinite sum of c_n. Noting that c_n^{comp} is given by (7.36), we can easily verify that $\pi_0^{\mathrm{comp}}\ge\pi_0$, and therefore

$$P_{\mathrm{loss}}^{\mathrm{comp}}\ge P_{\mathrm{loss}}, \tag{7.41}$$

i.e., $P_{\mathrm{loss}}^{\mathrm{comp}}$ gives an upper bound of P_{loss}.

Next we obtain a lower bound of P_{loss} from $\{[\boldsymbol{a}_n^{\mathrm{comp}}]_0\}_{n=1,2,\ldots,N_{\mathrm{trunc}}}$, which, along with the upper bound (7.41), yields the error bound of P_{loss}. To this end, we first construct an upper bound of c_n $(n=1,2,\ldots)$ using $\{[\boldsymbol{a}_n^{\mathrm{comp}}]_0\}_{n=1,2,\ldots,N_{\mathrm{trunc}}}$. We define $P_{\mathrm{admit}}^{\mathrm{upper}}(n)$ $(n=0,1,\ldots)$ as $P_{\mathrm{admit}}^{\mathrm{upper}}(0)=1$ and

$$P_{\mathrm{admit}}^{\mathrm{upper}}(n)=\begin{cases}\min\left\{\dfrac{[\boldsymbol{a}_n^{\mathrm{comp}}]_0+\psi(n)}{[\boldsymbol{a}_{n-1}^{\mathrm{comp}}]_0},\,P_{\mathrm{admit}}^{\mathrm{upper}}(n-1)\right\}, & \\ & n=1,2,\ldots,N_{\mathrm{trunc}}-1,\\ P_{\mathrm{admit}}^{\mathrm{upper}}(N_{\mathrm{trunc}}-1), & n=N_{\mathrm{trunc}},N_{\mathrm{trunc}}+1,\ldots.\end{cases}$$

Lemma 7.4. *c_n $(n=1,2,\ldots)$ is bounded above as follows.*

$$c_n\le c_n^{\mathrm{upper}},\quad n=1,2,\ldots, \tag{7.42}$$

where $c_1^{\mathrm{upper}}=\rho$ and for $n=1,2,\ldots$,

$$c_n^{\mathrm{upper}}=c_{n-1}^{\mathrm{upper}}\cdot\rho P_{\mathrm{admit}}^{\mathrm{upper}}(n-1). \tag{7.43}$$

Proof. It follows from (7.17), (7.35), and (7.39) that

$$P_{\mathrm{admit}}(n)=\frac{[\boldsymbol{a}_n]_0}{[\boldsymbol{a}_{n-1}]_0}\le\frac{[\boldsymbol{a}_n^{\mathrm{comp}}]_0+\psi(n)}{[\boldsymbol{a}_{n-1}^{\mathrm{comp}}]_0}.$$

On the other hand, it follows from Theorem 6.2 (i) that for $n=2,3,\ldots$,

$$P_{\mathrm{admit}}(n-1)\le P_{\mathrm{admit}}^{\mathrm{upper}}(n-1)\ \Rightarrow\ P_{\mathrm{admit}}(k)\le P_{\mathrm{admit}}^{\mathrm{upper}}(n-1)\quad\text{for all } k=n,n+1,\ldots.$$

We thus have $P_{\mathrm{admit}}(n)\le P_{\mathrm{admit}}^{\mathrm{upper}}(n)$ $(n=0,1,\ldots)$. (7.42) now follows from (5.33). □

We define $P_{\text{loss}}^{\text{low}}$ as

$$P_{\text{loss}}^{\text{low}} = \frac{\rho - (1 - \pi_0^{\text{low}})}{\rho},$$

where π_0^{low} denotes a lower bound of π_0, which is given by

$$\begin{aligned} \pi_0^{\text{low}} &= \left(1 + \rho + \sum_{n=2}^{\infty} c_n^{\text{upper}}\right)^{-1} \qquad (7.44) \\ &= \begin{cases} \left(1 + \rho + \displaystyle\sum_{n=2}^{N_{\text{trunc}}-2} c_n^{\text{upper}} + \frac{c_{N_{\text{trunc}}-1}^{\text{upper}}}{1 - \rho P_{\text{admit}}^{\text{upper}}(N_{\text{trunc}} - 1)}\right)^{-1}, & \\ & \rho P_{\text{admit}}^{\text{upper}}(N_{\text{trunc}} - 1) < 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Theorem 7.2. P_{loss} *is bounded as follows.*

$$\left(\frac{\rho - 1}{\rho}\right)^{+} \le P_{\text{loss}}^{\text{low}} \le P_{\text{loss}} \le P_{\text{loss}}^{\text{comp}} \le \frac{\rho}{1 + \rho}, \qquad (7.45)$$

where $(x)^{+} = \max(0, x)$*. We thus have*

$$0 \le P_{\text{loss}}^{\text{comp}} - P_{\text{loss}} \le \Delta P_{\text{loss}}, \qquad (7.46)$$

where

$$\Delta P_{\text{loss}} = P_{\text{loss}}^{\text{comp}} - P_{\text{loss}}^{\text{low}}. \qquad (7.47)$$

Proof. Because (7.46) immediately follows from (7.45), we consider (7.45). From Lemma 7.4 and the definition of π_0^{low}, we have $\pi_0^{\text{low}} \le \pi_0$, and therefore $P_{\text{loss}}^{\text{low}} \le P_{\text{loss}}$. Note that $P_{\text{loss}} \le P_{\text{loss}}^{\text{comp}}$ is given in (7.41). Therefore the remaining is to prove the first and the last inequalities in (7.45).

Because $P_{\text{admit}}^{\text{upper}}(n) \le P_{\text{admit}}^{\text{upper}}(0) = 1$ ($n = 1, 2, \ldots$), we have from (7.43) and (7.44),

$$\begin{aligned} \pi_0^{\text{low}} = \left(1 + \rho + \sum_{n=2}^{\infty} \rho^n \prod_{i=1}^{n-1} P_{\text{admit}}^{\text{upper}}(i)\right)^{-1} &\ge \left(1 + \rho + \sum_{n=2}^{\infty} \rho^n\right)^{-1} \\ &= (1 - \rho)^{+}. \end{aligned}$$

We thus have

$$P_{\text{loss}}^{\text{low}} \ge \frac{\rho - 1 + (1 - \rho)^{+}}{\rho} = \left(\frac{\rho - 1}{\rho}\right)^{+}.$$

On the other hand, $c_n^{\text{comp}} \ge 0$ ($n = 2, 3, \ldots$) implies

$$\pi_0^{\text{comp}} = \left(1 + \rho + \sum_{n=2}^{N_{\text{trunc}}} c_n^{\text{comp}}\right)^{-1} \le \frac{1}{1 + \rho},$$

which completes the proof. □

Remark 7.5. $((\rho-1)/\rho)^+$ *and* $\rho/(1+\rho)$ *in (7.45) are identical to the theoretical upper and lower bounds for* P_{loss} *shown in Theorem 6.3.*

Remark 7.6. *(7.45) implies that* ΔP_{loss} *is bounded above by*

$$\Delta P_{\text{loss}} \le \frac{\rho}{1+\rho} - \left(\frac{\rho-1}{\rho}\right)^+ = \begin{cases} \dfrac{\rho}{1+\rho}, & \rho < 1, \\ \dfrac{1}{\rho(1+\rho)}, & \rho \ge 1. \end{cases}$$

Therefore, $\Delta P_{\text{loss}} \le 1/2$ *in general, and when* ρ *takes a very large value,* $P_{\text{loss}}^{\text{comp}}$ *computed with an arbitrary sequence* $\{c_n^{\text{comp}}\}_{n=2,3,\ldots,N_{\text{trunc}}}$ *such that* $0 \le c_n^{\text{comp}} \le c_n$ *well approximates* P_{loss}*; for example,* $\Delta P_{\text{loss}} < 0.0091$ *if* $\rho \ge 10$.

Suppose we set a target accuracy ϵ_{c} in advance and try to choose the truncation point N_{trunc} for the infinite sum of c_n so that

$$\Delta P_{\text{loss}} \le \epsilon_{\text{c}}. \tag{7.48}$$

In this case, there does not necessarily exist N_{trunc} satisfying (7.48) when $\psi_{\text{c}}(n)$ $(n = 1,2,\ldots)$ is not sufficiently small. On the other hand, if c_n^{comp}'s are such accurate that $\psi_{\text{c}}(n)$'s are negligible, we have from (5.11), (6.1), and (7.40),

$$\begin{aligned} P_{\text{loss}}^{\text{comp}} - P_{\text{loss}} &= \frac{1}{\rho} \cdot \frac{\displaystyle\sum_{n=2}^{N_{\text{trunc}}} (c_n - c_n^{\text{comp}}) + \sum_{n=N_{\text{trunc}}+1}^{\infty} c_n}{\left(1+\rho+\displaystyle\sum_{n=2}^{N_{\text{trunc}}} c_n^{\text{comp}}\right)\left(1+\rho+\displaystyle\sum_{n=2}^{\infty} c_n\right)} \\ &\simeq \frac{1}{\rho} \cdot \frac{\displaystyle\sum_{n=N_{\text{trunc}}+1}^{\infty} c_n^{\text{comp}}}{\left(1+\rho+\displaystyle\sum_{n=2}^{N_{\text{trunc}}} c_n^{\text{comp}}\right)\left(1+\rho+\displaystyle\sum_{n=2}^{\infty} c_n^{\text{comp}}\right)} \\ &\le \frac{1}{\rho(1+\rho)^2} \sum_{n=N_{\text{trunc}}+1}^{\infty} c_n^{\text{comp}}, \end{aligned}$$

and therefore, if $\rho P_{\text{admit}}^{\text{comp}}(N_{\text{trunc}}-1) < 1$, Lemma 7.4 implies

$$\begin{aligned} P_{\text{loss}}^{\text{comp}} - P_{\text{loss}} &\le \frac{1}{\rho(1+\rho)^2} \cdot \frac{c_{N_{\text{trunc}}+1}^{\text{comp}}}{1-\rho P_{\text{admit}}^{\text{comp}}(N_{\text{trunc}}+1)} \\ &\le \frac{1}{(1+\rho)^2} \cdot \frac{c_{N_{\text{trunc}}}^{\text{comp}} P_{\text{admit}}^{\text{comp}}(N_{\text{trunc}}-1)}{1-\rho P_{\text{admit}}^{\text{comp}}(N_{\text{trunc}}-1)}. \end{aligned}$$

We can thus use a stopping criteria

$$\rho P_{\mathrm{admit}}^{\mathrm{comp}}(N_{\mathrm{trunc}}-1)<1, \qquad \frac{P_{\mathrm{admit}}^{\mathrm{comp}}(N_{\mathrm{trunc}}-1)}{1-\rho P_{\mathrm{admit}}^{\mathrm{comp}}(N_{\mathrm{trunc}}-1)}\cdot c_{N_{\mathrm{trunc}}}^{\mathrm{comp}} \le (1+\rho)^2\epsilon_{\mathrm{c}}, \tag{7.49}$$

to ensure $\Delta P_{\mathrm{loss}} \le \epsilon_{\mathrm{c}}$. Note that there exists N_{trunc} satisfying (7.49) for any $\epsilon_{\mathrm{c}} > 0$ because $\lim_{n\to\infty} c_n = 0$ (see Corollary 6.1).

7.3 Numerical examples

In this section, we present some numerical examples. Let E[G] and Cv[G] denote the mean and the coefficient of variation of impatience times. For the impatience time distribution, we employ the following distributions, which are determined only by the first two moments.

(i) Mixed Erlang distribution ($0 < \mathrm{Cv}[G] < 1$, denoted by $\mathrm{Er}_{k,k+1}$):
The p.d.f. $g(x)$ of impatience times is given by

$$g(x) = p\mu\cdot\frac{\exp[-\mu x](\mu x)^{k-1}}{(k-1)!} + (1-p)\mu\cdot\frac{\exp[-\mu x](\mu x)^{k}}{k!},$$

where

$$\begin{aligned} k &= \lfloor 1/(\mathrm{Cv}[G])^2 \rfloor, \\ p &= \frac{k+1}{1+(\mathrm{Cv}[G])^2}\left((\mathrm{Cv}[G])^2 - \sqrt{\frac{1-k(\mathrm{Cv}[G])^2}{k+1}}\right), \\ \mu &= \frac{pk+(1-p)(k+1)}{\mathrm{E}[G]}, \end{aligned}$$

and the phase-type representation is given by the following $1\times k$ vector $\boldsymbol{\alpha}$ and $k\times k$ matrix $\boldsymbol{T}$.

$$\boldsymbol{\alpha} = (1\ 0\ \ldots\ 0), \qquad \boldsymbol{T} = \begin{pmatrix} -\mu & \mu & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\mu & \mu & \cdots & 0 & 0 & 0 \\ 0 & 0 & -\mu & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\mu & \mu & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\mu & (1-p)\mu \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\mu \end{pmatrix}.$$

(ii) Balanced hyper-exponential distribution ($\mathrm{Cv}[G] > 1$, denoted by H_2):
The p.d.f. $g(x)$ of impatience times is given by

$$g(x) = p\mu_1\exp[-\mu_1 x] + (1-p)\mu_2\exp[-\mu_2 x],$$

where

$$p = \frac{1}{2}\left(1+\sqrt{\frac{(\mathrm{Cv}[G])^2-1}{(\mathrm{Cv}[G])^2+1}}\right), \qquad \mu_1 = \frac{2p}{\mathrm{E}[G]}, \qquad \mu_2 = \frac{2(1-p)}{\mathrm{E}[G]},$$

and the phase-type representation is given by

$$\boldsymbol{\alpha} = (p \quad 1-p), \qquad \boldsymbol{T} = \begin{pmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{pmatrix}.$$

For the service time distribution, we consider constant and Pareto distributions, where the latter is denoted by "Pareto". In all numerical examples, we choose the mean service time as a unit time (i.e., $\mathrm{E}[H] = 1$) and we set the truncation criterion ϵ_{g} for $\overline{\boldsymbol{g}}^{\star}$ to be 10^{-11} and the stopping criterion ϵ_{a} to be 10^{-9}.

Remark 7.7. *As mentioned in Example 7.1, when the impatience time distribution is either the hyper-exponential or mixed Erlang distributions, the explicit formula for c_n consists of the exponentially growing number of terms.*

Note that $\overline{\boldsymbol{g}}^{\star\overline{\mathrm{comp}}} = \overline{\boldsymbol{g}}^{\star}$ in all numerical examples with mixed Erlang impatience times, because $\epsilon_{\mathrm{a}} = 10^{-9}$ is small enough and the mixed Erlang distribution is represented by a finite-stage Coxian distribution. Note also that in all numerical examples with $\rho \le 1$, the error bound ΔP_{loss} is smaller than $\epsilon_{\mathrm{a}} = 10^{-9}$. On the other hand, in some examples with $\rho > 1$, the error bound ΔP_{loss} takes a larger value than ϵ_{a}, and in the worst case, $\Delta P_{\mathrm{loss}} \simeq 0.0013$. In Appendix 7.E, we provide such results for ΔP_{loss} as a reference.

Figures 7.5 and 7.6 show $\{P_{\mathrm{admit}}^{\mathrm{comp}}(n)\}_{n=0,1,\dots,N_{\mathrm{trunc}}-1}$ and $\{c_n^{\mathrm{comp}}\}_{n=1,2,\dots,N_{\mathrm{trunc}}}$ in the M/D/1+H$_2$ queue, where $\rho = 0.8$, $\mathrm{E}[G] \in \{1, 10, 100, 1000, 10000\}$, and $\mathrm{Cv}[G] = 3$. The non-increasing property of $\{P_{\mathrm{admit}}(n)\}_{n=1,2,\dots}$ is observed in Figure 7.5 (cf. Theorem 6.2 (i)). It is also interesting to observe that $P_{\mathrm{admit}}(N_{\mathrm{trunc}}-1) \gg \epsilon_{\mathrm{a}}$, while the stopping criterion $c_{N_{\mathrm{trunc}}} \le \epsilon_{\mathrm{a}}$ is satisfied. Note that in Figure 7.6, ρ^n is also plotted as a reference, which is the asymptotic value of c_n when $\mathrm{E}[G] \to \infty$ in this case. We observe that c_n approaches ρ^n with an increase of $\mathrm{E}[G]$.

Figure 7.7 shows $\{c_n^{\mathrm{comp}}\}_{n=1,2,\dots,N_{\mathrm{trunc}}}$ in the M/D/1+H$_2$ queue, where $\mathrm{E}[G] = 10$, $\mathrm{Cv}[G] = 3$, and $\rho \in \{0.4, 0.8, 1.2, 1.6, 2, 2.4\}$. We can see that $\{c_n^{\mathrm{comp}}\}_{n=1,2,\dots,N_{\mathrm{trunc}}}$ is either non-increasing or unimodal as shown in Corollary 6.1. When $\rho > 1$, the value of $c_{N_{\mathrm{trunc}}}$ is not very small because the computation is stopped when $[\boldsymbol{a}_{N_{\mathrm{trunc}}-1}]_0 \le \epsilon_{\mathrm{a}}$. Even though this increases the truncation error in $P_{\mathrm{loss}}^{\mathrm{comp}}$, the error bound ΔP_{loss} is not greater than 10^{-6} in all cases.

Next we examine the case of less variable impatience times. Figures 7.8–7.10 show $P_{\mathrm{admit}}^{\mathrm{comp}}(n)$ and c_n^{comp} in the M/D/1+Er$_{k,k+1}$ queue with $\mathrm{Cv}[G] = 0.3$, which correspond to Figures 7.5–7.7 for the M/D/1+H$_2$ queue, respectively. Note that $\mathrm{E}[G] \in \{1, 10, 30, 50, 100\}$ in Figures 7.8 and 7.9, which is different from Figures 7.5 and

7.6. While the fundamental properties of $P_{\text{admit}}^{\text{comp}}(n)$ and c_n^{comp} in the M/D/1+Er$_{k,k+1}$ queue is similar to those in the M/D/1+H$_2$ queue, we observe some differences between these two models:

(i) From Figures 7.5 and 7.8, we observe that $P_{\text{admit}}(n)$ in the M/D/1+H$_2$ queue is convex, while $P_{\text{admit}}(n)$ in the M/D/1+Er$_{k,k+1}$ queue has a single inflection point.

(ii) From Figures 7.6 and 7.9, we observe that as E[G] increases, c_n in the M/D/1+Er$_{k,k+1}$ queue approaches to ρ^n far more rapidly than that in the M/D/1+H$_2$ queue.

(iii) From Figures 7.7 and 7.10, we observe that if $\rho > 1$, the maximum value of $\{c_n\}_{n=1,2,\ldots,N_{\text{trunc}}}$ in the M/D/1+Er$_{k,k+1}$ queue is far larger than that in the M/D/1+H$_2$ queue.

In particular, (ii) and (iii) imply that P_{loss} in the M/D/1+Er$_{k,k+1}$ queue is smaller than P_{loss} in the M/D/1+H$_2$ queue if ρ and E[G] are identical, which can be confirmed in Figures 7.11 and 7.12, where we show $P_{\text{loss}}^{\text{comp}}$ versus ρ for these models. In these figures, the theoretical lower bound $(\rho-1)/\rho$ of P_{loss} is also plotted as a reference (cf. Remark 7.5). In both queues, $P_{\text{loss}}^{\text{comp}}$ approaches the lower bound as ρ increases, and P_{loss} for large E[G] and ρ is well approximated by the lower bound.

Next we examine the case of Pareto service time distribution. Figure 7.13 shows P_{loss} versus ρ for the M/Pareto/1+Er$_{k,k+1}$ queue with E[G] = 100, Cv[G] = 0.3 and various values of shape parameter γ of Pareto service times. When $\gamma \geq 2$, the behavior of P_{loss} is similar to that in the M/D/1+Er$_{k,k+1}$ queue with E[G] = 100 shown in Figure 7.12, i.e., it is well approximated by the theoretical lower bound $((\rho-1)/\rho)^+$. When $1 < \gamma < 2$, on the other hand, P_{loss} dramatically increases with an decrease of γ, and when γ is close to 1, P_{loss} is well approximated by the theoretical upper bound $\rho/(1+\rho)$. Therefore, the theoretical lower and upper bounds shown in Theorem 6.3 are tight in this M/Pareto/1+Er$_{k,k+1}$ queue with E[G] = 100.

Finally, Figure 7.14 shows P_{loss} versus E[G] for the M/Pareto/1+Er$_{k,k+1}$ queue with $\rho = 0.6$ and Cv[G] = 0.3. We observe that the curves in the figure take various positions between the lower bound 0 and the upper bound 0.375. We also observe that when γ is close to 1, P_{loss} is almost insensitive to E[G].

7.4 Conclusion

We developed a computational algorithm for the stationary loss probability P_{loss} in the M/G/1+PH queue. Using the uniformization and the results of Chapters 5 and 6, we proved that P_{loss} in the M/G/1+PH queue is given in terms of the sequence of $\infty \times 1$ vectors $\boldsymbol{a}_n$, which can be efficiently computed. The developed algorithm for P_{loss} is summarized in Figure 7.4. The particular feature of this algorithm is that it

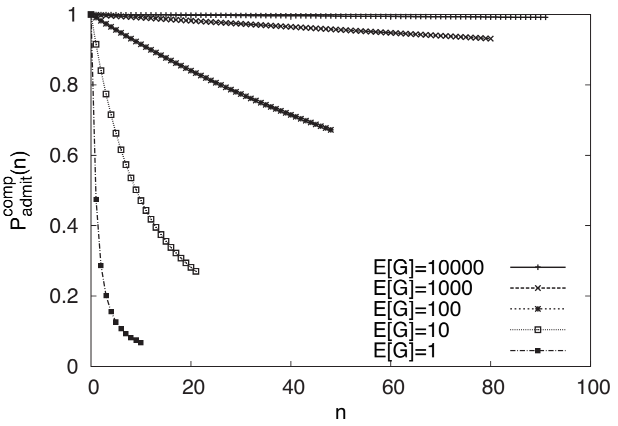

Figure 7.5: $P^{\text{comp}}_{\text{admit}}(n)$ in the M/D/1+H$_2$ queue with $\rho = 0.8$ and Cv$[G] = 3.0$.

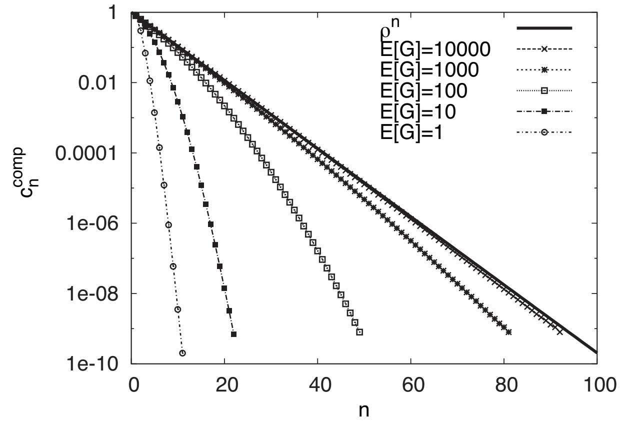

Figure 7.6: c^{comp}_n in the M/D/1+H$_2$ queue with $\rho = 0.8$ and Cv$[G] = 3.0$.

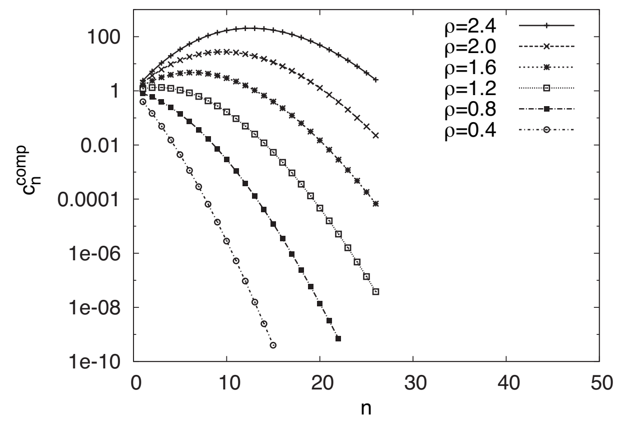

Figure 7.7: c^{comp}_n in the M/D/1+H$_2$ queue with E$[G] = 10$ and Cv$[G] = 3.0$.

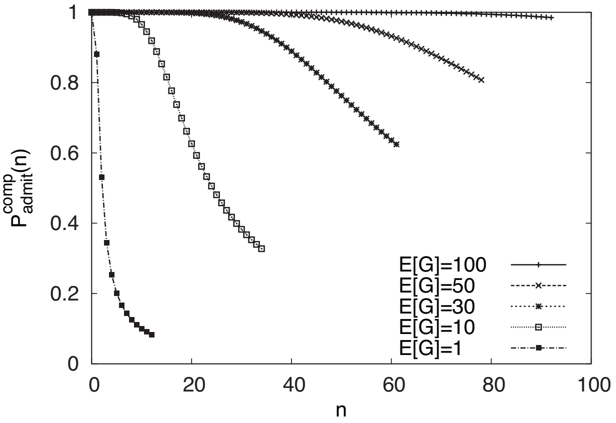

Figure 7.8: $P^{\text{comp}}_{\text{admit}}(n)$ in the M/D/1+Er$_{k,k+1}$ queue with Cv$[G] = 0.3$.

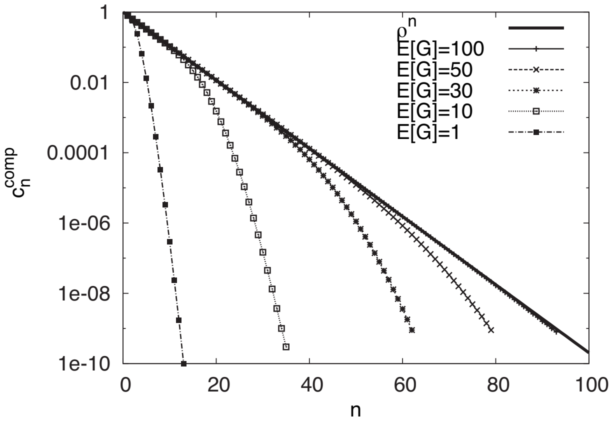

Figure 7.9: c_n^{comp} in the M/D/1+Er$_{k,k+1}$ queue with $\rho = 0.8$ and Cv$[G] = 0.3$.

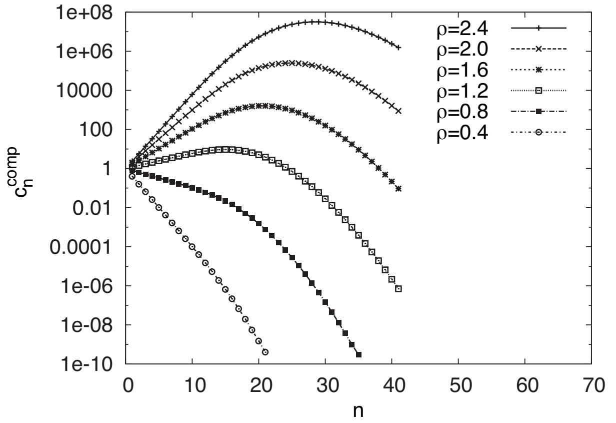

Figure 7.10: c_n^{comp} in the M/D/1+Er$_{k,k+1}$ queue with E$[G] = 10$ and Cv$[G] = 0.3$.

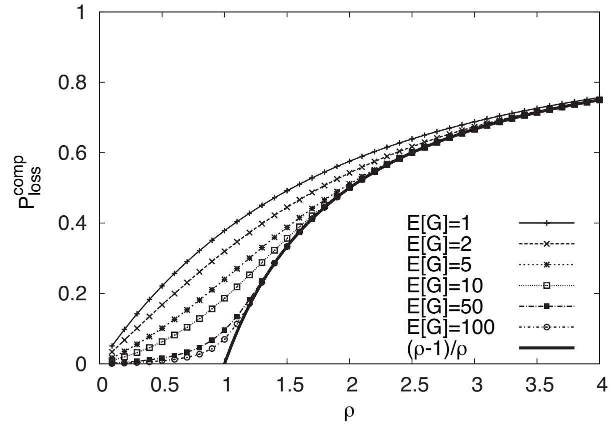

Figure 7.11: $P_{\text{loss}}^{\text{comp}}$ in the M/D/1+H$_2$ queue with Cv[G] = 3.0.

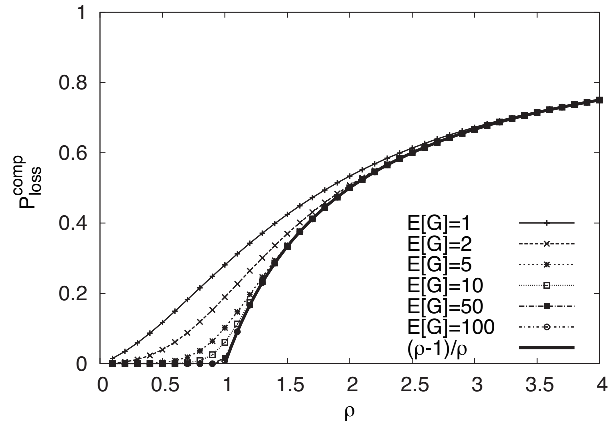

Figure 7.12: $P_{\text{loss}}^{\text{comp}}$ in the M/D/1+Er$_{k,k+1}$ queue with Cv[G] = 0.3.

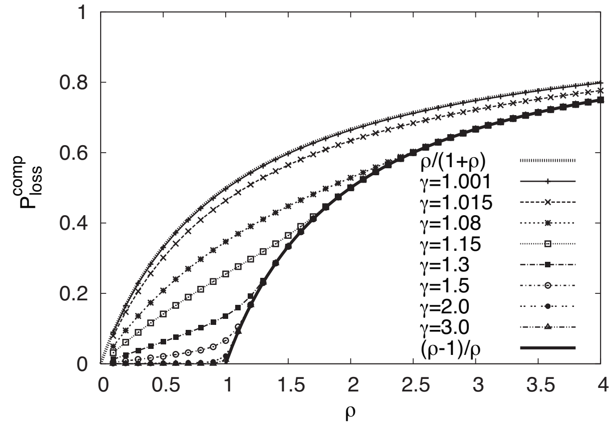

Figure 7.13: $P_{\text{loss}}^{\text{comp}}$ in the M/Pareto/1+Er$_{k,k+1}$ queue with E[G] = 100 and Cv[G] = 0.3.

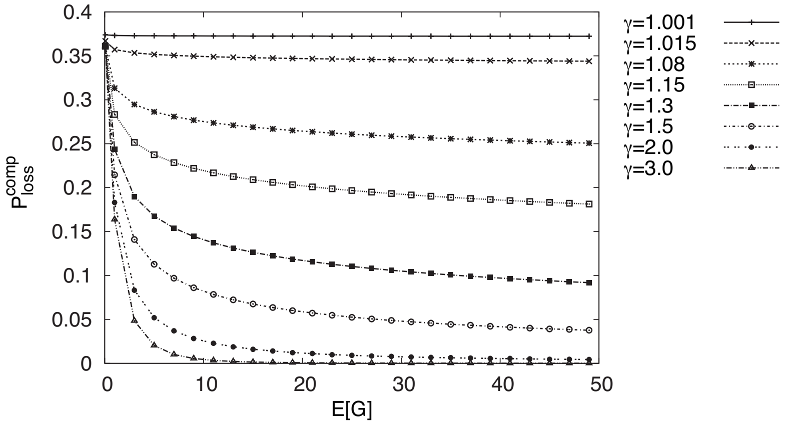

Figure 7.14: $P_{\text{loss}}^{\text{comp}}$ in the M/Pareto/1+Er$_{k,k+1}$ queue with $\rho = 0.6$ and Cv$[G] = 0.3$.

also outputs an upper bound of the numerical error in computed P_{loss}. We further presented some numerical examples in Section 7.3, some of which are not easy to compute if we followed previously known results.

Appendices

7.A Proof of Lemma 7.1

Straightforward calculations with (7.5) and (7.8) yield

$$\begin{aligned}
&\Pr(N_\zeta(\hat{V}_n) = m) \\
&= \frac{1}{P_{\text{admit}}(n)} \sum_{i=0}^{\infty} \overline{g}_i \int_{0+}^{\infty} \frac{\exp[-\theta x](\theta x)^i}{i!} \cdot \frac{\exp[-\zeta x](\zeta x)^m}{m!} \cdot v(x \mid n) dx \\
&= \frac{1}{P_{\text{admit}}(n)} \sum_{i=0}^{\infty} \overline{g}_i \cdot \frac{(i+m)!}{i!m!} \left[\frac{\zeta}{\zeta+\theta}\right]^m \left[\frac{\theta}{\zeta+\theta}\right]^i \\
&\qquad\qquad \cdot \int_{0+}^{\infty} \frac{\exp[-(\zeta+\theta)x][(\zeta+\theta)x]^{i+m}}{(i+m)!} \cdot v(x \mid n) dx \\
&= \frac{1}{P_{\text{admit}}(n)} \sum_{i=0}^{\infty} \Pr(N_{\zeta+\theta}(V_n) = i+m) b_{\zeta,\theta}(i+m,m) \cdot \overline{g}_i,
\end{aligned}$$

from which (7.19) follows. □

7.B Proof of Lemma 7.2

7.B.1 Proof of (7.27)

Let $\mathscr{F}$ denote a set of $\infty \times 1$ non-negative vectors whose elements are in a descending order.

$$\mathscr{F} = \{(x_0 \ x_1 \ \cdots)^\top \geq \mathbf{0}; x_i \geq x_{i+1} \, (i = 0,1,\ldots)\}.$$

We will show that for $n = 1,2,\ldots$,

(i) $\boldsymbol{x} \in \mathscr{F} \Rightarrow \tilde{\boldsymbol{H}}_n \boldsymbol{x} \in \mathscr{F}$, and

(ii) $\boldsymbol{x} \in \mathscr{F} \Rightarrow \overline{\boldsymbol{B}}_n \boldsymbol{x} \in \mathscr{F}$.

Note that if (i) and (ii) hold, we can obtain (7.27) recursively, using $\overline{\boldsymbol{g}}^\star \in \mathscr{F}$, (7.14), and (7.15).

It follows from the assumption $\boldsymbol{x} \in \mathscr{F}$ and the definition (7.11) of $\tilde{\boldsymbol{H}}_n$ that

$$\begin{aligned}
&[\tilde{\boldsymbol{H}}_n \boldsymbol{x}]_m - [\tilde{\boldsymbol{H}}_n \boldsymbol{x}]_{m+1} \\
&\quad = \sum_{i=0}^{\infty} \tilde{h}^{[i]}(n\theta)[\boldsymbol{x}]_{m+i} - \sum_{i=0}^{\infty} \tilde{h}^{[i]}(n\theta)[\boldsymbol{x}]_{m+i+1} \\
&\quad = \sum_{i=0}^{\infty} \tilde{h}^{[i]}(n\theta)([\boldsymbol{x}]_{m+i} - [\boldsymbol{x}]_{m+i+1}) \geq 0, \quad n = 1,2,\ldots, m = 0,1,\ldots,
\end{aligned}$$

from which (i) follows. Next we consider (ii). By definition (7.12) of $\overline{\boldsymbol{B}}_n$, we have for $n = 1,2,\ldots$ and $m = 0,1,\ldots,$

$$[\overline{\boldsymbol{B}}_n \boldsymbol{x}]_{m+1} = \sum_{i=0}^{m+1} \binom{m+1}{i} \left[\frac{n}{n+1}\right]^i \left[\frac{1}{n+1}\right]^{m+1-i} \overline{g}_{m+1-i}[\boldsymbol{x}]_i. \tag{7.50}$$

We then have

$$\begin{aligned}
[\overline{\boldsymbol{B}}_n \boldsymbol{x}]_1 - [\overline{\boldsymbol{B}}_n \boldsymbol{x}]_0 &= \left(\left[\frac{1}{n+1}\right] \overline{g}_1 [\boldsymbol{x}]_0 + \left[\frac{n}{n+1}\right] \overline{g}_0 [\boldsymbol{x}]_1 \right) - \overline{g}_0 [\boldsymbol{x}]_0 \\
&= \left[\frac{1}{n+1}\right] (\overline{g}_1 - \overline{g}_0)[\boldsymbol{x}]_0 + \left[\frac{n}{n+1}\right] \overline{g}_0 ([\boldsymbol{x}]_1 - [\boldsymbol{x}]_0) \leq 0.
\end{aligned}$$

Using (7.50) and

$$\binom{m+1}{i} = \binom{m}{i} + \binom{m}{i-1}, \quad m = 1,2,\ldots, i = 1,2,\ldots,m,$$

we can obtain for $m = 1,2,\ldots,$

$$[\overline{\boldsymbol{B}}_n \boldsymbol{x}]_{m+1} = \frac{1}{n+1} \sum_{i=0}^{m} \binom{m}{i} \left[\frac{n}{n+1}\right]^i \left[\frac{1}{n+1}\right]^{m-i} \overline{g}_{m+1-i}[\boldsymbol{x}]_i$$

$$+\frac{n}{n+1}\sum_{i=0}^{m}\binom{m}{i}\left[\frac{n}{n+1}\right]^{i}\left[\frac{1}{n+1}\right]^{m-i}\overline{g}_{m-i}[\boldsymbol{x}]_{i+1}.$$

Therefore, noting

$$[\overline{\boldsymbol{B}}_n\boldsymbol{x}]_m = \left[\frac{1}{n+1}+\frac{n}{n+1}\right]\sum_{i=0}^{m}\binom{m}{i}\left[\frac{n}{n+1}\right]^{i}\left[\frac{1}{n+1}\right]^{m-i}\overline{g}_{m-i}[\boldsymbol{x}]_{i},$$
$$n=1,2,\ldots,\ m=0,1,\ldots,$$

we have for $m=1,2,\ldots,$

$$\begin{aligned}[\overline{\boldsymbol{B}}_n\boldsymbol{x}]_{m+1}-[\overline{\boldsymbol{B}}_n\boldsymbol{x}]_m &\\ = \frac{1}{n+1}\sum_{i=0}^{m}\binom{m}{i}\left[\frac{n}{n+1}\right]^{i}\left[\frac{1}{n+1}\right]^{m-i}&(\overline{g}_{m+1-i}-\overline{g}_{m-i})[\boldsymbol{x}]_i\\ +\frac{n}{n+1}\sum_{i=0}^{m}\binom{m}{i}\left[\frac{n}{n+1}\right]^{i}\left[\frac{1}{n+1}\right]^{m-i}&\overline{g}_{m-i}([\boldsymbol{x}]_{i+1}-[\boldsymbol{x}]_i).\end{aligned}$$

(ii) then follows from $\boldsymbol{g}^\star\in\mathscr{F}$, $\boldsymbol{x}\in\mathscr{F}$, and the above equation. □

7.B.2 Proof of (7.28)

Owing to (7.27), for any $n=0,1,\ldots$, $\{[\boldsymbol{a}_n]_i\}_{i=0,1\ldots}$ is a non-increase sequence bounded below by 0, so that $\lim_{i\to\infty}[\boldsymbol{a}_n]_i$ exists. In what follows, we prove that

$$\boldsymbol{e}^\top\boldsymbol{a}_n<\infty,\quad n=0,1,\ldots, \tag{7.51}$$

where $\boldsymbol{e}^\top$ denotes a $1\times\infty$ vector whose elements are all equal to one. Note that (7.51) implies (7.28) because if $\lim_{i\to\infty}[\boldsymbol{a}_n]_i>0$, $\boldsymbol{e}^\top\boldsymbol{a}_n$ diverges to infinity.

First, (7.51) holds for $n=0$ because we have from (7.7) and (7.14),

$$\boldsymbol{e}^\top\boldsymbol{a}_0=\boldsymbol{e}^\top\overline{\boldsymbol{g}}^\star=\theta\mathrm{E}[G]<\infty.$$

We then assume that (7.51) holds for some $n=m$ ($m=0,1,\ldots$). Using (7.7), (7.11), (7.12), and (7.15), we have

$$\begin{aligned}\boldsymbol{e}^\top\boldsymbol{a}_{m+1} &= \boldsymbol{e}^\top(\overline{\boldsymbol{B}}_{m+1}\tilde{\boldsymbol{H}}_{m+1}\boldsymbol{a}_m)\\ &= (\boldsymbol{e}^\top\overline{\boldsymbol{B}}_{m+1})\tilde{\boldsymbol{H}}_{m+1}\boldsymbol{a}_m\\ &\le \left(\sum_{m=0}^{\infty}\overline{g}_m\right)\boldsymbol{e}^\top\cdot\tilde{\boldsymbol{H}}_{m+1}\boldsymbol{a}_m\\ &\le \theta\mathrm{E}[G](\boldsymbol{e}^\top\tilde{\boldsymbol{H}}_{m+1})\boldsymbol{a}_m\\ &\le \theta\mathrm{E}[G]\cdot\boldsymbol{e}^\top\boldsymbol{a}_m\\ &< \infty.\end{aligned}$$

Therefore (7.51) also holds for $n=m+1$, which completes the proof. □

7.C Proof of Lemma 7.3

Because $\{\overline{g}_m\}_{m=0,1,\ldots}$ is a non-increasing sequence, we have

$$\|\boldsymbol{a}_0^{\overline{\text{error}}}\|_\infty = \|\overline{\boldsymbol{g}}^\star - \overline{\boldsymbol{g}}^{\star\overline{\text{comp}}}\|_\infty = \overline{g}_{m_{\text{g}}^*} \le \epsilon_{\text{g}}.$$

Next, we consider $n = 1,2,\ldots$. We represent $\tilde{\boldsymbol{H}}_n$ $(n = 1,2,\ldots)$ and $\overline{\boldsymbol{B}}_n$ $(n = 1,2,\ldots)$ in the following form (see Figure 7.3).

$$\begin{aligned}
\tilde{\boldsymbol{H}}_n &= \begin{pmatrix} \tilde{\boldsymbol{H}}_n^{\text{comp}} & \tilde{\boldsymbol{H}}_n^{\text{error1}} \\ \boldsymbol{O} & \tilde{\boldsymbol{H}}_n^{\text{error2}} \end{pmatrix}, \\
\overline{\boldsymbol{B}}_n &= \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{comp}} + \overline{\boldsymbol{B}}_n^{\text{error1}} & \overline{\boldsymbol{B}}_n^{\text{rest1}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} & \overline{\boldsymbol{B}}_n^{\text{rest2}} + \overline{\boldsymbol{B}}_n^{\text{error3}} \end{pmatrix}.
\end{aligned}$$

We then have

$$\begin{aligned}
\boldsymbol{a}_n &= \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \left[\boldsymbol{a}_{n-1}^{\overline{\text{comp}}} + \boldsymbol{a}_{n-1}^{\overline{\text{error}}}\right] \\
&= \overline{\boldsymbol{B}}_n \begin{pmatrix} \tilde{\boldsymbol{H}}_n^{\text{comp}} & \tilde{\boldsymbol{H}}_n^{\text{error1}} \\ \boldsymbol{O} & \tilde{\boldsymbol{H}}_n^{\text{error2}} \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_{n-1}^{\text{comp}} \\ \boldsymbol{0} \end{pmatrix} + \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{a}_{n-1}^{\overline{\text{error}}} \\
&= \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{comp}} + \overline{\boldsymbol{B}}_n^{\text{error1}} & \overline{\boldsymbol{B}}_n^{\text{rest1}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} & \overline{\boldsymbol{B}}_n^{\text{rest2}} + \overline{\boldsymbol{B}}_n^{\text{error3}} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{H}}_n^{\text{comp}} \boldsymbol{a}_{n-1}^{\text{comp}} \\ \boldsymbol{0} \end{pmatrix} \\
&\quad + \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{a}_{n-1}^{\overline{\text{error}}} \\
&= \boldsymbol{a}_n^{\overline{\text{comp}}} + \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{error1}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \boldsymbol{a}_{n-1}^{\text{comp}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \boldsymbol{a}_{n-1}^{\text{comp}} \end{pmatrix} + \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{a}_{n-1}^{\overline{\text{error}}}. \qquad (7.52)
\end{aligned}$$

Note here that in exactly the same way as in the proof of Lemma 7.2, we can show that

$$[\boldsymbol{a}_n^{\overline{\text{comp}}}]_i \ge [\boldsymbol{a}_n^{\overline{\text{comp}}}]_{i+1}, \quad i = 0,1,\ldots. \qquad (7.53)$$

Therefore, from (7.52) and (7.53), we obtain

$$\begin{aligned}
\|\boldsymbol{a}_n^{\overline{\text{error}}}\|_\infty &= \left\| \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{error1}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \boldsymbol{a}_{n-1}^{\text{comp}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \boldsymbol{a}_{n-1}^{\text{comp}} \end{pmatrix} + \overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{a}_{n-1}^{\overline{\text{error}}} \right\|_\infty \\
&\le \left\| \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{error1}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} \tilde{\boldsymbol{H}}_n^{\text{comp}} \end{pmatrix} \boldsymbol{e} \right\|_\infty \cdot [\boldsymbol{a}_{n-1}^{\text{comp}}]_0 + \|\overline{\boldsymbol{B}}_n \tilde{\boldsymbol{H}}_n \boldsymbol{e}\|_\infty \cdot \|\boldsymbol{a}_{n-1}^{\overline{\text{error}}}\|_\infty \\
&\le \left\| \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{error1}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} \end{pmatrix} \boldsymbol{e} \right\|_\infty \cdot [\boldsymbol{a}_{n-1}^{\text{comp}}]_0 + \|\boldsymbol{a}_{n-1}^{\overline{\text{error}}}\|_\infty.
\end{aligned}$$

By definition, $\overline{\boldsymbol{B}}_n^{\text{error1}}\boldsymbol{e}$ denotes an $\{(n+1)(m_{\text{g}}^*-1)+1\} \times 1$ vector whose i-th ($i = 0,1,\dots,(n+1)(m_{\text{g}}^*-1)$) element is given by

$$[\overline{\boldsymbol{B}}_n^{\text{error1}}\boldsymbol{e}]_i = \begin{cases} 0, & i = 0,1,\dots,m_{\text{g}}^*-1, \\ \displaystyle\sum_{k=m_{\text{g}}^*}^{i} b_n(i,i-k)\overline{g}_k, & i = m_{\text{g}}^*, m_{\text{g}}^*+1,\dots,(n+1)(m_{\text{g}}^*-1), \end{cases}$$

and $\overline{\boldsymbol{B}}_n^{\text{error2}}\boldsymbol{e}$ is an $\infty \times 1$ vector whose i-th ($i = 0,1,\dots$) element is given by

$$\begin{aligned}[\overline{\boldsymbol{B}}_n^{\text{error2}}\boldsymbol{e}]_i &= \sum_{k=m_{\text{g}}^*+i}^{(n+1)(m_{\text{g}}^*-1)+i+1} b_n\Big((n+1)(m_{\text{g}}^*-1)+i+1, \\ &\qquad\qquad (n+1)(m_{\text{g}}^*-1)+i-k+1\Big)\overline{g}_k.\end{aligned}$$

Therefore we have

$$\begin{aligned}\left\| \begin{pmatrix} \overline{\boldsymbol{B}}_n^{\text{error1}} \\ \overline{\boldsymbol{B}}_n^{\text{error2}} \end{pmatrix} \boldsymbol{e} \right\|_\infty &\le \sum_{k=m_{\text{g}}^*}^{\infty} 1 \cdot \overline{g}_k \\ &\le \epsilon_g,\end{aligned}$$

which completes the proof. □

7.D Computation of $\tilde{h}^{[m]}(\zeta)$

In this appendix, we provide computational methods of $\tilde{h}^{[i]}(\zeta)$ ($\zeta > 0$) for three types of service time distributions: (i) constant, (ii) phase-type distribution, and (iii) Pareto distribution. In the case of constant service times, we have

$$\begin{aligned}\tilde{h}^{[m]}(\zeta) &= \frac{1}{\mathrm{E}[H]}\int_0^{\mathrm{E}[H]} \frac{\exp[-\zeta x](\zeta x)^m}{m!}dx \\ &= \frac{1}{\zeta \mathrm{E}[H]}\left[1 - \sum_{i=0}^{m} \frac{\exp[-\zeta \mathrm{E}[H]](\zeta \mathrm{E}[H])^i}{i!}\right], \quad m = 0,1,\dots.\end{aligned}$$

Next we consider the case of phase-type service times with PDF $H(x) = 1 - \boldsymbol{\beta}\exp[\boldsymbol{S}x]\boldsymbol{e}$ ($x \ge 0$). Note that

$$\tilde{h}(x) = \boldsymbol{\pi}\exp[\boldsymbol{S}x](-\boldsymbol{S})\boldsymbol{e}, \quad x \ge 0,$$

where

$$\boldsymbol{\pi} = \frac{\boldsymbol{\beta}(-\boldsymbol{S})^{-1}}{\boldsymbol{\beta}(-\boldsymbol{S})^{-1}\boldsymbol{e}}.$$

We thus have

$$\begin{aligned}\tilde{h}^{[m]}(\zeta) &= \int_0^\infty \frac{\exp[-\zeta x](\zeta x)^m}{m!} \boldsymbol{\pi} \exp[\boldsymbol{S}x](-\boldsymbol{S})\boldsymbol{e}dx \\ &= \frac{1}{\zeta} \cdot \boldsymbol{\pi} \left\{ \left[\boldsymbol{I} - \zeta^{-1}\boldsymbol{S} \right]^{-1} \right\}^{m+1} (-\boldsymbol{S})\boldsymbol{e}.\end{aligned}$$

Therefore we can compute $\tilde{h}^{[m]}(\zeta)$ ($m = 0, 1, \ldots$) by $\tilde{h}^{[m]}(\zeta) = \boldsymbol{\pi}\boldsymbol{y}_m$, where

$$\boldsymbol{y}_0 = \frac{1}{\zeta} \left[\boldsymbol{I} - \zeta^{-1}\boldsymbol{S} \right]^{-1} (-\boldsymbol{S})\boldsymbol{e}, \quad \boldsymbol{y}_m = \left[\boldsymbol{I} - \zeta^{-1}\boldsymbol{S} \right]^{-1} \boldsymbol{y}_{m-1}, \quad m = 1, 2, \ldots.$$

Finally, we consider the case that service times are i.i.d. according to a Pareto distribution with shape parameter γ ($\gamma > 1$) and location parameter $x_{\min}$ ($x_{\min} > 0$).

$$H(x) = \begin{cases} 0, & 0 \le x < x_{\min}, \\ 1 - \left(\dfrac{x_{\min}}{x} \right)^\gamma, & x \ge x_{\min}. \end{cases}$$

We then have

$$\tilde{h}(x) = \begin{cases} \dfrac{1}{\mathrm{E}[H]}, & 0 \le x < x_{\min}, \\ \dfrac{1}{\mathrm{E}[H]} \left(\dfrac{x_{\min}}{x} \right)^\gamma, & x \ge x_{\min}, \end{cases}$$

where $\mathrm{E}[H] = \gamma x_{\min}/(\gamma - 1)$. It then follows that

$$\begin{aligned}\tilde{h}^{[m]}(\zeta) &= \int_0^{x_{\min}} \frac{\exp[-\zeta x](\zeta x)^m}{m!} \cdot \frac{dx}{\mathrm{E}[H]} + \int_{x_{\min}}^\infty \exp[-\zeta x] \frac{(\zeta x)^m}{m!} \cdot \frac{1}{\mathrm{E}[H]} \left(\frac{x_{\min}}{x} \right)^\gamma dx \\ &= \frac{1}{\zeta \mathrm{E}[H]} \left[1 - \sum_{i=0}^m \exp[-\zeta x_{\min}] \frac{(\zeta x_{\min})^i}{i!} \right] \\ &\qquad + \frac{x_{\min}(\zeta x_{\min})^{\gamma-1}}{\mathrm{E}[H]m!} \int_{\zeta x_{\min}}^\infty \exp[-x] x^{m-\gamma} dx. \qquad (7.54)\end{aligned}$$

The integral on the right hand side of (7.54) is an incomplete gamma function, which can be computed with high accuracy by means of numerical libraries (see e.g., [GSL]).

7.E Error bounds ΔP_{loss} for numerical examples

Figures 7.15–7.17 show ΔP_{loss} in some numerical examples in Section 7.3, where $\Delta P_{\text{loss}} \ge 10^{-9}$. Note that ΔP_{loss} in all other examples is less than 10^{-9}.

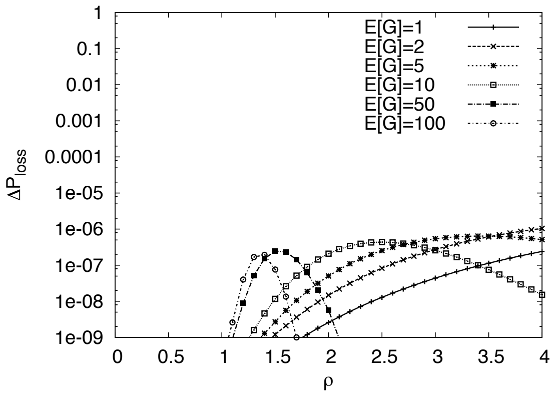

Figure 7.15: ΔP_{loss} in the M/D/1+H$_2$ queue with Cv[G] = 3.0, corresponding to Figure 7.11.

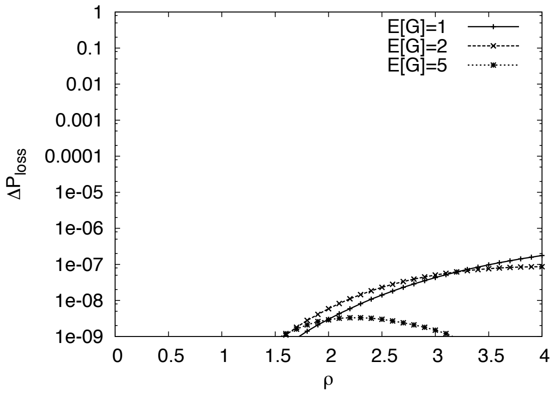

Figure 7.16: ΔP_{loss} in the M/D/1+Er$_{k,k+1}$ queue with Cv[G] = 0.3, corresponding to Figure 7.12.

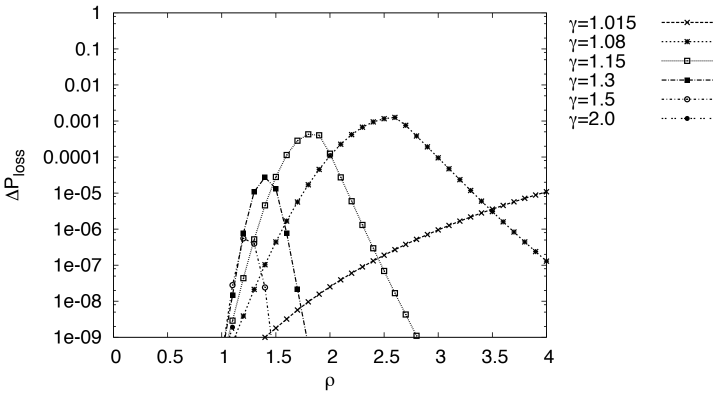

Figure 7.17: ΔP_{loss} in the M/Pareto/1+Er$_{k,k+1}$ queue with E[G] = 100 and Cv[G] = 0.3, corresponding to Figure 7.13.

8 Conclusion

In this dissertation, we developed analytical methods for two kinds of queueing models interacting with underlying processes, which are regarded as fundamental models for communication systems with adaptive resource allocation mechanisms. In the first model, the state of the underlying process is assumed to be switched when the system becomes empty. On the other hand, in the second model, the state of the underlying process is assumed to change continually according to the workload in system.

We analyzed the multi-class FCFS M/G/1 queue with exponential working vacations in Chapter 2. This model is a special case of our fundamental model of the first kind, where there exists only two underlying states. We developed an analytical method of this model based on an approach that we first analyze the workload process and then derive other performance measures using the workload distribution. With this approach, we derived various performance measures including the waiting time and sojourn time distributions, the joint distribution of the queue lengths and the workload in system in respective classes, and the joint transform associated to the busy cycle.

Using results in Chapter 2, we can see the difference between the queueing model with working vacations and the queueing model embedded in a random environment, where the processing rate is assumed to change according to an independent underlying process. To be more specific, consider a special case of our model where the processing rate is proportional to the arrival rate, and compare it to the corresponding queue embedded in a random environment of a two-state Markov chain. In the latter model, the stationary number of customers in the system is independent of the underlying Markov chain and its conditional distribution given a specific state of the Markov chain is the same as that of the ordinary M/G/1 queue (Section 6 in [Tak05]). On the other hand, it is verified that the model we considered does not have such a property. Therefore, the queueing model with working vacation (and therefore queuing models interacting with underlying processes) is essentially different from the queueing model embedded in a random environmental process.

In Chapter 3, we considered the multi-class MAP/G/1 queue with disasters, which corresponds to a censored process of the multi-class MAP/G/1 queue with working vacations obtained by observing only working vacation periods. We generalized the approach based on an analysis of the workload process taken in Chapter 2, and developed computational algorithms for the joint queue length distribution and the moments of the waiting time and sojourn time distributions, which are summarized in Figure 3.1.

In Chapter 4, we extended analytical methods for the M/G/1-type Markov process with respect to the irreducibility of the underlying process. By allowing $\boldsymbol{C}+\boldsymbol{D}$ to be reducible, this Markov process become applicable to a wider class of queueing models including our fundamental model of the first kind. We first proved that previously known results for the case of irreducible $\boldsymbol{C}+\boldsymbol{D}$ cannot be directly applied to that of reducible $\boldsymbol{C}+\boldsymbol{D}$. We then derived formulas for the stationary distribution applicable to reducible $\boldsymbol{C}+\boldsymbol{D}$, and further developed an efficient computational algorithm for the fundamental matrix and the moments of the stationary distribution.

In Chapter 5, we considered the M/G/1 queue with general impatient customers (M/G/1+G), which is equivalent to our fundamental model of the second kind. We revisited the formal series solution of the p.d.f. of the stationary workload in system, and provided an probabilistic interpretation to it through an analysis of the LCFS-PR M/G/1 queue with workload dependent loss. As demonstrated for the M/G/1+M, M/G/1+D, and M/M/1+G queues, this new perspective leads to a unified understanding of special cases of the M/G/1+G queue.

In Chapter 6, we analyzed the stationary loss probability in the M/G/1+G queue. We derived theoretical upper and lower bounds of the loss probability and some stochastic ordering relations, based the results in Chapter 5. To prove stochastic ordering relations, we derived new results for the excess wealth and dispersive orders, which seem to be basic results although we could not find them in the literature. We also proved that the loss probability in the M/D/1+D queue is smallest among all M/G/1+G queues with the same and finite arrival rate, mean service time, and mean impatience time.

Finally in Chapter 7, we developed a computational algorithm for the loss probability in the M/G/1+PH queue, which is summarized in Figure 7.4. Our algorithm is based on the uniformization technique and the results in Chapters 5 and 6, which can efficiently compute the loss probability along with an upper bound of numerical error. Using this computational algorithm, we can examine the impact of the model parameters on the loss probability, which is the most important quantity of interest in our fundamental model of the second kind.

This dissertation separately analyzed the two kinds of queueing models interacting with underlying processes so that their mathematical structures to be well understood. In general, communication systems may be equipped with multiple adaptive resource allocation mechanisms, which leads to a more complicated situation that these two kinds of models are blended together. An analysis of such extended model is a challenging problem, and it is worth investigating as a future work.

Bibliography

[AG62] Ancker, C. J. and Gafarian, A. V. Queueing with impatient customers who leave at random. *The Journal of Industrial Engineering*, **13** (1962), 84–90.

[AK93] Asmussen, S. and Koole, G. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, **30**, (1993), 365–372.

[BH81] Baccelli, F. and Hebuterne, G. On queues with impatient customers. *Proceedings of Performance '81* (Kylstra, F. J. ed.), North-Holland, Amsterdam, 159–179, 1981.

[BBH84] Baccelli, F., Boyer, P., and Hebuterne, G. Single-server queues with impatient customers. *Advances in Applied Probability*, **16** (1984), 887–905.

[BKL01] Bae, J., Kim, S., and Lee, E. Y. The virtual waiting time of the M/G/1 queue with impatient customers. *Queueing Systems*, **38** (2001), 485–494.

[Bae13] Bae, J. The virtual waiting time of the M/G/1 queue with impatient customers of n times of impatience. *Journal of the Korean Statistical Society*, **42** (2013), 375–385.

[Bal03] Balter, M. H., Li, C., Osogami, T., Scheller-Wolf, A., and Squillante, M. S. Analysis of task assignment with cycle stealing under central queue. *Proceedings of the 23rd International Conference on Distributed Computing Systems*, 628–637, 2003.

[Bar57] Barrer, D. Y. Queueing with impatient customers and ordered service. *Operations Research*, **5** (1957), 650–656.

[BS06] Bartoszewicz, J. and Skolimowska M. Preservation of classes of life distributions and stochastic orders under weighting. *Statistics & Probability Letters*, **76** (2006), 587–596.

[Bar86] Bartoszewicz, J. Dispersive ordering and the total time on test transformation. *Statistics & Probability Letter*, **4** (1986), 285–288.

[BS12] Baumann, H. and Sandmann, W. Numerical solution of level dependent quasi-birth-and-death processes. *Procedia Computer Science*, **1** (2012), 1561–1569.

[BPS11] Boxma, O., Perry, D., and Stadje, W. The M/G/1+G queue revisited. *Queueing Systems*, **67** (2011), 207–220.

[BPSZ10] Boxma, O., Perry, D., Stadje, W., and Zacks, S. The busy period of an M/G/1 queue with customer impatience. *Journal of Applied Probability*, **47** (2010), 130–145.

[BT03] Boxma, O. J. and Takine, T. The M/G/1 FIFO queue with several customer classes. *Queueing Systems*, **45** (2003), 185–189.

[BB99] Brandt, A. and Brandt, M. On the M(n)/M(n)/s queue with impatient calls. *Performance Evaluation*, **35** (1999), 1–18.

[BB13] Brandt, A. and Brandt, M. Workload and busy period for M/GI/1 with a general impatience mechanism. *Queueing Systems*, **75** (2013), 189–209.

[BP77] Brill, P. H. and Posner, M. J. M. Level crossings in point processes applied to queues: Single-server case. *Operations Research*, **25** (1977), 662–674.

[Cin75] Çinlar, E. *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, 1975.

[Coh69] Cohen, J. W. Single server queues with restricted accessibility. *Journal of Engineering Mathematics*, **3** (1969), 265–284.

[Coh77] Cohen, J. W. On up- and downcrossing. *Journal of Applied Probability*, **14** (1977), 405–410.

[Coh82] Cohen, J. W. *The Single-Server Queue*, North-Holland, Amsterdam, 1982.

[Dos90] Doshi, B. Conditional and unconditional distributions for M/G/1 type queues with server vacations. *Queueing Systems*, **7** (1990), 229–252.

[Gan59] Gantmacher, F. R. *The Theory of Matrices, Vol. 2*, Chelsea, New York, 1959.

[Dal65] Daley, D. J. General customer impatience in the queue GI/G/1. *Journal of Applied Probability*, **2** (1965), 186–205.

[DR88] van Doorn, E. A. and Regterschot, G. J. K. Conditional PASTA. *Operations Research Letters*, **7** (1988), 229–232.

[DN99] Dudin, A. and Nishimura, S. A BMAP/SM/1 queueing system with Markovian arrival input of disasters. *Journal of Applied Probability*, **36** (1999), 868–881.

[DS04] Dudin, A. and Semenova, O. A stable algorithm for stationary distribution calculation for a BMAP/SM/1 queueing system with Markovian arrival input of disasters. *Journal of Applied Probability*, **41** (2004), 547–556.

[He96] He, Q. M. Queues with marked customers. *Advances in Applied Probability*, **28** (1996), 567–587.

[HS80] Heyman, D. P. and Stidham, S., Jr., The relation between customer and time averages in queues. *Operations Research*, **28** (1980), 983–994.

[JS96] Jain, G. and Sigman, K. A Pollaczek-Khintchine formula for M/G/1 queues with disasters. *Journal of Applied Probability*, **33** (1996), 1191–1200.

[KS88] Keilson, J. and Servi, L. D. A distributional form of Little's law. *Operations Research Letters*, **7** (1988), 223–227.

[Kel79] Kelly, F. P. *Reversibility and Stochastic Networks*, John Wiley & Sons, Chichester, NY, 1979.

[KBL01] Kim, S., Bae, J., and Lee, E. Y. Busy periods of Poisson arrival queues with loss. *Queueing Systems*, **39** (2001), 201–212.

[KCC03] Kim, J. D., Choi, D. W., and Chae, K. C. Analysis of queue-length distribution of the M/G/1 queue with working vacations (M/G/1/WV). *Proceedings of Hawaii International Conference on Statistics and Related Fields*, 1191–1200, 2003.

[KK13] Kim, J. and Kim, J. M/PH/1 queue with deterministic impatience time. *Communications of the Korean Mathematical Society*, **28** (2013), 383–396.

[Kle75] L. Kleinrock. *Queueing Systems, Volume I: Theory*, John Wiley & Sons, New York, 1975.

[Kov61] Kovalenko, I. N. Some queueing problems with restrictions. *Theory of Probability and Its Applications*, **6** (1961), 205–208.

[GSL] Galassi, M. et al. *GNU Scientific Library Reference Manual (3rd Ed.)*, `http://www.gnu.org/software/gsl/`.

[LTZL09] Li, J.-H., Tian, N.-S., Zhang, A. G., and Luh, H. P. Analysis of the M/G/1 queue with exponentially working vacations – A matrix analytic approach. *Queueing Systems*, **61** (2009), 139–166.

[LK08] Liu, L. and Kulkarni, V. G. Busy period analysis for M/PH/1 queues with workload dependent balking. *Queueing Systems*, **59** (2008), 37–51.

[LXT07] Liu, W.-Y., Xu, X.-L, and Tian, N.-S. Stochastic decompositions in the M/M/1 queue with working vacations. *Operations Research Letters*, **35** (2007), 595–600.

[LMN90] Lucantoni, D. M., Meier-Hellstern, K. S., and Neuts, M. F. A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability*, **22** (1990), 676–705.

[MT03] Masuyama, H. and Takine, T. Analysis and computation of the joint queue length distribution in a FIFO single-server queue with multiple batch Markovian arrival streams. *Stochastic Models*, **19** (2003), 349–381.

[Mas15] Masuyama, H. Error bounds for augmented truncations of discrete-time block-monotone Markov chains under geometric drift conditions. *Advances in Applied Probability*, **47** (2015), 83–105.

[Neu89] Neuts, M. F. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, 1989.

[PA95] Perry, D. and Asmussen, S. Rejection rules in the M/G/1 queue. *Queueing Systems*, **19** (1995), 105–130.

[PMKT10] Phung-Duc, T., Masuyama, H., Kasahara, S., and Takahashi, Y. A simple algorithm for the rate matrices of level-dependent QBD processes. *In proceedings of the 5th International Conference on Queueing Theory and Network Applications*, 2010.

[PSZ00] Perry, D., Stadje, W., and Zacks S. Busy period analysis for M/G/1 and G/M/1 type queues with restricted accessibility. *Operations Research Letters*, **27** (2000), 163–174.

[Ram88] Ramaswami, V. A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Stochastic Models*, **4** (1988), 183–188.

[Rao67] Rao, S. S. Queueing with balking and reneging in M/G/1 systems. *Metrika*, **12** (1967), 173–188.

[Sen89] Sengupta, B. An invariance relationship for the G/G/1 queue. *Advances in Applied Probability*, **21** (1989), 956–957.

[SF02] Servi, L. D. and Finn, S. G. M/M/1 queues with working vacations (M/M/1/WV). *Performance Evaluation*, **50** (2002), 41–52.

[SS07] Shaked, M. and Shanthikumar, J. G. *Stochastic Order*, Springer, New York, NY, 2007.

[Shi04] Shin, Y. W. BMAP/G/1 queue with correlated arrivals of customers and disasters. *Operations Research Letters*, **32** (2004), 364–373.

[Sta79] Stanford, R. E. Reneging phenomena in single channel queues. *Mathematics of Operations Research*, **4** (1979), 162–178.

[Sta90] Stanford, R. E. On queues with impatience. *Advances in Applied Probability*, **22** (1990), 768–769.

[Tak96] Takine, T. A continuous version of matrix-analytic methods with the skip-free to the left property. *Stochastic Models*, **12** (1996), 673–682.

[Tak01] Takine, T. Queue length distribution in a FIFO single-sever queue with multiple arrival streams having different service time distributions. *Queueing Systems*, **39** (2001), 349–375.

[Tak02] Takine, T. Matrix product-form solution for an LCFS-PR single-server queue with multiple arrival streams governed by a Markov chain. *Queueing Systems*, **42** (2002), 131–151.

[Tak05] Takine, T. Single-server queues with Markov-modulated arrivals and service speed. *Queueing Systems*, **49** (2005), 7–22.

[TH92] Takine, T. and Hasegawa, T. A generalization of the decomposition property in the M/G/1 queue with server vacations. *Operations Research Letters*, **12** (1992), 97–99.

[TH94] Takine, T. and Hasegawa, T. The workload in the MAP/G/1 queue with state-dependent services: Its application to a queue with preemptive resume priority. *Stochastic Models*, **10** (1994), 183–204.

[TMSH94] Takine, T., Matsumoto, Y., Suda, T., and Hasegawa, T. Mean waiting times in nonpreemptive priority queues with Markovian arrival and i.i.d. service processes. *Performance Evaluation*, **20** (1994), 131–149.

[Tij94] Tijms, H. C. *Stochastic Models, An Algorithmic Approach*, Wiley, Chichester, 1994.

[VVB06] van Velthoven, J., van Houdt, B., and Blondia, C. On the probability of abandonment in queues with limited sojourn and waiting times. *Operations Research Letters*, **34** (2006) 333–338.

[WLJ10] Wang, K.; Li, N.; Jiang, Z. Queueing system with impatient customers: A review. *Proceedings of IEEE International Conference on Service Operations and Logistics and Informatics*, 82–97, 2010, [Wol82] Wolff, R. W. Poisson arrivals see time averages. *Operations Research*, **30** (1982) 223–231.

[WT06] Wu, D.-A. and Takagi, H. M/G/1 queue with multiple working vacations. *Performance Evaluation*, **63** (2006), 654–681.

[YKC02] Yang, W. S., Kim, J. D., and Chae, K. C. Analysis of M/G/1 stochastic clearing systems. *Stochastic Analysis and Applications*, **20** (2002), 1083–1100.

Publication List

A. Journal Papers

1. Y. Inoue and T. Takine, The multi-class FIFO M/G/1 queue with exponential working vacations. *Journal of the Operations Research Society of Japan*, **56** (2013), 111–136.

2. Y. Inoue and T. Takine, The FIFO single-server queue with disasters and multiple Markovian arrival streams. *Journal of Industrial and Management Optimization*, **10** (2014), 57–87.

3. Y. Inoue and T. Takine, Analysis of the loss probability in the M/G/1+G queue. *Queueing Systems*, **80** (2015), 363–386.

4. Y. Inoue and T. Takine, An extension of the matrix-analytic method for M/G/1-type Markov processes. *Journal of the Operations Research Society of Japan*, **58** (2015), 376–393.

5. Y. Inoue and T. Takine, The M/D/1+D queue has the minimum loss probability among M/G/1+G queues. *Operations Research Letters*, **43** (2015), 629–632.

B. International Conferences (refereed)

1 Y. Inoue and T. Takine, The workload distribution in a MAP/G/1 queue with disasters. Presented at the 7th International Conference on Queueing Theory and Network Applications (QTNA2012), 8 pages, Kyoto, Japan, August 2012 (**Best Paper Award**).

C. Domestic Conferences (non-refereed, written in Japanese)

1 Y. Inoue and T. Takine, The M/G/1 queue with working vacations. Presented at The Queueing Symposium 2011 : Stochastic Models and Their Applications, Kyoto, Japan, January 2011.

2 Y. Inoue and T. Takine, The M^{X}/G/1 queue with working vacations. Presented at The 2011 Autumn National Conference of The Operations Research Society of Japan, Kobe, Japan, September 2011.

3 Y. Inoue and T. Takine, The workload distribution in the MAP/G/1 queue with disasters. Presented at The Queueing Symposium 2012 : Stochastic Models and Their Applications, Hamamatsu, Japan, January 2012 (**Research Encouragement Award**).

4 Y. Inoue and T. Takine, The joint queue length distribution in the multi-class MAP/G/1 queue with disasters. Presented at The 2012 Autumn National Conference of The Operations Research Society of Japan, Nagoya, Japan, September 2012.

5 Y. Inoue and T. Takine, Queueing models with multiple Markovian customer arrival streams. Presented at Seminar for Young Researchers in Operations Research of Japan, Hikone, Japan, October 2012.

6 Y. Inoue and T. Takine, The workload distribution in the MAP/G/1 queue with working vacations. Presented at The Queueing Symposium 2013 : Stochastic Models and Their Applications, Nagasaki, Japan, January 2013.

7 Y. Inoue and T. Takine, An extension of matrix analytic methods for bivariate Markov processes with the skip-free to the left property. Presented at The 2013 Autumn National Conference of The Operations Research Society of Japan, Tokushima, Japan, September 2013.

8 Y. Inoue and T. Takine, On the virtual waiting time distribution in the M/G/1+G queue. Presented at The Symposium on Stochastic Models 2014, Tokyo, Japan, January 2014 (**Research Encouragement Award**).

9 Y. Inoue and T. Takine, The virtual waiting time distribution in the multi-class M/G/1+G queue. Presented at The 245-th Workshop of Special Interest Group of Queueing Theory, The Operations Research Society of Japan, Tokyo, Japan, February 2014.

10 Y. Inoue and T. Takine, On numerical computation of the loss probability in the M/G/1+G queue. Presented at The 2014 Autumn National Conference of The Operations Research Society of Japan, Sapporo, Japan, August 2014.

11 Y. Inoue and T. Takine, Analysis and computation of the loss probability in the M/G/1+G queue. Presented at The Symposium on Stochastic Models 2015, Sendai, Japan, January 2015. (**Research Encouragement Award**)