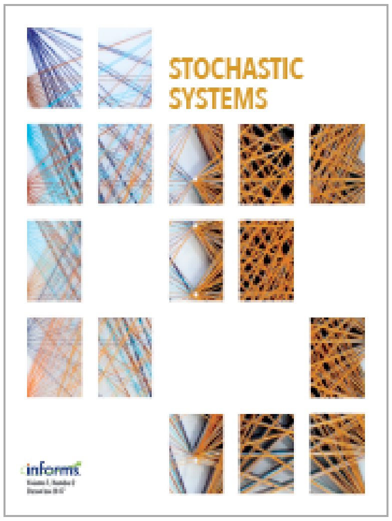


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The BAR Approach for Multiclass Queueing Networks with SBP Service Policies

Anton Braverman; , J. G. Dai; , Masakiyo Miyazawa

https://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 15, No. 1, March 2025, pp. 1–49
ISSN 1946-5238 (online)

The BAR Approach for Multiclass Queueing Networks with SBP Service Policies

Anton Braverman,[a,*] J. G. Dai,[b] Masakiyo Miyazawa[c]

[a] Kellogg School of Management, Northwestern University, Evanston, Illlinois 60208; [b] School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853; [c] School of Data Science, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China

*Corresponding author

Contact: anton.braverman@kellogg.northwestern.edu, https://orcid.org/0000-0003-4030-3172 (AB); jd694@cornell.edu, https://orcid.org/0000-0002-5223-0129 (JGD); masakiyo.miyazawa@gmail.com (MM)

Received: February 11, 2023
Revised: January 7, 2024
Accepted: March 20, 2024
Published Online in Articles in Advance: May 2, 2024

https://doi.org/10.1287/stsy.2023.0011

Abstract. The basic adjoint relationship (BAR) approach is an analysis technique based on the stationary equation of a Markov process. This approach was introduced to study heavy-traffic, steady-state convergence of generalized Jackson networks in which each service station has a single job class. We extend it to multiclass queueing networks operating under static-buffer-priority (SBP) service disciplines. Our extension makes a connection with Palm distributions that allows one to attack a difficulty arising from queue-length truncation, which appears to be unavoidable in the multiclass setting. For multiclass queueing networks operating under SBP service disciplines, our BAR approach provides an alternative to the "interchange of limits" approach that has dominated the literature in the last twenty years. The BAR approach can produce sharp results and allows one to establish steady-state convergence under three additional conditions: stability, state space collapse (SSC) and a certain matrix being "tight." These three conditions do not appear to depend on the interarrival and service-time distributions beyond their means, and their verification can be studied as three separate modules. In particular, they can be studied in a simpler, continuous-time Markov chain setting when all distributions are exponential. As an example, these three conditions are shown to hold in reentrant lines operating under last-buffer-first-serve discipline. In a two-station, five-class reentrant line, under the heavy-traffic condition, the tight-matrix condition implies both the stability condition and the SSC condition. Whether such a relationship holds generally is an open problem.

Keywords: piecewise deterministic Markov process • stationary distribution • heavy traffic approximation

1. Introduction

In this paper, we prove that the stationary distribution of a multiclass queueing network converges to the stationary distribution of a semimartingale reflecting Brownian motion (SRBM) in heavy traffic or as the load at each service station becomes "critical," where it is assumed that the network operates under a static-buffer-priority (SBP) service discipline (see Section 3 for its definition). For this proof, we extend the basic adjoint relationship (BAR) approach developed in Miyazawa (2017) and Braverman et al. (2017) and was coined in Harrison and Williams (1987) in the setting of charactering the stationary distribution of an SRBM. The main result of this paper is Theorem 5.1, which assumes three additional conditions: stability, state space collapse, and a certain matrix being "tight" (see Definition 4.2). As of now, it is difficult to characterize when each of these conditions holds in a general setting, but it is known that there are various examples, like reentrant lines under the last-buffer-first-server (LBFS) service discipline, that satisfy them. For a more gradual introduction to the machinery behind Theorem 5.1, we also work through a pilot example of a two-station, five-class reentrant line in Section 2. In what follows, we first introduce the background for our work, then explain the features of the BAR approach, and finally summarize the contributions of this paper.

The subject of this study is Brownian models for multiclass queueing networks. These Brownian models were introduced in Harrison (1988). Multiclass queueing networks were studied in classical papers such as Baskett et al.

(1975) and Kelly (1975). In these classical papers, the queueing networks are modeled as continuous-time Markov chains (CTMCs) with *discrete* state spaces. These CTMCs are shown to have "product-form" stationary distributions. Fueled by applications in computer systems and communications networks, product-form research was a dominant theme for more than two decades. Serfozo (1999) provides a summary of this line of research at the end of 1990s. Harrison's multiclass queueing networks have general interarrival and service-time distributions and accommodate arbitrary service disciplines. These queueing networks can be modeled as piecewise deterministic Markov processes that were formally introduced in Davis (1984). These continuous-time Markov processes have components with continuous state spaces. Obtaining the stationary distributions of these Markov processes, whether analytically or numerically, is often difficult. This difficulty motivates the study of Brownian models, which are often represented by SRBMs.

In addition to introducing multiclass queueing networks that model real-world systems, Harrison (1988) introduced Brownian system models that serve as alternative models of the same real-world systems. Since the publication of Harrison (1988), many papers proving that a certain "state" process of a multiclass queueing network converges in distribution to the corresponding process of the Brownian system model in heavy traffic have appeared (Bramson 1998; Williams 1998; Chen and Zhang 2000a, b; Chen and Ye 2001). These extend the pioneering works of Reiman (1984) and Johnson (1983) that prove a heavy-traffic limit theorem for generalized Jackson networks, a special class of queueing networks in which each service station has a single job class. These limit theorems are of the type of functional central limit theorems that approximate the dynamics of a queueing network by the dynamics of its Brownian counterpart, but they are silent on steady-state convergence: whether the stationary distribution of a multiclass queueing network converges to that of a Brownian model.

Gurvich (2014) proved steady-state convergence for multiclass queueing networks operating under a class of queue-ratio service disciplines that include SBP disciplines as special cases. This work was inspired by the pioneering paper of Gamarnik and Zeevi (2006) that proved steady-state convergence for generalized Jackson networks. Ye and Yao (2016, 2018) went further by (a) relaxing the conditions in Gurvich (2014) and, more importantly, (b) covering a wider class of service disciplines. Ye and Yao (2018) represents the state of the art in results for steady-state convergence of multiclass queueing networks. All these works proved the "interchange of limits" by using and extending the sophisticated "hydro-dynamic limits" methodology introduced in Bramson (1998) for process convergence, establishing rigorously that process convergence in functional central limit theorems is robust enough to carry over to steady-state convergence. Since Gamarnik and Zeevi (2006), interchange of limits has been proved for many other stochastic models; see the discussion on page 147 of Braverman et al. (2017), including the relevance of using Stein's method to study steady-state convergence.

This paper proves steady-state convergence for multiclass queueing networks directly, without working with the dynamics of either the prelimit or limit process. The logic for this possibility is simple: the generator of a Markov process, when well defined, governs both the dynamics and the stationary distribution of the Markov process. By working with the generator, one does not need to use the dynamics of a Markov process to understand its steady state. However, for a piecewise deterministic Markov process, the test functions in the domain of the generator need to satisfy a so-called boundary condition. For the generalized Jackson networks studied in Braverman et al. (2017), the test functions of interest are in the domain of the generator and the corresponding BAR does not have any boundary terms. Taking advantage of this fact, the authors were able develop the BAR approach to reproduce the Gamarnik and Zeevi (2006) result under a weaker condition. In the multiclass queueing networks considered in this paper, one needs to truncate the queue length terms in the test functions. As a result, they are no longer in the domain of the generator. The BAR in multiclass queueing networks involves boundary terms through Palm distributions, which are generated by counting processes of the jumps (see Section 6.1). A key step in our proof of Theorem 5.1 is to show, using Palm measures, that those boundary terms are negligible in heavy traffic, and the asymptotic BAR similar to the one in Braverman et al. (2017) still holds.

The BAR approach promotes modularity. It separates the stability and steady-state state space collapse (SSC) results from steady-state convergence. The stability of multiclass queueing networks has been extensively studied in the literature (Dai 1995, Chen and Zhang 1997). Sufficient conditions for steady-state state space collapse in multiclass queueing networks were established in Cao et al. (2022), and the conditions were verified to hold in that paper for reentrant lines under the first-buffer-first-serve discipline (FBFS) and LBFS service discipline. Steady-state SSC was proved for a bandwidth-sharing network in Wang et al. (2022).

The multifold contributions of this paper are summarized here.

- The BAR approach has been demonstrated to be a natural approach to proving heavy-traffic, steady-state convergence, as opposed to the limit-interchange approach widely used in the literature.
    (a) It makes the heavy-traffic, steady-state analysis essentially not sensitive to the distributions of interarrival and service times, thus allowing a researcher to start the analysis in a CTMC setting.

(b) It can produce the sharpest results with minimal moment conditions; our approach assumes the existence of the $(2+\delta_0)$th moments of interarrival and service times, where Ye and Yao (2018) requires the seventh moment.

(c) It was successfully used in Dai et al. (2023) to establish asymptotic steady-state independence for generalized Jackson networks in multiscale heavy traffic. It is unclear how the "limit-interchange" approach in Gurvich (2014) and Ye and Yao (2018) can be extended to the multiscale setting.

- The BAR approach developed in this paper goes significantly beyond the restrictive version in Braverman et al. (2017).

(a) It takes care of both the queue-length truncation and interarrival and service-time truncation that will likely be encountered in many other stochastic processing networks.

(b) It connects with Palm distributions in a way that was not explored in Braverman et al. (2017); see Lemmas 6.4 and 8.5 in this paper. Guang et al. (2024) has already made critical use of this Palm connection.

In the discrete-time setting, the BAR approach has been studied extensively in the literature. For example, Eryilmaz and Srikant (2012) and Maguluri and Srikant (2016) used carefully engineered polynomial functions as test functions to get asymptotically tight bounds on the steady-state moments. Characterizing all moments allowed Eryilmaz and Srikant (2012) to also establish steady-state convergence to a limiting distribution. They coined the term "drift method" for their approach. By using a family of exponential test functions (closely related to the ones used in our paper), Hurtado-Lange and Maguluri (2020) proved steady-state convergence for a "generalized switch" that was first studied in Stolyar (2004). The authors called their approach the "transform method," which is essentially our BAR approach in the discrete-time setting. Discrete time offers simplifications not available in our continuous-time setting. We emphasize that our approach is complicated not only by continuous time, but also by the presence of general interarrival and service-time distributions. Indeed, Wang et al. (2022) is one example of the drift method being applied to the famous bandwidth-sharing model; by assuming phase-type job size distributions, the authors were able to study the model in the CTMC setting and therefore did not need to deal with the added complexity of general job size distributions.

This paper is composed of eight sections. In Section 2, we exemplify the BAR approach for a two-station, five-class reentrant line with SBP service discipline. To simplify the analysis, it is assumed that all the interarrival and service times are either exponentially distributed or generally distributed but bounded. Proposition 2.1 is a main result of this section, which is a special case of Theorem 5.1. Although this network is simple, it illustrates the main ideas of the BAR approach. In Section 3, multiclass queueing networks are introduced, while SRBM and its BAR are discussed in Section 4. Then, the main result, Theorem 5.1, and its corollary are presented in Section 5. The preliminary results for proving Theorem 5.1 are given in Section 6. The proof of Theorem 5.1 is divided into six steps. Steps 2 through 6 are proved in Section 7 whereas step 1 is proved in Section 8, where SSC under the Palm distributions is obtained. This SSC is a key result in this step and may be interesting itself. This is the reason why the first step is separately proved in Section 8. Some auxiliary results are given in Appendices A, B, and C.

2. Two-Station, Five-Class Queueing Network

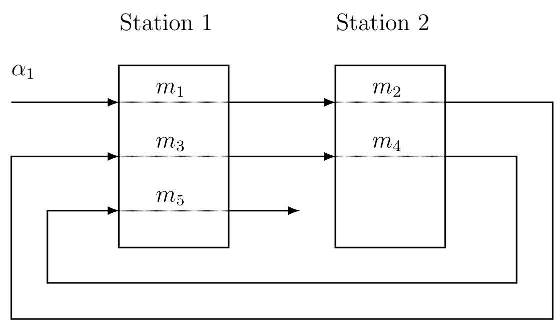
In this section, we first introduce a pilot example of a two-station, five-class queueing network operating under a SBP service discipline. We then state the main result of this paper in the setting of this two-station network. Finally, we prove the result in two steps: (i) when the interarrival and service-time distributions are exponential and (ii) when the interarrival and service-time distributions are general with bounded supports. By focusing first on the two-station setting, we avoid an elaborate notational system that is required for a general queueing network but are able to highlight the key technical contributions of this paper.

2.1. Network Description

Figure 1 depicts a two-station, five-class queueing network. Each rectangle represents a single-server station that processes jobs one at a time. Jobs arrive to the network exogeneously following a renewal process. Each job has five processing steps in the network that follow the flow indicated by the arrows in the figure; server 1 performs steps 1, 3, and 5 at station 1, whereas server 2 performs steps 2 and 4 at station 2. When a job completes its processing at step k and the server at step $k+1$ is busy, the job moves to buffer $k+1$ and waits for its turn to be processed at step $k+1$. After finishing step 5 processing, jobs exit the network.

Each buffer is assumed to have infinite capacity. Following Harrison (1988), we adopt the notion of job classes. A job belongs to class k if it is either processing in step k or waiting in buffer k. We use the terms "class" and "buffer" interchangeably, with the understanding that a job in step k processing still belongs to buffer k. Let $m_k T_{s,k}(i)$ be the processing time of the ith class k job and let $\{m_k T_{s,k}(i), i \geq 1\}$ be the corresponding sequence of

Figure 1. Two-Station, Five-Class Reentrant Line



processing times. We assume that the elements of this sequence are independent and identically distributed (i.i.d.) with mean m_k and $\mathbb{E}[T_{s,k}(i)] = 1$. The interarrival times $\{(1/\lambda_1)T_{e,1}(i), i \geq 1\}$ of the exogenous arrival process are assumed to be i.i.d. with mean $1/\lambda_1$ and $\mathbb{E}[T_{e,1}(i)] = 1$. We assume these sequences are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We also assume that different i.i.d. sequences are independent. For a positive random variable U, its squared coefficient of variation (SCV), denoted as $c^2(U)$, is defined to be

$$c^2(U) = \frac{\mathrm{Var}(U)}{(\mathbb{E}[U])^2}.$$

When server 1 completes the processing of a class $k \in \{1, 3, 5\}$ job, it needs a service discipline to decide which buffer the next job should be picked from. For our pilot example, we specify service discipline by the following list:

$$\{(5, 3, 1), (2, 4)\}. \tag{2.1}$$

This list means that, at station 1, this discipline gives the highest priority to class 5, the next priority to class 3, and the lowest priority to class 1, whereas at station 2, the highest priority goes to class 2 and the lowest priority to class 4. We further assume that the service discipline is *preemptive-resume*: When a job with a higher rank than the one currently being served arrives at the server's station, the service of the current job is interrupted. When all jobs of higher rank are served, the interrupted service continues from where it left off. This service discipline is referred to as SBP, which is defined for a general multiclass queueing network in Section 3.

2.2. Markov Process and Its Stability

For $t \geq 0$ define

$$X(t) = \begin{pmatrix} Z(t) \\ U_1(t) \\ V(t) \end{pmatrix}, \quad Z(t) = \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \\ Z_5(t) \end{pmatrix}, \quad V(t) = \begin{pmatrix} V_1(t) \\ V_2(t) \\ V_3(t) \\ V_4(t) \\ V_5(t) \end{pmatrix}, \tag{2.2}$$

where $Z_k(t)$ is the number of class k jobs at time t, including possibly the one in service, $U_1(t)$ is the remaining interarrival time of the next class 1 job, and $V_k(t)$ is the remaining service time of the leading class k job at time t, assuming that the class k server devotes its entire service capacity to this job. If there is no class k job in service at time t, then $V_k(t)$ is the service time of the next class k job in service.

In this paper all vectors as column vectors, but, notationally, column vectors are bulkier than row vectors. Although we could write

$$X(t) = (Z^T(t), U_1^T(t), V^T(t))^T,$$

keeping track of all the transpose scripts ^{T} is also cumbersome. Therefore, going forward we omit the script ^{T} and envision vectors to be column vectors, unless we state otherwise. For example, we write $X(t) = (Z(t), U_1(t), V(t))$ to denote the column vector in (2.2).

It is known that $\{X(t), t \geq 0\}$ is a continuous-time Markov process with state space $S = \mathbb{Z}_+^5 \times \mathbb{R}_+^6$, where $\mathbb{Z}_+ = \{0, 1, \ldots\}$, and we call $X(t)$ the state of the queueing network at time t. This process is piecewise deterministic because between jumps, $X(t)$ evolves deterministically in t. We adopt the convention that each sample path of the state process is right continuous.

It follows from theorem 4.1 of Dai (1995) and section 8.7 of Dai and Harrison (2020) that, under a mild assumption on the interarrival time distribution, the Markov process $\{X(t), t \geq 0\}$ is positive Harris recurrent and thus has a unique stationary distribution when the conditions

$$\rho_1 = \lambda_1(m_1 + m_3 + m_5) < 1, \tag{2.3}$$

$$\rho_2 = \lambda_1(m_2 + m_4) < 1, \tag{2.4}$$

$$\rho_v = \lambda_1(m_2 + m_5) < 1, \tag{2.5}$$

are satisfied. In such a case, we use

$$X = (Z, U_1, V), \quad \text{where } Z = (Z_1, Z_2, Z_3, Z_4, Z_5) \text{ and } V = (V_1, V_2, V_3, V_4, V_5),$$

to denote the random vector distributed according to the stationary distribution.

Lemma 2.1. *When Conditions (2.3)–(2.5) are satisfied, the following are satisfied:*

$$\beta_1 \equiv \mathbb{P}\{Z_1 = 0, Z_3 = 0, Z_5 = 0\} = 1 - \lambda_1(m_1 + m_3 + m_5) = 1 - \rho_1, \tag{2.6}$$

$$\beta_3 \equiv \mathbb{P}\{Z_3 = 0, Z_5 = 0\} = 1 - \lambda_1(m_3 + m_5), \quad \beta_5 \equiv \mathbb{P}\{Z_5 = 0\} = 1 - \lambda_1 m_5, \tag{2.7}$$

$$\beta_4 \equiv \mathbb{P}\{Z_2 = 0, Z_4 = 0\} = 1 - \lambda_1(m_2 + m_4) = 1 - \rho_2, \quad \beta_2 \equiv \mathbb{P}\{Z_2 = 0\} = 1 - \lambda_1 m_2. \tag{2.8}$$

For a proof of this lemma, see Lemma 6.6 in Section 6.2. Thus, when (2.3)–(2.5) hold, the quantity ρ_i is the long-run utilization of server $i \in \{1, 2\}$. Conditions (2.3) and (2.4) ensure that servers 1 and 2 are not overloaded in the long run. Condition (2.5) is known as the virtual station condition, where ρ_v is the traffic intensity of the virtual station and is *unusual*. As explained in Dai and Vande Vate (2000), under the SBP discipline (2.1), classes 2 and 5 form a virtual station for which Condition (2.5) is the load condition. When the Markov process has a stationary distribution, we call the queueing network stable. For a general queueing network (to be introduced in Section 3) operating under an arbitrary SBP discipline, characterizing its stability region in a manner similar to (2.3)–(2.5) remains an open problem.

2.3. Heavy-Traffic Limit Theorem

We consider a sequence of queueing networks indexed by $r \in (0, 1]$. Readers are referred to Section 3 for a motivation for studying a sequence of networks. For notational simplicity, only the arrival rate $\lambda_1^{(r)}$ is assumed to depend on r. We assume that $\lambda_1^{(r)} = 1 - r$ for $r \in (0, 1]$ and that

$$m_1 + m_3 + m_5 = m_2 + m_4 = 1, \tag{2.9}$$

$$m_2 + m_5 < 1. \tag{2.10}$$

Under Condition (2.9), (2.10) is equivalent to

$$m_5 < m_4. \tag{2.11}$$

Under Condition (2.9), $\rho_1^{(r)} = \rho_2^{(r)} = 1 - r$. Thus,

$$r^{-1}(1 - \rho_1^{(r)}) = 1 \quad \text{and} \quad r^{-1}(1 - \rho_2^{(r)}) = 1, \quad r \in (0, 1). \tag{2.12}$$

In particular, $\rho_i^{(r)} \uparrow 1$ for $i = 1, 2$, as $r \to 0$. Condition (2.12) is a special case of the heavy-traffic Condition (5.4)–(5.6) to be introduced in Section 5 for a general sequence of networks. Condition (2.10) implies stability of the queueing network for any $r \in (0, 1)$, and we let $X^{(r)}$ denote the random element having the stationary distribution of the Markov process $\{X^{(r)}(t), t \geq 0\}$. We let

$$Z^{(r)} = (Z_1^{(r)}, \ldots, Z_5^{(r)})^T$$

be the column vector of steady-state job counts. The following proposition is a special case of Theorem 5.1 in Section 5. The SRBM in the proposition has been well studied; see, for example, section 2.3 of Braverman et al. (2017). To make this paper as self-contained as possible, the background materials on SRBM will be presented in Section 4.

Proposition 2.1. *There exists a random element* $(Z_1^*, Z_4^*) \in \mathbb{R}_+^2$ *such that*

$$rZ^{(r)} \Rightarrow Z^* = (Z_1^*, 0, 0, Z_4^*, 0)^T, \quad \text{as } r \to 0, \tag{2.13}$$

where "⇒" denotes convergence in distribution. Furthermore, the distribution of (Z_1^*, Z_4^*) *on* $\mathbb{R}_+^2$ *is the unique stationary distribution of a semimartingale reflecting Brownian motion (SRBM) with covariance matrix* Σ*, reflection matrix* R*, and drift vector* $-Rb$*, where*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{14} \\ \Sigma_{41} & \Sigma_{44} \end{pmatrix}, \quad R = \begin{pmatrix} R_{11} & R_{14} \\ R_{41} & R_{44} \end{pmatrix} = \frac{1}{m_4 - m_5}\begin{pmatrix} m_4 & -m_5 \\ -1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{2.14}$$

and

$$\Sigma_{11} = \frac{1}{2(\mu_5 - \mu_4)^2}((\mu_5 - \mu_4)^2 c_{e,1}^2 + m_1^2\mu_5^2 c_{s,1}^2 + (\mu_4 - 1)^2 c_{s,2}^2 + m_3^2\mu_5^2 c_{s,3}^2 + c_{s,4}^2 + c_{s,5}^2), \tag{2.15}$$

$$\Sigma_{14} = \Sigma_{41} = \frac{-1}{(\mu_5 - \mu_4)^2}(m_1^2\mu_5^2\mu_4 c_{s,1}^2 + (\mu_4 - 1)^2\mu_5 c_{s,2}^2 + m_3^2\mu_5^2\mu_4 c_{s,3}^2 + \mu_5 c_{s,4}^2 + \mu_4 c_{s,5}^2), \tag{2.16}$$

$$\Sigma_{44} = \frac{1}{2(\mu_5 - \mu_4)^2}(m_1^2\mu_5^2\mu_4^2 c_{s,1}^2 + (\mu_4 - 1)^2\mu_5^2 c_{s,2}^2 + m_3^2\mu_5^2\mu_4^2 c_{s,3}^2 + \mu_5^2 c_{s,4}^2 + \mu_4^2 c_{s,5}^2). \tag{2.17}$$

The remainder of this section is dedicated to proving the proposition, first for the case when interarrival and service-time distributions are exponential and then for the case when interarrival and service time distributions are general with bounded support.

2.4. Exponential Distributions

In this section, we prove Proposition 2.1, under the assumption that interarrival and service-time distributions are exponential. In such a case, we can drop the components $(U_1^{(r)}(t), V^{(r)}(t))$ in the state description $X^{(r)}(t)$ because $\{Z^{(r)}(t), t \geq 0\}$ is a CTMC on the state space $S = \mathbb{Z}_+^5$, where $\mathbb{Z}_+ = \{0, 1, \ldots\}$. When $\lambda_1^{(r)} = 1 - r$ and (2.9)–(2.10) are satisfied, each CTMC in the sequence has a unique stationary distribution, and we recall that $Z^{(r)} \in \mathbb{Z}_+^5$ denotes the steady-state job count.

2.4.1. BAR. For the moment, we focus on a single network within the sequence of networks. We omit the index r for convenience; for example, $\lambda_1^{(r)}$ is denoted by λ_1. The main purpose of this section is to derive the Laplace transform version of the BAR (2.27). We use the terminology "Laplace transform" for a moment-generating function (MGF) if its domain is nonpositive.

It is well known that the stationary distribution π of the CTMC is characterized by the basic adjoint relationship (BAR)

$$\mathbb{E}[Gf(Z)] = 0 \quad \text{for each bounded function} \quad f: S \to \mathbb{R}, \tag{2.18}$$

where for each state $z \in S$ and each function $f: S \to \mathbb{R}$,

$$\begin{aligned} Gf(z) = {} & \lambda_1(f(z + e^{(1)}) - f(z)) + \mu_1(f(z - e^{(1)} + e^{(2)}) - f(z))1(z_5 = 0, z_3 = 0, z_1 > 0) \\ & + \mu_2(f(z - e^{(2)} + e^{(3)}) - f(z))1(z_2 > 0) + \mu_3(f(z - e^{(3)} + e^{(4)}) - f(z))1(z_5 = 0, z_3 > 0) \\ & + \mu_4(f(z - e^{(4)} + e^{(5)}) - f(z))1(z_2 = 0, z_4 > 0) + \mu_5(f(z - e^{(5)}) - f(z))1(z_5 > 0), \end{aligned} \tag{2.19}$$

and $e^{(j)} \in \mathbb{R}^5$ is the vector with a one in the jth component and zeros elsewhere. For a proof of (2.18), see, for example, Glynn and Zeevi (2008). When all states in S are linearly ordered, each test function $f: S \to \mathbb{R}$ is equivalent to a column vector of infinite dimensions and Gf is the usual matrix-vector product, where G is the corresponding square matrix known as the generator matrix of the CTMC.

The term

$$\mu_3(f(z - e^{(3)} + e^{(4)}) - f(z))1(z_5 = 0, z_3 > 0)$$

represents the state transition from state z to state $z - e^{(3)} + e^{(4)}$ due to the service completion of a class 3 job by server 1. Because of the SBP discipline in (2.1), this can happen only when class 5 has no job and class 3 has jobs.

Hence, the term has the indicator function of the set $(z_5 = 0, z_3 > 0)$. The service completion triggers a deletion of a job in class 3 (the $-e^{(3)}$ term) and an addition of a job to class 4 (the $+e^{(4)}$ term). The μ_3 term reflects the service rate when server 1 is fully dedicated to a class 3 job. Other terms in the definition of Gf can be understood similarly.

Equation (2.18) is a shorthand for

$$\sum_{z\in S}\mathbb{P}\{Z=z\}Gf(z)=0 \quad \text{or} \quad \sum_{z\in S}\pi(z)Gf(z)=0 \quad \text{for each } f: S\to\mathbb{R}.$$

The latter sum is equal to πGf when the stationary distribution π is viewed as a row vector, G as a square matrix, and f as a column vector. Clearly, $\pi Gf = 0$ for each bounded $f : S \to \mathbb{R}$ is equivalent to

$$\pi G = 0,$$

which is known as the balance equations that characterize the stationary distribution π of a CTMC with generator G.

Throughout this paper, we use the following notion. For each integer $d > 0$, let

$$\mathbb{R}^d_- = \{x=(x_1,\dots,x_d)^T \in \mathbb{R}^d : x_i \le 0 \text{ for } = 1,\dots,d\}.$$

Fixing a $\theta \in \mathbb{R}^5_-$, we define the bounded test function $g_\theta : S \to \mathbb{R}$ by

$$g_\theta(z) = e^{\langle\theta,z\rangle}, \tag{2.20}$$

where for $a, b \in \mathbb{R}^5$, $\langle a,b\rangle = \sum_{k=1}^5 a_k b_k$. Applying G to this function, one has

$$\begin{aligned} Gg_\theta(z) = [&\lambda_1\eta_1(\theta_1) + \mu_1\xi_1(\theta)1(z_1>0, z_3=0, z_5=0) + \mu_2\xi_2(\theta)1(z_2>0) \\ &+ \mu_3\xi_3(\theta)1(z_3>0, z_5=0) + \mu_4\xi_4(\theta)1(z_2=0, z_4>0) + \mu_5\xi_5(\theta)1(z_5>0)]g_\theta(z). \end{aligned} \tag{2.21}$$

Here,

$$\eta_1(\theta_1) = e^{\theta_1} - 1, \quad \xi_k(\theta) = e^{\theta_{k+1}-\theta_k} - 1 \quad \text{for } k \in \{1,2,3,4\}, \quad \xi_5(\theta) = e^{-\theta_5} - 1. \tag{2.22}$$

It follows from the BAR (2.18) and (2.21) that

$$\begin{aligned} &\lambda_1\eta_1(\theta_1)\mathbb{E}[g_\theta(X)] + \mu_1\xi_1(\theta)\mathbb{E}[g_\theta(X)1(Z_1>0, Z_3=0, Z_5=0)] \\ &+ \mu_3\xi_3(\theta)\mathbb{E}[g_\theta(X)1(Z_3>0, Z_5=0)] + \mu_5\xi_5(\theta)\mathbb{E}[g_\theta(X)1(Z_5>0)] \\ &+ \mu_2\xi_2(\theta)\mathbb{E}[g_\theta(X)1(Z_2>0)] + \mu_4\xi_4(\theta)\mathbb{E}[g_\theta(X)1(Z_2=0, Z_4>0)] = 0 \text{ for } \theta \in \mathbb{R}^5_-. \end{aligned} \tag{2.23}$$

We call this the Laplace transform version of the BAR (2.18). Let us define

$$\phi(\theta) = \mathbb{E}[g_\theta(Z)], \quad \phi_1(\theta) = \mathbb{E}[g_\theta(Z) | Z_1 = 0, Z_3 = 0, Z_5 = 0], \tag{2.24}$$

$$\phi_3(\theta) = \mathbb{E}[g_\theta(Z) | Z_3 = 0, Z_5 = 0], \quad \phi_5(\theta) = \mathbb{E}[g_\theta(Z) | Z_5 = 0], \tag{2.25}$$

$$\phi_2(\theta) = \mathbb{E}[g_\theta(Z) | Z_2 = 0], \quad \phi_4(\theta) = \mathbb{E}[g_\theta(Z) | Z_2 = 0, Z_4 = 0]. \tag{2.26}$$

The following lemma rewrites (2.23) in a more convenient form. Namely, (2.23) is written as a linear combination form of $\phi(\theta)$ and $\phi(\theta) - \phi_i(\theta)$ for $i = 1, 2, \dots 5$.

Lemma 2.2. *For each* $\theta \in \mathbb{R}^5_-$,

$$\begin{aligned} &\left(\lambda_1\eta_1(\theta_1) + \sum_{k=1}^5 \lambda_1\xi_k(\theta)\right)\phi(\theta) + [\mu_5\xi_5(\theta) - \mu_3\xi_3(\theta)]\beta_5(\phi(\theta) - \phi_5(\theta)) \\ &+ [\mu_3\xi_3(\theta) - \mu_1\xi_1(\theta)]\beta_3(\phi(\theta) - \phi_3(\theta)) + \mu_1\xi_1(\theta)(1-\rho_1)(\phi(\theta) - \phi_1(\theta)) \\ &+ \mu_4\xi_4(\theta)(1-\rho_2)(\phi(\theta) - \phi_4(\theta)) + [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(\phi(\theta) - \phi_2(\theta)) = 0, \end{aligned} \tag{2.27}$$

where β_k*, defined in Lemma 2.1, is the steady-state probability that all classes with priority greater or equal to class k have no customers.*

Proof. Our starting point is (2.23). Consider the last line in (2.23). Lemma 2.1 implies that

$$\mathbb{E}[g_\theta(Z)1(Z_2 > 0)] = \mathbb{E}[g_\theta(Z)] - \mathbb{E}[g_\theta(Z)1(Z_2 = 0)] = \phi(\theta) - \beta_2\phi_2(\theta),$$
$$\mathbb{E}[g_\theta(Z)1(Z_2 = 0, Z_4 > 0)] = \beta_2\phi_2(\theta) - \beta_4\phi_4(\theta).$$

Hence,

$$\begin{aligned}
&\mathbb{E}[(\mu_2\xi_2(\theta)1(Z_2 > 0) + \mu_4\xi_4(\theta)1(Z_2 = 0, Z_4 > 0))g_\theta(Z)]\\
&= \mu_2\xi_2(\theta)\phi(\theta) + [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(-\phi_2(\theta)) + \mu_4\xi_4(\theta)\beta_4(-\phi_4(\theta))\\
&= (\mu_2\xi_2(\theta) - [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2 - \mu_4\xi_4(\theta)\beta_4)\phi(\theta)\\
&\quad + [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(\phi(\theta) - \phi_2(\theta)) + \mu_4\xi_4(\theta)\beta_4(\phi(\theta) - \phi_4(\theta))\\
&= (\mu_2\xi_2(\theta)(1 - \beta_2) + \mu_4\xi_4(\theta)(\beta_2 - \beta_4))\phi(\theta)\\
&\quad + [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(\phi(\theta) - \phi_2(\theta)) + \mu_4\xi_4(\theta)\beta_4(\phi(\theta) - \phi_4(\theta)).
\end{aligned}$$

From Lemma 2.1, we know that $\beta_4 = 1 - \lambda_1(m_2 + m_4) = 1 - \rho_2$ and $\beta_2 = 1 - \lambda_1 m_2$, implying that the right-hand side equals

$$\begin{aligned}
&(\lambda_1\xi_2(\theta) + \lambda_1\xi_4(\theta))\phi(\theta)\\
&+ [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(\phi(\theta) - \phi_2(\theta)) + \mu_4\xi_4(\theta)(1 - \rho_2)(\phi(\theta) - \phi_4(\theta)).
\end{aligned} \tag{2.28}$$

Similarly, one can show that

$$\begin{aligned}
&\mathbb{E}[(\mu_5\xi_5(\theta)1(Z_5 > 0) + \mu_3\xi_3(\theta)1(Z_3 > 0, Z_5 = 0) + \mu_1\xi_1(\theta)1(Z_1 > 0, Z_3 = 0, Z_5 = 0))g_\theta(Z)]\\
&= (\lambda_1\xi_1(\theta) + \lambda_1\xi_3(\theta) + \lambda_1\xi_5(\theta))\phi(\theta) + [\mu_5\xi_5(\theta) - \mu_3\xi_3(\theta)]\beta_5(\phi(\theta) - \phi_5(\theta))\\
&\quad + [\mu_3\xi_3(\theta) - \mu_1\xi_1(\theta)]\beta_3(\phi(\theta) - \phi_3(\theta)) + \mu_1\xi_1(\theta)(1 - \rho_1)(\phi(\theta) - \phi_1(\theta)).
\end{aligned} \tag{2.29}$$

Finally, (2.27) follows from (2.23) and the expressions in (2.28) and (2.29). □

2.4.2. Taylor Expansion and Asymptotic BAR.

For each $\theta \in \mathbb{R}^5_-$, define

$$\phi^{(r)}(\theta) = \mathbb{E}[g_\theta(rZ^{(r)})] = \mathbb{E}[g_{r\theta}(Z^{(r)})],$$

which is analogous to the definition of $\phi(\theta)$ in (2.24). Define $\phi_k^{(r)}(\theta)$ similarly for $k \in \{1, \dots, 5\}$. We now present Lemma 2.3, in which we start with (2.27) and replace $\eta_1(\theta_1)$ and $\xi_k(\theta)$ by their second-order Taylor expansions, which are simpler to work with, to derive what we call the asymptotic BAR in (2.34). Later we will see how the asymptotic BAR allows us to characterize the limiting distribution of $rZ^{(r)}$ as $r \to 0$.

To state Lemma 2.3, we define

$$\overline{\eta}_1(\theta_1) = \theta_1 \quad \text{and} \quad \tilde{\eta}_1(\theta_1) = \frac{1}{2}\theta_1^2. \tag{2.30}$$

We will see in the proof of Lemma 2.3 that $\overline{\eta}_1(\theta_1)$ and $\tilde{\eta}_1(\theta_1)$ are the first- and second-order terms, respectively, of the Taylor expansion of $\eta_1(\theta_1)$. Similarly, we define

$$\overline{\xi}_k(\theta) = \theta_{k+1} - \theta_k, \quad \tilde{\xi}_k(\theta) = \frac{1}{2}(\theta_{k+1} - \theta_k)^2 \quad k \in \{1, \dots, 4\}, \tag{2.31}$$

$$\overline{\xi}_5(\theta) = -\theta_5, \quad \tilde{\xi}_5(\theta) = \frac{1}{2}\theta_5^2. \tag{2.32}$$

Last, we define

$$\eta_1^*(\theta_1) = \overline{\eta}_1(\theta_1) + \tilde{\eta}_1(\theta_1), \quad \xi_k^*(\theta) = \overline{\xi}_k(\theta) + \tilde{\xi}_k(\theta), \quad k \in \{1, \dots, 5\}, \tag{2.33}$$

to be the second-order approximations of $\eta_1(\theta_1)$ and $\xi_k(\theta)$, respectively.

Lemma 2.3. *For each $\theta \in \mathbb{R}^5_-$, as $r \to 0$,*

$$\begin{aligned} &r^2\left(\lambda_1^{(r)}\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5}\lambda_1^{(r)}\tilde{\xi}_k(\theta)\right)\phi^{(r)}(\theta) \\ &\quad + r^2\mu_1\overline{\xi}_1(\theta)(\phi^{(r)}(\theta) - \phi_1^{(r)}(\theta)) + \mu_4 r^2\overline{\xi}_4(\theta)(\phi^{(r)}(\theta) - \phi_4^{(r)}(\theta)) \\ &\quad + [\mu_3\xi_3^*(r\theta) - \mu_1\xi_1^*(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) \\ &\quad + [\mu_5\xi_5^*(r\theta) - \mu_3\xi_3^*(r\theta)]\beta_5^{(r)}(\phi^{(r)}(\theta) - \phi_5^{(r)}(\theta)) \\ &\quad + [\mu_2\xi_2^*(r\theta) - \mu_4\xi_4^*(r\theta)]\beta_2^{(r)}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta)) = o(r^2). \end{aligned} \tag{2.34}$$

Proof. Replacing θ in (2.27) by $r\theta$, one has that for each $\theta \in \mathbb{R}^5_-$ and each $r \in (0,1)$,

$$\begin{aligned} &\left(\lambda_1^{(r)}\eta_1(r\theta_1) + \sum_{k=1}^{5}\lambda_1^{(r)}\xi_k(r\theta)\right)\phi^{(r)}(\theta) + [\mu_5\xi_5(r\theta) - \mu_3\xi_3(r\theta)]\beta_5^{(r)}(\phi^{(r)}(\theta) - \phi_5^{(r)}(\theta)) \\ &+ [\mu_3\xi_3(r\theta) - \mu_1\xi_1(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) + \mu_1\xi_1(r\theta)(1-\rho_1^{(r)})(\phi^{(r)}(\theta) - \phi_1^{(r)}(\theta)) \\ &+ \mu_4\xi_4(r\theta)(1-\rho_2^{(r)})(\phi^{(r)}(\theta) - \phi_4^{(r)}(\theta)) + [\mu_2\xi_2(r\theta) - \mu_4\xi_4(r\theta)]\beta_2^{(r)}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta)) = 0. \end{aligned}$$

Using the Taylor expansion $e^x = 1 + x + \frac{1}{2}x^2 + o(x)$ when $x \to 0$, one has

$$\eta_1(\theta_1) = \eta_1^*(\theta_1) + o(\theta_1^2) \text{ as } \theta_1 \to 0, \tag{2.35}$$

$$\xi_k(\theta) = \xi_k^*(\theta) + o(|\theta|^2) \text{ as } \theta \to 0 \text{ for } k \in \{1,\dots,5\}. \tag{2.36}$$

Therefore, for each $\theta \in \mathbb{R}^5_-$ as $r \to 0$,

$$\begin{aligned} &\left(\lambda_1^{(r)}\eta_1^*(r\theta_1) + \sum_{k=1}^{5}\lambda_1^{(r)}\xi_k^*(r\theta)\right)\phi^{(r)}(\theta) + [\mu_5\xi_5^*(r\theta) - \mu_3\xi_3^*(r\theta)]\beta_5^{(r)}(\phi^{(r)}(\theta) - \phi_5^{(r)}(\theta)) \\ &+ [\mu_3\xi_3^*(r\theta) - \mu_1\xi_1^*(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) + \mu_1\xi_1^*(r\theta)(1-\rho_1^{(r)})(\phi^{(r)}(\theta) - \phi_1^{(r)}(\theta)) \\ &+ \mu_4\xi_4^*(r\theta)(1-\rho_2^{(r)})(\phi^{(r)}(\theta) - \phi_4^{(r)}(\theta)) + [\mu_2\xi_2^*(r\theta) - \mu_4\xi_4^*(r\theta)]\beta_2^{(r)}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta)) = o(r^2). \end{aligned}$$

Using the facts that $\overline{\eta}_1(\theta_1) + \sum_{k=1}^5 \overline{\xi}_k(\theta) = 0$, that $\tilde{\xi}_k(r\theta) = r^2\tilde{\xi}_k(\theta)$ for each $\theta \in \mathbb{R}^5_-$, and that $1 - \rho_i^{(r)} = r$, we have (2.34), proving the lemma. □

2.4.3. SSC. It follows from theorem 3.7 and section 4.1 of Cao et al. (2022) that the following moment SSC holds:

$$\limsup_{r\to 0} \ \mathbb{E}[Z_2^{(r)} + Z_3^{(r)} + Z_5^{(r)}]^2 < \infty. \tag{2.37}$$

In fact, we now argue that as a consequence of (2.37), for any $\theta \in \mathbb{R}^5_-$,

$$\lim_{r\to 0}(\phi^{(r)}(\theta) - \phi^{(r)}(\theta_L, 0)) = 0, \quad \lim_{r\to 0}(\phi_k^{(r)}(\theta) - \phi_k^{(r)}(\theta_L, 0)) = 0, \quad k \in \{1,\dots,5\}, \tag{2.38}$$

where, for a function $f : \mathbb{R}^5 \to \mathbb{R}$, $f(\theta_L, 0)$ is a shorthand for $f(\theta_1, 0, 0, \theta_4, 0)$ with $\theta_L = (\theta_1, \theta_4)^T$. We refer to (2.38) as the Laplace transform version of SSC. To prove the first equality in (2.38),

$$|\phi^{(r)}(\theta) - \phi^{(r)}(\theta_L, 0)| \le \mathbb{E}[1 - (e^{\sum_{k\in\{2,3,5\}}\theta_k r Z_k^{(r)}})] \le \mathbb{E}\left[\sum_{k\in\{2,3,5\}} |\theta_k| r Z_k^{(r)}\right],$$

where the second inequality follows from $1 - e^{-x} \le x$ for $x \ge 0$. The last term converges to zero because of (2.37) and Jensen's inequality. The rest of (2.38) is proved similarly.

Proof of Proposition 2.1. The SSC in the preceding paragraph implies that $\lim_{r\to 0} rZ_k^{(r)} \Rightarrow 0$ for $k \in \{2,3,5\}$. To prove Proposition 2.1, it remains to show that $(rZ_1^{(r)}, rZ_4^{(r)})$ converges and characterizes the limit. We begin with the following lemma, which is stated for the general queueing network setting in Lemma 7.1.

Lemma 2.4. *For any sequence* $\{(\phi^{(r_n)}(\theta), \phi_1^{(r_n)}(\theta), \ldots, \phi_5^{(r_n)}(\theta))\}_{n=1}^{\infty}$ *with* $r_n \in (0,1)$ *and* $\lim_{n\to\infty} r_n = 0$, *there exists a subsequence indexed by* $\{r_{n_k}\}$ *such that*

$$\lim_{k\to\infty} (\phi^{(r_{n_k})}(\theta), \phi_1^{(r_{n_k})}(\theta), \ldots, \phi_5^{(r_{n_k})}(\theta)) = (\phi^*(\theta), \phi_1^*(\theta), \ldots, \phi_5^*(\theta)) \quad \text{for each } \theta \in \mathbb{R}^5_-.$$

We call $(\phi^*(\theta), \phi_1^*(\theta), \ldots, \phi_5^*(\theta))$ *a limit point of* $\{(\phi^{(r)}(\theta), \phi_1^{(r)}(\theta), \ldots, \phi_5^{(r)}(\theta))\}_{r\in(0,1)}$.

We show that the set of all limit points in Lemma 2.4 is a singleton by proving that there is a random vector $(Z_1^*, Z_4^*) \in \mathbb{R}^2_+$, independent of the subsequence $\{r_{n_k}\}$, such that

$$\phi^*(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = \mathbb{E}[e^{\theta_1 Z_1^* + \theta_4 Z_4^*}] \quad \text{for each } \theta \in \mathbb{R}^5_-. \tag{2.39}$$

Because every sequence $\{\phi^{(r_n)}(\theta)\}_{n=1}^{\infty}$ contains a convergent subsequence that converges to the limit point defined by (2.39), it follows that

$$\lim_{r\to 0} \phi^{(r)}(\theta) = \phi^*(\theta) = \mathbb{E}[e^{\theta_1 Z_1^* + \theta_4 Z_4^*}] \quad \text{for each } (\theta_1, \theta_4)^T \in \mathbb{R}^2_-,$$

which is equivalent to the convergence in (2.13). The following informal discussion outlines how we prove (2.39).

Let us assume for simplicity that $\phi^{(r)}(\theta), \phi_1^{(r)}(\theta), \ldots, \phi_5^{(r)}(\theta)$ converge pointwise as $r \to 0$. Otherwise, we can replace r by r_{n_k} and be assured that $\phi^{(r_{n_k})}(\theta), \phi_1^{(r_{n_k})}(\theta), \ldots, \phi_5^{(r_{n_k})}(\theta)$ converge. We characterize the limit point using the asymptotic BAR (2.34) as follows. Dividing both sides of (2.34) by r^2 and letting $r \to 0$ yields

$$\begin{aligned}
&\left(\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5} \tilde{\xi}_k(\theta)\right)\phi^*(\theta) + \mu_1 \bar{\xi}_1(\theta)(\phi^*(\theta) - \phi_1^*(\theta)) + \mu_4 \bar{\xi}_4(\theta)(\phi^*(\theta) - \phi_4^*(\theta)) \\
&= -\lim_{r\downarrow 0} \frac{1}{r^2}([\mu_3 \xi_3^*(r\theta) - \mu_1 \xi_1^*(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) \\
&\qquad + [\mu_5 \xi_5^*(r\theta) - \mu_3 \xi_3^*(r\theta)]\beta_5^{(r)}(\phi^{(r)}(\theta) - \phi_5^{(r)}(\theta)) \\
&\qquad + [\mu_2 \xi_2^*(r\theta) - \mu_4 \xi_4^*(r\theta)]\beta_2^{(r)}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta))), \quad \theta \in \mathbb{R}^5_-.
\end{aligned} \tag{2.40}$$

We proceed in two steps. In step one, we identify a subset of $\mathbb{R}^5_-$ such that the right-hand side of (2.40) is zero for all θ in this subset. We do this because we are unable to characterize the right-hand side outside this subset. Step 1 requires Lemmas 2.5, 2.6, and 2.7, which are stated later. Following these lemmas, we use (2.40), the right-hand side of which now equals zero, to characterize $\phi^*(\theta)$ and prove (2.39)—this is step 2.

Let us compare our example to Braverman et al. (2017), who applied the BAR approach with Laplace transforms to generalized Jackson networks (GJNs). In that paper, the authors derived an asymptotic BAR for GJNs, which allowed them to obtain an equation that is analogous to (2.40). However, because GJNs are single-class queueing networks, the right-hand side of their equation equals zero. This means that Braverman et al. (2017) did not need to perform step one of the previous paragraph, whereas we do because our two-station example is a multiclass queueing network.

We now carry out step 1. To understand how to choose θ so the right-hand side of (2.40) equals zero, we examine the first term inside the parentheses. Namely,

$$\begin{aligned}
&\frac{1}{r^2}[\mu_3 \xi_3^*(r\theta) - \mu_1 \xi_1^*(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) \\
&= \frac{1}{r^2}[\mu_3 \bar{\xi}_3(r\theta) - \mu_1 \bar{\xi}_1(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) \\
&\quad + \frac{1}{r^2}[\mu_3 \tilde{\xi}_3(r\theta) - \mu_1 \tilde{\xi}_1(r\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)).
\end{aligned}$$

Because $\bar{\xi}_k(\theta)$ are linear in θ, we show in Lemma 2.5 that we can choose θ to make the first term on the right-hand side equal zero. Furthermore, $\sup_{r\in(0,1)} |\tilde{\xi}_k(r\theta)|/r^2 < \infty$ because $\tilde{\xi}_k(\theta)$ are quadratic in θ; see (2.31). Therefore, to prove that the second term vanishes as $r \to 0$, we show in Lemma 2.6 that $\lim_{r\to 0}(\phi^{(r)}(\theta) - \phi_k^{(r)}(\theta)) = 0$ for $k \in \{2,3,5\}$.

Recall that all vectors are envisioned as column vectors, and recall our convention of writing column vectors discussed in Section 2.2.

Lemma 2.5. *Recall the definition of $\overline{\xi}_k(\theta)$ from (2.31) and consider the system of linear equations:*

$$\mu_3\overline{\xi}_3(\theta) - \mu_1\overline{\xi}_1(\theta) = \mu_3(\theta_4 - \theta_3) - \mu_1(\theta_2 - \theta_1) = 0, \tag{2.41}$$

$$\mu_5\overline{\xi}_5(\theta) - \mu_3\overline{\xi}_3(\theta) = -\mu_5\theta_5 - \mu_3(\theta_4 - \theta_3) = 0, \tag{2.42}$$

$$\mu_2\overline{\xi}_2(\theta) - \mu_4\overline{\xi}_4(\theta) = \mu_2(\theta_3 - \theta_2) - \mu_4(\theta_5 - \theta_4) = 0. \tag{2.43}$$

For each fixed $\theta_L = (\theta_1, \theta_4) \in \mathbb{R}^2$, there exists a unique $h(\theta_L) = (\theta_2, \theta_3, \theta_5)$ such that $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ satisfies (2.41)–(2.43). Furthermore, the set Θ_L, defined as

$$\Theta_L = \left\{(\theta_1, \theta_4) \in \mathbb{R}^2_- : m_4 - \frac{\mu_5 m_4 - 1}{m_1\mu_5} < \frac{\theta_4}{\theta_1} < m_4\right\}, \tag{2.44}$$

is a nonempty and open set, and

$$h(\theta_L) = (\theta_2, \theta_3, \theta_5) < 0 \quad \text{for all} \quad \theta_L = (\theta_1, \theta_4) \in \Theta_L. \tag{2.45}$$

Proof. Fix a $\theta_L = (\theta_1, \theta_4) \in \mathbb{R}^2$. One can verify that $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ with

$$\theta_5 = \frac{1}{\mu_5 m_4 - 1}[m_4\theta_1 - \theta_4], \tag{2.46}$$

$$\theta_2 = \theta_1 - m_1\mu_5\theta_5, \tag{2.47}$$

$$\theta_3 = \theta_4 + m_3\mu_5\theta_5. \tag{2.48}$$

This satisfies Equations (2.41)–(2.43). Now for $\theta_L \in \Theta_L$, we have $\theta_4 > m_4\theta_1$ and $\theta_4 < 0$, which implies that $\theta_5 < 0$ and $\theta_3 < 0$. Finally, $\theta_2 < 0$ follows from $m_4 - (\mu_5 m_4 - 1)/(m_1\mu_5) < \frac{\theta_4}{\theta_1}$ and $\theta_1 < 0$. □

Lemma 2.6. *For each $\theta \in \mathbb{R}^5_-$,*

$$\lim_{r\to 0}(\phi^{(r)}(\theta) - \phi_k^{(r)}(\theta)) = 0, \quad k \in \{2,3,5\}. \tag{2.49}$$

Proof. We prove the lemma for $k = 2$. Other cases can be proved similarly. Recalling from (2.33) that $\xi_k^*(\theta) = \overline{\xi}_k(\theta) + \tilde{\xi}_k(\theta)$, it follows from (2.34) that for each $\theta \in \mathbb{R}^5_-$,

$$\begin{aligned}&[\mu_3\overline{\xi}_3(\theta) - \mu_1\overline{\xi}_1(\theta)]\beta_3^{(r)}(\phi^{(r)}(\theta) - \phi_3^{(r)}(\theta)) + [\mu_5\overline{\xi}_5(\theta) - \mu_3\overline{\xi}_3(\theta)]\beta_5^{(r)}(\phi^{(r)}(\theta) - \phi_5^{(r)}(\theta))\\ &+ [\mu_2\overline{\xi}_2(\theta) - \mu_4\overline{\xi}_4(\theta)]\beta_2^{(r)}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta)) = o(1).\end{aligned} \tag{2.50}$$

For each fixed $\theta_L = (\theta_1, \theta_4) \in \mathbb{R}^2_-$ and $\theta_5 \in \mathbb{R}$, set θ_2 and θ_3 follows (2.47) and (2.48), respectively. One can verify that $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ satisfies (2.41) and (2.42). Furthermore, it follows from (2.46) that one can choose $\theta_5 < 0$ small enough so that $\theta_2 < 0$, $\theta_3 < 0$ and

$$\mu_2\overline{\xi}_2(\theta) - \mu_4\overline{\xi}_4(\theta) \neq 0. \tag{2.51}$$

For this choice of $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$, (2.50) gives

$$\lim_{r\to 0}(\phi^{(r)}(\theta) - \phi_2^{(r)}(\theta)) = 0,$$

which, together with SSC (2.38), yields

$$\lim_{r\to 0}(\phi^{(r)}(\theta_L, 0) - \phi_2^{(r)}(\theta_L, 0)) = 0 \quad \text{for each} \quad \theta_L \in \mathbb{R}^2_-. \tag{2.52}$$

Now for any $\theta \in \mathbb{R}^5_-$, Equation (2.52) and SSC (2.38) imply (2.49) for $k = 2$. □

Lemma 2.7. *For each $\theta_L \in \Theta_L$, let $\theta = (\theta_L, \theta_H)$ be the unique θ that satisfies (2.41)–(2.43). Then any limit point $(\phi^*(\theta), \phi_1^*(\theta), \ldots, \phi_5^*(\theta))$ satisfies*

$$\begin{aligned} 0 &= \left(\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5} \tilde{\xi}_k(\theta)\right)\phi^*(\theta) + \mu_1 \overline{\xi}_1(\theta)(\phi^*(\theta) - \phi_1^*(\theta)) + \mu_4 \overline{\xi}_4(\theta)(\phi^*(\theta) - \phi_4^*(\theta)) \\ &= \left(\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5} \tilde{\xi}_k(\theta)\right)\phi^*(\theta_L, 0) + \mu_1 \overline{\xi}_1(\theta)(\phi^*(\theta_L, 0) - \phi_1^*(\theta_L, 0)) \\ &\quad + \mu_4 \overline{\xi}_4(\theta)(\phi^*(\theta_L, 0) - \phi_4^*(\theta_L, 0)), \qquad \theta_L \in \Theta_L. \end{aligned} \tag{2.53}$$

Proof. The first equality follows by combining Lemmas 2.5 and 2.6 with the discussion preceding Lemma 2.5. The second equality follows from the Laplace transform version of SSC (2.38). □

We now prove that for each limit point ϕ^*, (2.39) holds for some random vector (Z_1^*, Z_4^*) that is independent of the subsequence that generates the limit point.

Our starting point is (2.53) in Lemma 2.7. In (2.53), for each $\theta_L = (\theta_1, \theta_4) \in \Theta_L$, θ is set to be the vector $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ with θ_5, θ_2, and with θ_3 being defined through (2.46)–(2.48). Observe that

$$\mu_1 \overline{\xi}_1(\theta) = \mu_1(-\theta_1 + \theta_2) = -\mu_5 \theta_5 = -\frac{1}{m_4 - m_5}[m_4 \theta_1 - \theta_4] = -\langle \theta_L, R^{(1)} \rangle,$$

$$\mu_4 \overline{\xi}_4(\theta) = \mu_4(-\theta_4 + \theta_5) = -\frac{1}{m_4 - m_5}[-m_5 \theta_1 + \theta_4] = -\langle \theta_L, R^{(4)} \rangle,$$

where the 2×2 matrix R is given in (2.14), and $R^{(1)}$ and $R^{(4)}$ are the first and fourth columns of R, respectively. Also,

$$\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5} \tilde{\xi}_k(\theta) = \frac{1}{2}\left(\theta_1^2 + \sum_{k=1}^{4}(-\theta_k + \theta_{k+1})^2 + \theta_5^2\right) = \Sigma_{11}\theta_1^2 + 2\Sigma_{14}\theta_1\theta_4 + \Sigma_{44}\theta_4^2, \tag{2.54}$$

where the second equality follows from Lemma 2.8 and Σ_{11}, Σ_{14} and Σ_{44} are given by (2.15)–(2.17) with $c_{e,1}^2 = c_{s,k}^2 = 1$ for $k = 1, \ldots, 5$; the latter is true because the interarrival and service-time distributions are exponential. Therefore, (2.53) is reduced to

$$\begin{aligned} &(\Sigma_{11}\theta_1^2 + 2\Sigma_{14}\theta_1\theta_4 + \Sigma_{44}\theta_4^2)\phi^*(\theta_L, 0) + \langle \theta_L, R^{(1)} \rangle (\phi_1^*(\theta_L, 0) - \phi^*(\theta_L, 0)) + \\ &+ \langle \theta_L, R^{(4)} \rangle (\phi_4^*(\theta_L, 0) - \phi^*(\theta_L, 0)) = 0 \quad \text{for each } \theta_L = (\theta_1, \theta_4) \in \Theta_L. \end{aligned} \tag{2.55}$$

Because R in (2.14) is an $\mathcal{M}$ matrix, it follows from Proposition 5.1 and the proof of (6.3) and (6.4) in Braverman et al. (2017) that $\phi^*(0-, 0, 0, 0-, 0) = 1$, $\phi_1^*(0, 0, 0, 0-, 0) = 1$, and $\phi_4^*(0-, 0, 0, 0, 0) = 1$, where

$$\phi^*(0-, 0, 0, 0-, 0) = \lim_{\theta_1 \uparrow 0, \theta_4 \uparrow 0} \phi^*(\theta_1, 0, 0, \theta_4, 0),$$

$$\phi_1^*(0, 0, 0, 0-, 0) = \lim_{\theta_4 \uparrow 0} \phi_1^*(0, 0, 0, \theta_4, 0),$$

$$\phi_4^*(0, 0, 0, 0-, 0) = \lim_{\theta_1 \uparrow 0} \phi_4^*(\theta_1, 0, 0, 0, 0).$$

It follows that $\phi^*(\theta_1, 0, 0, \theta_4, 0)$ is the Laplace transform of a probability measure ν on $\mathbb{R}_+^2$, $\phi_1^*(0, 0, 0, \theta_4, 0)$ is the Laplace transform of a probability measure ν_1 on $\mathbb{R}_+$, and $\phi_4^*(\theta_1, 0, 0, 0, 0)$ is the Laplace transform of a probability measure ν_4 on $\mathbb{R}_+$, namely

$$\phi^*(\theta_1, 0, 0, \theta_4, 0) = \int_{\mathbb{R}_+^2} e^{\theta_1 x_1 + \theta_4 x_4} d\nu(x_1, x_4) \text{ for } (\theta_1, \theta_4) < 0$$

$$\phi_1^*(0, 0, 0, \theta_4, 0) = \int_{\mathbb{R}_+} e^{\theta_4 x_4} d\nu_1(x_4) \text{ for } \theta_4 < 0, \quad \phi_4^*(\theta_1, 0, 0, 0, 0) = \int_{\mathbb{R}_+} e^{\theta_1 x_1} d\nu_4(x_1) \text{ for } \theta_1 < 0;$$

see, for example, lemma 6.1 of Braverman et al. (2017) for an argument. Furthermore, it follows from Lemma 4.1 in Section 4 that the probability measures ν, ν_1, and ν_4 are unique. If we consider θ_1 and θ_4 as complex variables, then the aforementioned Laplace transforms have analytic extensions from Θ_L to $\mathbb{R}^2_-$. Let (Z_1^*, Z_4^*) be a random vector that has the distribution of ν. Then,

$$\phi^*(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) = \phi^*(\theta_1,0,0,\theta_4,0) = \mathbb{E}[e^{\theta_1 Z_1^* + \theta_4 Z_4^*}] \quad \text{for any} \quad \theta \in \mathbb{R}^5_-,$$

where the first equality follows the Laplace version of SSC (2.38). The uniqueness of the probability measure ν proves (2.39). □

Thus, the proof of Proposition 2.1 is completed by Lemma 2.8 below, which computes $\Sigma_{i,j}$ for $i,j = 1,4$.

Lemma 2.8. *For each $(\theta_1,\theta_4) \in \mathbb{R}^2$, let θ_2, θ_3, and θ_5 be defined through (2.46)–(2.48). Then, the quadratic equation*

$$\frac{1}{2}\left(c_{e,1}^2\theta_1^2 + \sum_{k=1}^{4} c_{s,k}^2(-\theta_k+\theta_{k+1})^2 + c_{s,5}^2\theta_5^2\right) = \Sigma_{11}\theta_1^2 + 2\Sigma_{14}\theta_1\theta_4 + \Sigma_{44}\theta_4^2 \tag{2.56}$$

holds for each $(\theta_1,\theta_4) \in \mathbb{R}^2$ if and only if Σ_{11}, Σ_{14}, and Σ_{44} are given by (2.15)–(2.17).

Proof. Because $\theta_5 = (\theta_1 - \mu_4\theta_4)/(\mu_5-\mu_4)$ and $(m_1+m_3)\mu_5 = \mu_5 - 1$ by $m_1+m_3+m_5=1$, quadratic terms in the left side of (2.56) are computed as

$$\begin{aligned}
(\theta_2-\theta_1)^2 &= m_1^2\mu_5^2\theta_5^2 = \frac{m_1^2\mu_5^2}{(\mu_5-\mu_4)^2}(\theta_1-\mu_4\theta_4)^2,\\
(\theta_3-\theta_2)^2 &= (\theta_4-\theta_1+(m_1+m_3)\mu_5\theta_5)^2 = (\theta_4-\theta_1+(\mu_5-1)\theta_5)^2\\
&= \left(\theta_4-\theta_1+\frac{\mu_5-1}{\mu_5-\mu_4}(\theta_1-\mu_4\theta_4)\right)^2 = \frac{(\mu_4-1)^2}{(\mu_5-\mu_4)^2}(\theta_1-\mu_5\theta_4)^2,\\
(\theta_4-\theta_3)^2 &= \frac{m_3^2\mu_5^2}{(\mu_5-\mu_4)^2}(\theta_1-\mu_4\theta_4)^2,\\
(\theta_5-\theta_4)^2 &= \left(\frac{\theta_1-\mu_4\theta_4}{\mu_5-\mu_4}-\theta_4\right)^2 = \frac{1}{(\mu_5-\mu_4)^2}(\theta_1-\mu_5\theta_4)^2,\\
\theta_5^2 &= \frac{1}{(\mu_5-\mu_4)^2}(\theta_1-\mu_4\theta_4)^2.
\end{aligned}$$

Hence, collecting the coefficients of θ_1^2, we have

$$\begin{aligned}
\Sigma_{11} &= \frac{1}{2}\left(c_{e,1}^2 + c_{s,1}^2\frac{m_1^2\mu_5^2}{(\mu_5-\mu_4)^2} + c_{s,2}^2\left(\frac{\mu_4-1}{\mu_5-\mu_4}\right)^2\right.\\
&\qquad \left. + c_{s,3}^2\frac{m_3^2\mu_5^2}{(\mu_5-\mu_4)^2} + c_{s,4}^2\frac{1}{(\mu_5-\mu_4)^2} + c_{s,5}^2\frac{1}{(\mu_5-\mu_4)^2}\right)\\
&= \frac{1}{2(\mu_5-\mu_4)^2}((\mu_5-\mu_4)^2c_{e,1}^2 + m_1^2\mu_5^2c_{s,1}^2 + (\mu_4-1)^2c_{s,2}^2 + m_3^2\mu_5^2c_{s,3}^2 + c_{s,4}^2 + c_{s,5}^2).
\end{aligned}$$

(2.16) and (2.17) are similarly obtained. □

2.5. General Bounded Distributions

In this section, we prove Proposition 2.1 when interarrival and service-time distributions are general. To keep our notational system simple, we further assume these distributions have bounded supports. The bounded support assumption will be replaced with a moment condition in Sections 7 and 8.

Define

$$\tilde{\eta}_1(\theta_1) = \frac{1}{2}c_{e,1}^2\theta_1^2, \quad \tilde{\xi}_k(\theta) = \frac{1}{2}c_{s,k}^2(\theta_{k+1}-\theta_k)^2 \quad k\in\{1,\ldots,4\}, \quad \tilde{\xi}_5(\theta) = \frac{1}{2}c_{s,5}^2\theta_5^2, \tag{2.57}$$

where $c_{e,1}^2$ is the SCV of the interarrival time distribution, and $c_{s,k}^2$ is the SCV of the class k service-time distribution.

The main purpose of this section is to prove the following lemma.

Lemma 2.9. *Assume that interarrival and service-time distributions have bounded supports. Then Lemma 2.7 continues to hold with $\tilde{\eta}_1(\theta_1)$ and $\tilde{\xi}_k(\theta)$ defined in (2.57) and $\overline{\eta}_1(\theta_1)$ and $\overline{\xi}_k(\theta)$ defined in (2.30)* and (2.31).

Once Lemma 2.9 is proved, the remaining steps in the proof of Proposition 2.1 for the general distribution case are the same as for the exponential case. To prove Lemma 2.9, we define, for each $\theta \in \mathbb{R}^5$, $\eta_1(\theta_1)$ and $\xi_k(\theta)$ as the solutions to

$$e^{\theta_1}\mathbb{E}(e^{-\eta_1(\theta_1)T_{e,1}}) = 1, \tag{2.58}$$

$$e^{-\theta_k+\theta_{k+1}}\mathbb{E}(e^{-\xi_k(\theta)T_{s,k}}) = 1, \quad k \in \{1,\ldots,4\}, \tag{2.59}$$

$$e^{-\theta_5}\mathbb{E}(e^{-\xi_5(\theta)T_{s,5}}) = 1. \tag{2.60}$$

We intentionally reuse the notation $\eta_1(\theta)$ and $\xi_k(\theta)$ from (2.22). This causes no harm because these two sets of definitions are identical when $T_{e,1}$ and $T_{s,k}$ are exponentially distributed. It is proved in Braverman et al. (2017) that when $T_{e,1}$ and $T_{s,k}$ have bounded support, then $\eta_1(\theta_1)$ and $\xi_k(\theta)$ are well defined for each $\theta \in \mathbb{R}^5_-$. Furthermore, the following lemma holds.

Lemma 2.10. *Taylor expansions* (2.35)–(2.36) *continue to hold with $\tilde{\eta}_1(\theta_1)$ and $\tilde{\xi}_k(\theta)$ defined in* (2.57).

Recall that we assume the sequence of two-station, five-class networks has arrival rates $\lambda_1^{(r)} = 1 - r$ and mean service times satisfying (2.9) and (2.10). Let $\kappa > 0$ be the constant such that the support of each distribution is contained in the interval $[0,\kappa]$. Recall that $X^{(r)}$ is the random vector representing the unique stationary distribution on $S \equiv \mathbb{Z}^5_+ \times [0,\kappa]^6$ of the corresponding Markov process. In the following, we use

$$x \equiv (z_1,\ldots,z_5,u_1,v_1,\ldots,v_5) \in S$$

to denote a generic state. For each $\theta \in \mathbb{R}^5_-$, define

$$f_\theta(x) = \exp(\langle\theta,z\rangle)\exp\left(-\lambda_1\eta_1(\theta_1)u_1 - \sum_{k=1}^{5}\mu_k\xi_k(\theta)v_k\right), \quad x \in S. \tag{2.61}$$

For each fixed $\theta \in \mathbb{R}^5_-$, it is clear that $f_\theta(x)$ is a bounded function of $x \in S$. For each $\theta \in \mathbb{R}^5_-$, define

$$\psi^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)})], \quad \psi_1^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)}) \mid Z_1^{(r)} = 0, Z_3^{(r)} = 0, Z_5^{(r)} = 0],$$

$$\psi_3^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)}) \mid Z_3^{(r)} = 0, Z_5^{(r)} = 0], \quad \psi_5^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)}) \mid Z_5^{(r)} = 0],$$

$$\psi_2^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)}) \mid Z_2^{(r)} = 0], \quad \psi_4^{(r)}(\theta) = \mathbb{E}[f_{r\theta}(X^{(r)}) \mid Z_2^{(r)} = 0, Z_4^{(r)} = 0].$$

Lemma 2.9 follows immediately from the following two lemmas.

Lemma 2.11. *For each $\theta_L \in \Theta_L$, let $\theta = (\theta_L,\theta_H)$ be the unique θ that satisfies* (2.41)–(2.43)*. Then, any limit point $(\psi^*(\theta),\psi_1^*(\theta),\ldots,\psi_5^*(\theta))$ of $\{(\psi^{(r)}(\theta),\psi_1^{(r)}(\theta),\ldots,\psi_5^{(r)}(\theta))\}_{r\in(0,1)}$ satisfies*

$$\left(\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5}\tilde{\xi}_k(\theta)\right)\psi^*(\theta) + \mu_1\overline{\xi}_1(\theta)(\psi^*(\theta) - \psi_1^*(\theta)) + \mu_4\overline{\xi}_4(\theta)(\psi^*(\theta) - \psi_4^*(\theta)) = 0,$$

where $\tilde{\eta}_1(\theta_1)$ and $\tilde{\xi}_k(\theta)$ are defined in (2.57) *and $\overline{\xi}_k(\theta)$ is defined in* (2.31).

This lemma is similar to Lemma 2.7, with $\psi^{(r)}(\theta)$ replacing $\phi^{(r)}(\theta)$. The proof of Lemma 2.11 will be at the end of this section after we introduce the BAR for $X^{(r)}$. The following lemma allows one to derive Lemma 2.7 from Lemma 2.11 immediately.

Lemma 2.12. *For each $\theta \in \mathbb{R}^5_-$, as $r \to 0$,*

$$\phi^{(r)}(\theta) - \psi^{(r)}(\theta) = o(1) \quad \text{and} \quad \phi_k^{(r)}(\theta) - \psi_k^{(r)}(\theta) = o(1).$$

Proof. Fix a $\theta \in \mathbb{R}^5_-$. For each $x \in S$ and each $r \in (0,1)$,

$$|g_{r\theta}(z) - f_{r\theta}(x)| = g_{r\theta}(z)\left|1 - \exp\left(-\lambda_1^{(r)}\eta_1(r\theta_1)u_1 - \sum_{k=1}^{5}\mu_k\xi_k(r\theta)v_k\right)\right|$$
$$\leq e^{|\Lambda(r,\theta,u_1,v)|}|\Lambda(r,\theta,u_1,v)| \leq e^{\Lambda(r\theta)}\Lambda(r\theta),$$

where

$$\Lambda(r,\theta,u_1,v) \equiv -\lambda_1^{(r)}\eta_1(r\theta_1)u_1 - \sum_{k=1}^{5}\mu_k\xi_k(r\theta)v_k,$$

$$\Lambda(\theta) \equiv \kappa|\eta_1(\theta_1)| + \sum_{k=1}^{5}\kappa\mu_k|\xi_k(\theta)|.$$

Therefore,

$$|\phi^{(r)}(\theta) - \psi^{(r)}(\theta)| \le e^{\Lambda(r\theta)}\Lambda(r\theta) \text{ and } |\phi_k^{(r)}(\theta) - \psi_k^{(r)}(\theta)| \le e^{\Lambda(r\theta)}\Lambda(r\theta).$$

It follows from Lemma 2.10 that

$$\lim_{r\to 0}\eta_1(r\theta_1) = 0 \text{ and } \lim_{r\to 0}\xi_k(r\theta) = 0,$$

which implies that $\lim_{r\to 0}\Lambda(r\theta) = 0$ and the lemma is proved. □

2.4.4. BAR. This section proves that for each $\theta \in \mathbb{R}_-^5$,

$$\begin{aligned}
&r^2\left(\lambda_1^{(r)}\tilde{\eta}_1(\theta_1) + \sum_{k=1}^{5}\lambda_1^{(r)}\tilde{\xi}_k(\theta)\right)\psi^{(r)}(\theta) + r^2\mu_1\xi_1^*(\theta)(\psi^{(r)}(\theta) - \psi_1^{(r)}(\theta))\\
&\quad + \mu_4 r^2\xi_4^*(\theta)(\psi^{(r)}(\theta) - \psi_4^{(r)}(\theta)) + [\mu_3\xi_3^*(r\theta) - \mu_1\xi_1^*(r\theta)]\beta_3^{(r)}(\psi^{(r)}(\theta) - \psi_3^{(r)}(\theta))\\
&\quad + [\mu_5\xi_5^*(r\theta) - \mu_3\xi_3^*(r\theta)]\beta_5^{(r)}(\psi^{(r)}(\theta) - \psi_5^{(r)}(\theta))\\
&\quad + [\mu_2\xi_2^*(r\theta) - \mu_4\xi_4^*(r\theta)]\beta_2^{(r)}(\psi^{(r)}(\theta) - \psi_2^{(r)}(\theta)) = o(r^2),
\end{aligned} \tag{2.62}$$

where $\tilde{\eta}_1(\theta_1)$ and $\tilde{\xi}_k(\theta)$ are defined in (2.57) and $\xi_k^*(\theta)$ retains the definition in (2.33). Equation (2.62) is analogous to (2.34) in the exponential case. In the exponential case, (2.34) leads to the proof of Lemma 2.7. Copying exactly the same proof, one can readily prove Lemma 2.11 from (2.62), thereby proving Proposition 2.1 for the general distribution case.

In the remainder of this section, we prove (2.62). In the following, we drop the superscript (r) everywhere to focus on one queueing network within the family of queueing networks. Recall that

$$X = (Z, U_1, V), \text{ where } Z = (Z_1, Z_2, Z_3, Z_4, Z_5) \text{ and } V = (V_1, V_2, V_3, V_3, V_5),$$

is the random vector distributed according to the stationary distribution of $\{X(t), t \ge 0\}$. We first develop a Laplace transform version of the BAR for X. Readers should be aware that the rest of this section is a simplified version of Section 6, Section 7, and Section 8. The simplification comes from the simplified notational system offered by the two-station, five-class reentrant line and the bounded support assumption on interarrival and service-time distributions.

Let $\mathcal{D}$ be the set of bounded function $f : S \to \mathbb{R}$ satisfying the following conditions: (a) $f(x)$ is bounded in $x \equiv (z, u_1, v_1, \ldots, v_5) \in S$. (b) For each fixed $z \in \mathbb{Z}_+^5$, $f(z, u_1, v_1, \ldots, v_5)$ has partial derivative from the right in u_1 and v_k, and these partial derivatives are bounded. For each $f \in \mathcal{D}$, define "interior operator"

$$\begin{aligned}
\mathcal{A}f(x) = &-\frac{\partial f}{\partial u_1}(x) - \frac{\partial f}{\partial v_1}(x)1(z_1 > 0, z_3 = 0, z_5 = 0) - \frac{\partial f}{\partial v_3}(x)1(z_3 > 0, z_5 = 0)\\
&-\frac{\partial f}{\partial v_5}(x)1(z_5 > 0) - \frac{\partial f}{\partial v_2}(x)1(z_2 > 0) - \frac{\partial f}{\partial v_4}(x)1(z_2 = 0, z_4 > 0).
\end{aligned} \tag{2.63}$$

We intend to derive a BAR corresponding to (2.18) using this operator. We first note that $\mathcal{A}f(X(t))$ is the derivative of $f(X(u))$ at $u = t$ when $X(u)$ is continuous at $u = t$. However, $f(X(t))$ may change at jump instants of $X(t)$. Taking this into account, we observe that the total change of the sample path of $f(X(u))$ from $u = 0$ to $u = t$ is

$$f(X(t)) - f(X(0)) = \int_0^t \mathcal{A}f(X(u))du + \sum_{0<u\le t}(f(X(u)) - f(X(u-)), \tag{2.64}$$

where $X(u-) = \lim_{t\uparrow u} X(t)$ is the left limit of $X(\cdot)$ at u, and there are finitely many $u \in (0,t]$ such that $f(X(u)) \neq f(X(u-))$. Because $f \in \mathcal{D}$, the summation in (2.64) is well defined.

We fix $t = 1$ in (2.64). Assuming $X(0)$ follows the stationary distribution, $\{X(u), 0 \le u \le 1\}$ is a stationary process. Taking the expectations in both sides of (2.64) and using

$$\mathbb{E}\left[\int_0^1 \mathcal{A}f(X(u))du\right] = \int_0^1 \mathbb{E}[\mathcal{A}f(X(u))]du$$

due to the boundedness of $\mathcal{A}f(X(u))$, we have

$$\mathbb{E}[\mathcal{A}f(X)] + \mathbb{E}\left[\sum_{0<u\le 1}(f(X(u)) - f(X(u-))\right] = 0, \tag{2.65}$$

where all the expectations are well defined because f and its partial derivatives are bounded.

By our convention, $X(u)$ is right continuous. Therefore, $U_1(u) > 0$ and $V_k(u) > 0$ for each $u > 0$. When $f(X(u)) \neq f(X(u-))$ at $u > 0$, at least one of the following events happens:

(a) An external arrival occurs at u, which is equivalent to $U_1(u-) = 0$ or

(b) A service completion occurs at class k, which is equivalent to $V_k(u-) = 0$, $k \in \{1,2,3,4,5\}$.

For evaluating the second expectation in (2.65), we separate different event types and define probability distributions $\mathbb{P}_{e,1}$ and $\mathbb{P}_{s,k}$ for $k \in \mathcal{K} \equiv \{1,2,\ldots,5\}$ on S^2 as, for $B \in \mathcal{B}(S^2)$,

$$\mathbb{P}_{e,1}[B] = \frac{1}{\lambda_1}\mathbb{E}\left[\sum_{0<u\le 1} 1((X(u-), X(t)) \in B)1(U_1(u-) = 0)\right], \tag{2.66}$$

$$\mathbb{P}_{s,k}[B] = \frac{1}{\lambda_1}\mathbb{E}\left[\sum_{0<u\le 1} 1((X(u-), X(t)) \in B)1(V_k(u-) = 0)\right], \qquad k \in \mathcal{K}, \tag{2.67}$$

Here $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$ are indeed probability distributions because $\mathbb{E}[\sum_{0<u\le 1} 1(U_1(u-) = 0)]$ and $\mathbb{E}[\sum_{0<u\le 1} 1(V_k(u-) = 0)]$ are the mean arrival rate of exogenous customers at station 1 and the mean departure rate at station k, respectively, and both of them are λ_1; see Lemma 6.1 for a proof. We call these distributions Palm distributions concerning the exogenous arrivals at station 1 and the departures from class k.

Denote an identity function from S^2 to S^2 by (X_-, X_+); then it can be considered as a pair of random variables taking values in S^2 on the measurable space $(S^2, \mathcal{B}(S^2))$. We consider it on the probability spaces $(S^2, \mathcal{B}(S^2), \mathbb{P}_{e,1})$ and $(S^2, \mathcal{B}(S^2), \mathbb{P}_{s,k})$. For $f \in \mathcal{D}$, let

$$\Delta f(X_-, X_+) = f(X_+) - f(X_-),$$

then we have

$$\mathbb{E}_{e,1}[\Delta f(X_-, X_+)] = \frac{1}{\lambda_1}\mathbb{E}\left[\sum_{0<u\le 1}(f(X(u-)) - f(X(t)))1(U_1(u-) = 0)\right],$$

$$\mathbb{E}_{s,k}[\Delta f(X_-, X_+)] = \frac{1}{\lambda_1}\mathbb{E}\left[\sum_{0<u\le 1}(f(X(u-)) - f(X(t)))1(V_k(u-) = 0)\right], \qquad k \in \{1,\ldots,5\},$$

where $\mathbb{E}_{e,1}$ is the expectation under the Palm distribution $\mathbb{P}_{e,1}$, and $\mathbb{E}_{s,k}$ is the expectation under the Palm distribution $\mathbb{P}_{s,k}$.

Substituting these formulas into (2.65), we have the following lemma.

Lemma 2.13. *The random vectors X and (X_-, X_+) satisfy the following BAR: for each $f \in \mathcal{D}$,*

$$\mathbb{E}[\mathcal{A}f(X)] + \lambda_1\mathbb{E}_{e,1}[\Delta f(X_+, X_-)] + \sum_{k=1}^{5}\lambda_1\mathbb{E}_{s,k}[\Delta f(X_+, X_-)] = 0. \tag{2.68}$$

Applying (2.68), it remains to evaluate expectations under the Palm distributions. From the definitions, (2.66) and (2.67), one can see that X_- represents the network state just before its jump instants under the Palm distributions, and X_+ does so just after the jump instants. More specifically, one can intuitively see that

$$X_+ = X_- + \left(e^{(1)}, \frac{1}{\lambda_1}T_{e,1}, 0\right), \qquad \text{under } \mathbb{P}_{e,1}, \tag{2.69}$$

$$X_+ = X_- + (-e^{(k)} + e^{(k+1)}1(k \le 4)), 0, m_k T_{s,k} e^{(k)}), \qquad \text{under } \mathbb{P}_{s,k}, k \in \mathcal{K}, \tag{2.70}$$

where, under $\mathbb{P}_{e,1}$, $T_{e,1}$ is independent of X_- and has the same distribution $T_{e,1}(1)$ under $\mathbb{P}$, and, under $\mathbb{P}_{s,k}$, $T_{s,k}$ is independent of X_- and has the same distribution $T_{e,1}(1)$ under $\mathbb{P}$. This representation is formally proved for the general multiclass network with SBP service discipline in Lemma 6.3. Note that X and (X_-, X_+) are defined on different probability spaces. Random vector X under $\mathbb{P}$ follows the stationary distribution of the Markov process.

Fix a $\theta \in \mathbb{R}^5_-$. For the f_θ in (2.61), one can check that $f_\theta \in \mathcal{D}$ and

$$\begin{aligned}\mathcal{A}f_\theta(x) &= f_\theta(x)\lambda_1\eta_1(\theta_1) + \mu_1\xi_1(\theta)f_\theta(x)(z_1 > 0, z_3 = 0, z_5 = 0)\\ &\quad + \mu_3\xi_3(\theta)f_\theta(x)(z_3 > 0, z_5 = 0) + \mu_5\xi_5(\theta)f_\theta(x)(z_5 > 0)\\ &\quad + \mu_2\xi_2(\theta)f_\theta(x)(z_2 > 0) + \mu_4\xi_4(\theta)f_\theta(x)(z_2 = 0, z_4 = 0).\end{aligned} \tag{2.71}$$

Setting $f = f_\theta$, it follows from (2.69), (2.70), and (2.58)–(2.60) that

$$\mathbb{E}_{e,1}[f(X_+) - f(X_-)] = 0, \qquad \mathbb{E}_{s,k}[f(X_+) - f(X_-)] = 0, \quad k = 1, 2, \ldots, 5. \tag{2.72}$$

Hence, (2.68) becomes that $\mathbb{E}[\mathcal{A}f(X)] = 0$. Define

$$\psi(\theta) = \mathbb{E}[f_\theta(X)], \quad \psi_1(\theta) = \mathbb{E}[f_\theta(X)|Z_1 = 0, Z_3 = 0, Z_5 = 0], \tag{2.73}$$

$$\psi_3(\theta) = \mathbb{E}[f_\theta(X)|Z_3 = 0, Z_5 = 0], \quad \psi_5(\theta) = \mathbb{E}[f_\theta(X)|Z_5 = 0], \tag{2.74}$$

$$\psi_2(\theta) = \mathbb{E}[f_\theta(X)|Z_2 = 0], \quad \psi_4(\theta) = \mathbb{E}[f_\theta(X)|Z_2 = 0, Z_4 = 0]. \tag{2.75}$$

Then, it follows from (2.71) that

$$\begin{aligned}&\lambda_1\eta_1(\theta_1)\mathbb{E}[f_\theta(X)] + \mu_1\xi_1(\theta)\mathbb{E}[f_\theta(X)(Z_1 > 0, Z_3 = 0, Z_5 = 0)]\\ &\quad + \mu_3\xi_3(\theta)\mathbb{E}[f_\theta(X)(Z_3 > 0, Z_5 = 0] + \mu_5\xi_5(\theta)\mathbb{E}[f_\theta(X)(Z_5 > 0)]\\ &\quad + \mu_2\xi_2(\theta)\mathbb{E}[f_\theta(X)(Z_2 > 0)] + \mu_4\xi_4(\theta)\mathbb{E}[f_\theta(X)(Z_2 = 0, Z_4 > 0)] = 0,\end{aligned} \tag{2.76}$$

which is analogous to (2.23) for the exponential case. Identical to the derivation of (2.27) from (2.23), one has

$$\begin{aligned}&\left(\lambda_1\eta_1(\theta_1) + \sum_{k=1}^{5}\lambda_1\xi_k(\theta)\right)\psi(\theta) + [\mu_5\xi_5(\theta) - \mu_3\xi_3(\theta)]\beta_5(\psi(\theta) - \psi_5(\theta))\\ &\quad + [\mu_3\xi_3(\theta) - \mu_1\xi_1(\theta)]\beta_3(\psi(\theta) - \psi_3(\theta)) + \mu_1\xi_1(\theta)(1 - \rho_1)(\psi(\theta) - \psi_1(\theta))\\ &\quad + \mu_4\xi_4(\theta)(1 - \rho_2)(\psi(\theta) - \psi_4(\theta)) + [\mu_2\xi_2(\theta) - \mu_4\xi_4(\theta)]\beta_2(\psi(\theta) - \psi_2(\theta)) = 0,\end{aligned} \tag{2.77}$$

where the β_k is the steady-state probability that all classes with priority greater or equal to class k have no customers, as defined in (2.6)–(2.8). Finally, (2.62) follows from Lemma 2.10 and (2.77) with θ being replaced by $r\theta$ and $\psi(r\theta)$ being replaced by $\psi^{(r)}(\theta)$.

3. Multiclass Queueing Networks

In this section, we introduce multiclass queueing networks that operate under SBP service disciplines. Our terminology and notation follow Bramson and Dai (2001) closely. In a multiclass queueing network, there are J service stations that process K classes of jobs, where J and K are positive integers such that $J < K$. (When $K = J$, our multiclass queueing networks become generalized Jackson networks, which were studied in Braverman et al. (2017).) Denote

$$\mathcal{J} = \{1, 2, \ldots, J\} \qquad \text{and} \qquad \mathcal{K} = \{1, 2, \ldots, K\}.$$

Each station is assumed to have a single server with unlimited waiting space. When a job arrives from outside the network, it receives service at a finite number of stations sequentially, after which it leaves the network. At any given time during its lifetime in the network, the job belongs to one of the job *classes*. It moves through the network, changing classes each time a service is completed; all jobs within a class are served at a unique station. Each job is assumed to eventually leave the network. The ordered sequence of classes that a job visits in the network is called its route; if all jobs follow the same route, the network is called a *reentrant line*. An example of a reentrant line is depicted in Figure 1.

Stations are labeled $j \in \mathcal{J}$, and classes are labeled $k \in \mathcal{K}$. We use $\mathcal{C}(j)$ to denote the set of classes belonging to station j, and $s(k)$ to denote the station to which class k belongs. Associated with each class k of a queueing network are two i.i.d. sequences of random variables, $T_{e,k}(\cdot) = \{T_{e,k}(i), i \geq 1\}$ and $T_{s,k}(\cdot) = \{T_{s,k}(i), i \geq 1\}$, one i.i.d. sequence of $\mathbb{R}^K$-valued random vectors $\Phi^{(k)}(\cdot) = \{\Phi^{(k)}(i), i \geq 1\}$, and two real numbers, $a_k \geq 0$ and $m_k > 0$.

We assume that the $3K$ sequences

$$T_{e,1}(\cdot), \ldots, T_{e,K}(\cdot),\ T_{s,1}(\cdot), \ldots, T_{s,K}(\cdot), \Phi^{(1)}(\cdot), \ldots, \Phi^{(K)}(\cdot) \tag{3.1}$$

are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and are mutually independent. We use $T_{e,k}$, $T_{s,k}$, and $\Phi^{(k)}$ to denote generic random element in sequences $T_{e,k}(\cdot)$, $T_{s,k}(\cdot)$, and $\Phi^{(k)}(\cdot)$, respectively. We assume that $T_{e,k}$ and $T_{s,k}$ are *unitized*; that is, $\mathbb{E}[T_{e,k}] = 1$ and $\mathbb{E}[T_{s,k}] = 1$, and $\Phi^{(k)}$ takes values in $\{e^{(0)}, e^{(\ell)}, \ell \in \mathcal{K}\}$, where $e^{(\ell)}$ is the K-vector with component ℓ being one and all other components being zero, and $e^{(0)}$ is the K-vector of zeros. For each i, $a_k T_{e,k}(i)$ will denote the interarrival time between the $(i-1)$th and the ith *externally* arriving job at class k, $m_k T_{s,k}(i)$ will denote the *service* time for the ith class k job, and $\Phi^{(k)}(i) = e^{(\ell)}$ means that the job that completes the ith class k service will join next as a class ℓ job for $\ell \in \mathcal{K}$ or will exit the network when $\ell = 0$.

Let $\lambda_k = 1/a_k$ and $\mu_k = 1/m_k$ for each class k. Then λ_k is the external arrival rate to class k, and m_k is the mean service time for class k jobs. We allow $\lambda_k = 0$ for some classes k, in which case class k has no external arrivals. We set

$$\mathcal{E} = \{k \in \mathcal{K} : \lambda_k \neq 0\}, \qquad E = |\mathcal{E}|,$$

where $|A|$ denotes the number of elements of a set A. We assume that there exists a $\delta_0 > 0$ such that

$$\mathbb{E}[T_{e,k}^{2+\delta_0}] < \infty, \quad k \in \mathcal{E}, \quad \text{and} \quad \mathbb{E}[T_{s,k}^{2+\delta_0}] < \infty \quad k \in \mathcal{K}, \tag{3.2}$$

and set

$$c_{e,k}^2 = \mathrm{Var}(T_{e,k}), \quad k \in \mathcal{E}, \qquad c_{s,k}^2 = \mathrm{Var}(T_{s,k}), \quad k \in \mathcal{K}.$$

Thus, $c_{e,k}^2$ and $c_{s,k}^2$ are the squared coefficients of variation for interarrival and service times. Let

$$P_{k\ell} = \mathbb{P}\{\Phi^{(k)} = e^{(\ell)}\}, \quad \ell \in \mathcal{K}, \quad P_{k0} = \mathbb{P}\{\Phi^{(k)} = e^{(0)}\} = 1 - \sum_{\ell \in \mathcal{K}} P_{k\ell}. \tag{3.3}$$

The $K \times K$ matrix $P = (P_{k\ell})$ is the *routing matrix* of the network. We assume our networks are *open*; that is, the matrix $(I - P)$ is invertible where I denotes the identity matrix. We note that P_{k0} in (3.3) is the probability of a job leaving the network after completing a class k service.

3.1. Service Discipline

A service discipline dictates the order in which jobs are served at each station. A service discipline is said to be *nonidling* if a server is always active when there are jobs waiting to be served at its station. In this paper, we restrict our discipline to SBP, which is defined later. Under an SBP discipline, the classes at each station are assigned a fixed ranking. When the server switches from one job to another, the new job will be taken from the leading (or longest-waiting) job at the highest-ranking nonempty class at the server's station. We assume that the ranking is strict; that is, there is no tie in the ranking. We also assume that the service discipline is *preemptive-resume*. That is, when a job with a higher rank than the one currently being served arrives at the server's station, the service of the current job is interrupted. When service of all jobs with higher ranks is completed, the interrupted service continues from where it left off.

Two SBP disciplines for reentrant lines that have been studied in the literature are FBFS and LBFS. Under the FBFS discipline, earlier classes along the route are assigned higher priorities. Under the LBFS discipline, later classes along the route are assigned higher priorities. For the two-station, five-class reentrant line pictured in Figure 1, we have $\mathcal{K} = \{1, 2, 3, 4, 5\}$, $\mathcal{E} = \{1\}$, $\mathcal{L} = \{1, 4\}$ and $\mathcal{H} = \{2, 3, 5\}$.

3.2. Notation Facilitating an SBP Discipline

For each class $k \in \mathcal{K}$, denote

$$H(k) \tag{3.4}$$

as the set of classes at station $s(k)$ whose priorities are at least as high as class k. Let

$$H_+(k) = H(k) \setminus \{k\} \tag{3.5}$$

be the set of classes at station $s(k)$ whose priorities are strictly higher than class k. $H_+(k)$ is empty when class k has the highest priority at station $s(k)$. Under our preemptive-resume priority discipline, class k jobs are processed only when there are no class ℓ jobs for *all* $\ell \in H_+(k)$. For each station $j \in \mathcal{J}$, define $\ell(j)$ to be the lowest-priority class at station j, and define $h(j)$ to be highest-priority class at station j. Define

$$\mathcal{K}_1 = \{h(1), h(2), \dots, h(J)\} \quad \text{and} \quad \mathcal{L} = \{\ell(1), \ell(2), \dots, \ell(J)\} \tag{3.6}$$

to be the sets of the highest and lowest classes, respectively, and

$$\mathcal{H} = \mathcal{K} \setminus \{\ell(1), \dots, \ell(J)\} \tag{3.7}$$

to be the set of "high priority" classes that *exclude* all the lowest priority classes. Clearly, $\mathcal{H} \cap \mathcal{L} = \emptyset$, but $\mathcal{K}_1 \cap \mathcal{L}$ is *not* necessarily empty as there may be stations serving only one job class. For a subset $\mathcal{C} \subset \mathcal{K}$, let $|\mathcal{C}|$ be its cardinality, that is, the number of its elements. Let $H = |\mathcal{H}|$ and $L = |\mathcal{L}|$. The latter is also equal to J.

For each class $k \in \mathcal{H}$, define

$$k- \text{ to be the highest class in } \{\ell \in \mathcal{K}; s(\ell) = s(k)\} \setminus H(k), \tag{3.8}$$

$$k+ \text{ to be the lowest class in } H_+(k), \text{ namely, in } H(k) \setminus \{k\}. \tag{3.9}$$

When $k-$ and $k+$ are undefined, the quantities indexed by them are explained there.

3.3. Traffic Equations

To investigate open multiclass queueing networks, one uses the solution α_ℓ, $\ell \in \mathcal{K}$, of the *traffic equations*

$$\alpha_k = \lambda_k + \sum_{\ell \in \mathcal{K}} \alpha_\ell P_{\ell k}, \qquad k \in \mathcal{K}, \tag{3.10}$$

or equivalently, in vector form, of $\alpha = \lambda + P^T \alpha$. All vectors in this paper are to be interpreted as column vectors unless we explicitly state otherwise. Because the network corresponding to P is open, the unique solution to (3.10) is $\alpha = (I - P^T)^{-1}\lambda$. The term α_k is referred to as the *nominal total arrival rate* at class k; it depends on both external and internal arrivals. If, for each class k, there is a long-run average rate of flow into the class that is equal to the long-run average rate out of that class, this rate will equal α_k.

Using m and α, one defines the *traffic intensity* ρ_j for the jth server as

$$\rho_j = \sum_{k \in \mathcal{C}(j)} \gamma_k, \quad \text{where} \tag{3.11}$$

$$\gamma_k = \alpha_k m_k. \tag{3.12}$$

In vector form, ρ is given by $\rho = CM\alpha$, where $M = \text{diag}(m)$ and C is the $\mathcal{J} \times \mathcal{K}$ *constituency matrix*

$$C_{jk} = \begin{cases} 1 & \text{if } \; j = s(k), \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

In our study, it is convenient to replace $\mathcal{J}$ with $\mathcal{L}$ using the fact that $\mathcal{J}$ and $\mathcal{L}$ have the same cardinality.

(For a d-dimensional vector x, diag(x) denotes the $d \times d$ matrix whose diagonal entries are given by the components of x and all other entries are zero.) When $\rho_j \le 1$, ρ_j is also referred to as the nominal fraction of time that server j is busy. In this paper, we are interested in networks in which ρ_j is close to one for each station j. Such networks are said to be "heavily loaded." The precise meaning of that term will be defined in Section 5.

3.4. Markov Process

At time $t \ge 0$, for $k \in \mathcal{K}$, let $Z_k(t)$ be the number of class k jobs including possibly the one in service, and let $R_{s,k}(t)$ be the remaining service time of a class k job in service or the service time of the next class k job if $Z_k(t) = 0$. For $k \in \mathcal{E}$, let $R_{e,k}(t)$ be the remaining time for a class k job to externally arrive. Let $Z(t), R_e(t), R_s(t)$ be the random vectors whose kth entries are $Z_k(t), R_{e,k}(t), R_{s,k}(t)$, respectively. Let

$$X(t) \equiv (Z(t), R_e(t), R_s(t)), \qquad t \ge 0. \tag{3.14}$$

Let $\mathbb{R}_+^{\mathcal{E}}$ be the set of all vectors $(u_k; k \in \mathcal{E})$ for $u_k \geq 0$, and let $X(\cdot) = \{X(t), t \geq 0\}$. Then, when dropping the bounded supports assumption on the interarrival and service-time distributions, $X(t)$ has state space

$$S \equiv \mathbb{Z}_+^K \times \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}_+^K,$$

and $X(\cdot)$ is a Markov process with respect to the filtration $\mathbb{F}^X \equiv \{\mathcal{F}_t^X; t \geq 0\}$, where $\mathcal{F}_t = \sigma(\{X(u); 0 \leq u \leq t\})$. We here assume that $X(\cdot)$ is right continuous on $[0, \infty)$ and has a limit from the left in $(0, \infty)$. When interarrival and service-time distributions are exponential, because of the memoryless property of an exponential distribution, $\{Z(t), t \geq 0\}$ itself is a continuous-time Markov chain with (discrete) state space $\mathbb{Z}_+^K$.

A distribution π on S is said to be a stationary distribution of the Markov process $X(\cdot)$ if $X(t)$ follows distribution π for any $t > 0$ when $X(0)$ is initialized with distribution π. A necessary condition for the existence of a stationary distribution is

$$\rho_j < 1, \quad j \in \mathcal{J}. \tag{3.15}$$

(See, for example, theorem 5.2 of Dai and Harrison (2020) for a proof when all distributions are phase type.) Dai (1995) provides a sufficient condition for the existence of a stationary distribution. The condition is in terms of the stability of a fluid model corresponding to the queueing network.

4. Basic Adjoint Relationship of an SRBM

In this paper, SRBMs are not used explicitly, but they are in the background somewhat prominently. For the definition of an SRBM, see, for example, section 2.3 of Braverman et al. (2017) or definition 3.1 in Bramson and Dai (2001). Recall that the queueing network defined in Section 3 has K classes and J stations. In the following, $\mathcal{L}$ can be any subset of $\mathcal{K}$ and $L = |\mathcal{L}|$. To be specific, the set $\mathcal{L} \subset \mathcal{K}$ is the lowest classes at stations in the queueing network. Therefore, $L = J$ the number of stations in the network. We use $\mathbb{R}^{\mathcal{L}}$ to denote the L-dimensional Euclidean space; for a vector $x = (x_\ell) \in \mathbb{R}^{\mathcal{L}}$, its components x_ℓ are indexed by $\ell \in \mathcal{L}$. Similarly, for an $\mathcal{L} \times \mathcal{L}$ matrix $A = (A_{ij})$, its entries A_{ij} are indexed by $i, j \in \mathcal{L}$. For a subset $\mathcal{C} \subset \mathcal{L}$, matrix $(A_{ij}, i, j \in \mathcal{C})$ is called a principal submatrix of A.

Definition 4.1 (Completely-$\mathcal{S}$ Matrix). Let R be an $\mathcal{L} \times \mathcal{L}$ matrix. Then R is called an $\mathcal{S}$ matrix if there exists $u \in \mathbb{R}_+^{\mathcal{L}}$ such that $Ru > 0$ (vector inequalities are to be interpreted componentwise). The matrix R is said to be *completely* $\mathcal{S}$ if each principal submatrix of R is an $\mathcal{S}$ matrix.

Given a finite measure ν on $\mathbb{R}_+^{\mathcal{L}} \equiv \{x \in \mathbb{R}^{\mathcal{L}} : x \geq 0\}$ whose total mass is not greater than 1, that is, ν is a subprobability distribution, let ϕ be its Laplace transform. Namely,

$$\phi(\theta) = \int_{\mathbb{R}_+^{\mathcal{L}}} e^{\langle \theta, x \rangle} \nu(dx), \quad \theta \in \mathbb{R}_-^{\mathcal{L}}, \tag{4.1}$$

where $\langle x, y \rangle = \sum_{i \in \mathcal{L}} x_i y_i$ is the inner product of of vectors $x, y \in \mathbb{R}^{\mathcal{L}}$.

The following lemma follows from the uniqueness of the stationary distribution of an SRBM in Dai and Kurtz (1994). The current form follows from lemma 2.1 of Braverman et al. (2017) and the appendix of the arXiv version of Dai et al. (2014).

Lemma 4.1. *Given an $\mathcal{L} \times \mathcal{L}$ positive definite matrix Σ, an $\mathcal{L} \times \mathcal{L}$ completely-$\mathcal{S}$ matrix R, and a positive $\mathcal{L}$-vector b, there is at most one set of probability measures ν and ν_j, $j \in \mathcal{L}$, such that ν_j has the support in $\{x \in \mathbb{R}_+^{\mathcal{L}} : x_j = 0\}$ for each $j \in \mathcal{L}$ and for Laplace transforms ϕ and ϕ_ℓ of ν and ν_ℓ, respectively,*

$$\langle \theta, \Sigma \theta \rangle \phi(\theta) + \sum_{\ell \in \mathcal{L}} b_\ell \langle \theta, R^{(\ell)} \rangle (\phi_\ell(\theta) - \phi(\theta)) = 0, \qquad \theta \in \mathbb{R}_-^{\mathcal{L}}, \tag{4.2}$$

where $R^{(\ell)}$ is the ℓth column of matrix R for $\ell \in \mathcal{L}$.

When the probability measure ν in the lemma exists, it is the unique stationary distribution of an SRBM with reflection matrix R, covariance matrix Σ, and drift vector $-Rb$. In what follows, we say ν is the distribution uniquely determined by the set of parameters (R, Σ, b).

In our application of Lemma 4.1, ϕ and ϕ_ℓ are obtained as the limits of a sequence of the Laplace transforms of probability distributions. Namely, they are the Laplace transforms of vague limits of the probability distributions. Hence, they may not be Laplace transforms of probability distributions. Thus, for successfully using Lemma 4.1, we need to verify that those ϕ and ϕ_ℓ are the Laplace transforms of probability distributions. We introduce a notion of a tight system for this verification.

Definition 4.2 (Tight System). Given an $\mathcal{L}\times\mathcal{L}$ matrix R and a positive $\mathcal{L}$-vector b, we say that (R, b) is a tight system if the set of linear equations and inequalities:

$$\begin{aligned}
&\sum_{j\in\mathcal{L}} b_j R_{ij}(x_A^{(j)} - x_A) = 0, && i\in A,\ A\subset\mathcal{L},\\
&x_A, x_A^{(j)} \in [0,1], && A\subset A'\subset\mathcal{L},\\
&x_A \ge x_{A'},\ x_A^{(j)} \ge x_{A'}^{(j)}, && j\in\mathcal{L}, A\subset A'\subset\mathcal{L},\\
&x_A^{(j)} = x_{A\setminus\{j\}}^{(j)}, && j\in A\subset\mathcal{L},\\
&x_\emptyset = x_\emptyset^{(j)} = 1, && j\in\mathcal{L}.
\end{aligned} \tag{4.3}$$

This system has a unique solution $x_A = x_A^{(j)} = 1$ for all $j\in\mathcal{L}$ and $A\subset\mathcal{L}$. □

The meaning of Condition (4.3) of tight system can be seen through the following lemma, which is proved similarly to lemma 5.1 of Braverman et al. (2017). For completeness, we prove it in Appendix A.

Lemma 4.2. *Let ϕ and ϕ_ℓ for $\ell\in\mathcal{L}$ be the Laplace transforms of finite measures on $\mathbb{R}_+^{\mathcal{L}}$ that satisfy* (4.2). *Then,* (i) *for $A\subset\mathcal{L}$,*

$$\phi_A(0-) = \lim_{\theta\uparrow 0}\phi(\theta_A), \qquad \phi_{A,\ell}(0-) = \lim_{\theta\uparrow 0}\phi_\ell(\theta_A), \quad \ell\in\mathcal{L}. \tag{4.4}$$

These are well defined, where θ_A is the L-dimensional vector θ whose entries $\ell\notin A$ are replaced by zero. (ii) *If (R, b) is a tight system, then $(x_A, x_A^{(\ell)}) \equiv (\phi_A, \phi_{A,\ell})$ for $A\subset\mathcal{L}$ satisfy condition* (4.3), *and therefore there exist unique probability distributions whose Laplace transforms ϕ and ϕ_ℓ satisfy* (4.2).

All our work is to find ϕ and ϕ_ℓ satisfying (4.2) and to verify (R, b) to be a tight system by Lemma 4.2. We list sufficient conditions for this tightness.

(4.a) If R is an $\mathcal{M}$ matrix and $b>0$, then (R, b) is a tight system for any b as long as $b>0$, where an $\mathcal{L}\times\mathcal{L}$ matrix is said be an $\mathcal{M}$ matrix if it is invertible and has nonpositive off-diagonal entries and positive diagonal entries.

(4.b) For $L=2$, $R\equiv\{R_{i,j}; i.j=1,2\}$ with $R_{i,i}>0$ for $i=1, 2$ and $b>0$ is a tight system if and only if one of the following conditions holds: (4b.1) $R_{12}\le 0$ and $R_{21}\le 0$, (4b.2) $R_{12}<0, R_{21}\ge 0$, and (4b.3) $R_{12}\ge 0, R_{21}<0$.

(4.c) (R, b) for any reentrant line operating LBFS is tight.

Here, (4.a) follows from the proof of proposition 5.1 of Braverman et al. (2017), whereas (4.b) and (4.c) are proved in Dai et al. (2024).

5. Heavy Traffic Assumption and the Main Result

In this section, we first introduce five assumptions that will be used in our main theorem. We will then state the main theorem. Finally, we discuss reasons for making these assumptions after the statement of the main theorem.

We consider a sequence of multiclass networks with SBP service discipline indexed by $r\in(0,1]$, where r monotonically tends to 0. (With some abuse of notation, we refer to such networks as a sequence of networks.) For the rth network, let $a_k^{(r)}T_{e,k}(n)$ and $m_k^{(r)}T_{s,k}(n)$ be the nth interarrival time of exogenous class k arrivals and the nth service times of class k customers, respectively, where $T_{e,k}(\cdot)\equiv\{T_{e,k}(i), i\ge 1\}$ and $T_{s,k}(\cdot)\equiv\{T_{s,k}(i), i\ge 1\}$ are independent sequences of i.i.d. random variables introduced in (3.1). Thus, we use the same primitive increments for the entire sequence of queueing networks. In a more general setting, these families of variables are given by *triangular* arrays of random variables, where the underlying $T_{e,k}^{(r)}(i), T_{s,k}^{(r)}(i)$ vary with r. Heavy-traffic limit theorems under this more general setup are robust under perturbations of the interarrival and service vectors. The purpose of the present setup is to keep the notation simple. Because Braverman et al. (2017) use the framework of triangular arrays, our main result, Theorem 5.1, can be generalized straightforwardly to that setting.

We assume the following moment condition on interarrival and service-time distributions.

Assumption 5.1. *Assume Condition* (3.2) *is satisfied. Namely, interarrival and service times have finite $2+\delta_0$ moments for some $\delta_0>0$.*

For $r\in(0,1]$ and $k\in\mathcal{K}$, let

$$\lambda_k^{(r)} = 1/a_k^{(r)}, \qquad \mu_k^{(r)} = 1/m_k^{(r)},$$

which are the exogenous arrival and service rates of class k customers, respectively. The squared coefficients of variation of the interarrival times and service times for class k, $c_{e,k}$ and $c_{s,k}$, do not depend on the index r. We assume that $\{k \in \mathcal{K}; \lambda_k^{(r)} > 0\}$ does not depend on $r \in (0,1]$ and is also denoted by $\mathcal{E}$. We also assume that the routing matrix P and the priority order do not depend on $r \in (0,1]$.

With vectors $\lambda^{(r)} \equiv (\lambda_k^{(r)}; k \in \mathcal{E})$ and $m^{(r)} \equiv (m_k^{(r)}; k \in \mathcal{K})$ replacing λ and m, define $\alpha_k^{(k)}$, $\gamma_k^{(r)}$, and $\rho_j^{(r)}$, following (3.10), (3.12), and (3.11), respectively, for $k \in \mathcal{K}$ and $j \in \mathcal{J}$. In vector form,

$$\alpha^{(r)} = (I - P^T)^{-1}\lambda^{(r)}, \qquad M^{(r)} = \text{diag}(m^{(r)}), \quad \gamma^{(r)} = M^{(r)}\alpha^{(r)}, \qquad \rho^{(r)} = C\gamma^{(r)}, \tag{5.1}$$

where C is the constituency matrix defined in (3.13). Define

$$\beta_k^{(r)} = 1 - \sum_{\ell \in H(k)} \gamma_\ell^{(r)}, \qquad k \in \mathcal{K}, \tag{5.2}$$

which is the probability that class k customers can be served. Recall that $\mathcal{L}$ is the set of low-priority classes defined in (3.6), and $s(k)$ is the station at which class k customers get service. Because $H(\ell) = \{k \in \mathcal{K}; s(k) = \ell\}$ for $\ell \in \mathcal{L}$, it follows from (5.1) and (5.2) that

$$\beta_\ell^{(r)} = 1 - \rho_{s(\ell)}^{(r)}, \qquad \ell \in \mathcal{L}. \tag{5.3}$$

Assumption 5.2. *We assume that there are K-vectors $\lambda \geq 0$, λ^*, $m > 0$ and m^* with $\lambda_k = \lambda_k^* = 0$ for $k \notin \mathcal{E}$ such that, for index $r \in (0,1]$,*

$$\lambda_k^{(r)} = \lambda_k - r\lambda_k^* > 0 \quad \textit{for } k \in \mathcal{E}, \qquad m_k^{(r)} = m_k - rm_k^* > 0 \quad \textit{for } k \in \mathcal{K}, \tag{5.4}$$

$$\rho = CM\alpha = e, \tag{5.5}$$

$$c \equiv C[\text{diag}(m^*)\alpha + \text{diag}(m)\alpha^*] > 0, \tag{5.6}$$

where $M = \text{diag}(m)$, e is the $\mathcal{J}$-vector of all 1's, $\alpha = (I - P^T)^{-1}\lambda$, $\alpha^* = (I - P^T)^{-1}\lambda^*$.

Condition (5.5) implies $\rho^{(r)} \to e$, which says that each station is *critically loaded* in the limit as $r \downarrow 0$. We do not impose a sign restriction on the "deviation vectors" $\lambda^* \in \mathbb{R}^K$ and $m^* \in \mathbb{R}^K$ in (5.4). One can check that

$$\rho^{(r)} = e - rc + r^2 C\, \text{diag}(m^*)\alpha^*. \tag{5.7}$$

We do require $c > 0$ in Condition (5.6) to make sure $\rho^{(r)} < e$, a necessary condition for the rth network to be stable, when $r > 0$ is small enough. Recall there is a one-to-one map between $\mathcal{J}$ and $\mathcal{L}$. Both ρ and c are $\mathcal{J}$-vectors. For notational convenience in the rest of the paper, we convert the $\mathcal{J}$-vector c into an equivalent $\mathcal{L}$-vector b via

$$b_\ell = c_{s(\ell)} \quad \text{for } \ell \in \mathcal{L}. \tag{5.8}$$

For the reentrant line considered in Section 2, the heavy-traffic conditions (5.4)–(5.6) are satisfied with $b_1 = b_4 = 1$.

We are concerned with the sequence of the multiclass queueing networks with SBP service disciplines that are indexed by $r \in (0,1]$. Denote the Markov process describing this network with index r by $X^{(r)}(\cdot) \equiv \{X^{(r)}(t); t \geq 0\}$, where

$$X^{(r)}(t) = (Z^{(r)}(t), R_e^{(r)}(t), R_s^{(r)}(t)) \in S \equiv \mathbb{Z}_+^K \times \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}_+^K, \qquad t \geq 0. \tag{5.9}$$

We are interested in the stationary distributions of the Markov process $X^{(r)}(\cdot)$. This motivates us to make the following assumption.

Assumption 5.3. *For each index $r \in (0,1]$, $X^{(r)}(\cdot)$ has a unique stationary distribution.*

It is well known that $\rho^{(r)} < e$ is not sufficient for Assumption 5.3 to hold. For example, for the two-station, five-class reentrant line in Section 2, we remarked that the virtual station stability condition (2.5) is needed in addition to $\rho^{(r)} < e$ for Assumption 5.3 to be satisfied.

For each $r \in (0,1]$, denote an S-valued random variable subject to the stationary distribution $\pi^{(r)}$ of $X^{(r)}(\cdot)$ by $X^{(r)} \equiv (Z^{(r)}, R_e^{(r)}, R_s^{(r)})$. Our main objective is to study the weak limit of the distribution of $rZ^{(r)}$ as $r \downarrow 0$ under the heavy-traffic condition. We wish to prove

$$rZ^{(r)} \Rightarrow Z^* = (Z_L^*, 0_H) \quad \text{as } r \to 0, \tag{5.10}$$

where "⇒" denotes convergence in distribution, and, for a vector $z \in \mathbb{R}^K$, we define $z_L = (z_k; k \in \mathcal{L}) \in \mathbb{R}^{\mathcal{L}}$ and $z_H = (z_k; k \in \mathcal{H}) \in \mathbb{R}^{\mathcal{H}}$. Readers are warned that we have abused the notation in (5.10) by adopting the convention that

$$z = (z_L, z_H) \tag{5.11}$$

for a vector $z \in \mathbb{R}^K$. Recall that vectors are considered column vectors unless stated otherwise. Thus z, z_L, and z_H are all column vectors with appropriate dimensions. We believe Notation (5.11) is more attractive than the cumbersome expression $z = (z_L^T, z_H^T)^T$, where the superscript ^{T} denotes transpose.

In (5.10), $rZ_H^{(r)} \Rightarrow 0$. This is an example of state space collapse (SSC) under the priority service discipline. The SSC is somewhat expected because heavy-traffic Conditions (5.4)–(5.6) imply that

$$\beta_k^{(r)} \to \beta_k = 1 - \sum_{\ell \in H(k)} \alpha_\ell m_\ell > 0, \quad k \in \mathcal{H},$$

which means that each server has excess capacity after serving all high-priority jobs. Therefore, the load from high-priority jobs is *not* in heavy traffic, and job counts in high-priority classes should not blow up even though the entire network goes into heavy traffic. However, the preceding intuition holds only when Assumption 5.3 holds, and proving SSC in steady state is an independent task, which can be difficult. In this paper, we assume the following.

Assumption 5.4. *For each $k \in \mathcal{H}$, the collection of the steady-state job-count vectors $\{Z_k^{(r)}; r \in (0,1]\}$ is uniformly integrable, namely,*

$$\lim_{a \to \infty} \sup_{r \in (0,1]} \mathbb{E}[Z_k^{(r)} 1(Z_k^{(r)} > a)] = 0. \tag{5.12}$$

As a consequence,

$$\sup_{r \in (0,1]} \mathbb{E}[Z_k^{(r)}] < \infty \quad \text{and} \quad \lim_{r \to 0} \mathbb{E}[Z_k^{(r)} 1(Z_k^{(r)} \geq h(r))] = 0, \quad k \in \mathcal{H} \tag{5.13}$$

for any $h : (0,1] \to \mathbb{R}_+$ satisfying $\lim_{r \to 0} h(r) = \infty$.

It is well known that

$$\sup_{r \in (0,1]} \mathbb{E}[(Z_k^{(r)})^{1+\delta}] < \infty, \tag{5.14}$$

for some $\delta > 0$, implies that the collection is uniformly integrable. Cao et al. (2022) develops a sufficient condition to prove (5.14) based on SSC of corresponding fluid models.

As stated in the main theorem, the random vector $Z_L^* \in \mathbb{R}_+^{\mathcal{L}}$ in (5.10) has the stationary distribution ν of an SRBM. For a given set of parameters (R, Σ, b), such a stationary distribution ν is characterized in Lemma 4.1. To state the theorem, we need to define two $\mathcal{L} \times \mathcal{L}$ matrices R and Σ. By imposing conditions on R and Σ, we will argue that probability measures ν and ν_j on $\mathbb{R}_+^{\mathcal{L}}$ for $j \in \mathcal{L}$ corresponding to (R, Σ, b) in Lemma 4.1 exist and are unique, where $b > 0$ is the vector in heavy-traffic Conditions (5.6) and (5.8).

To define R, let

$$A = (I - P^T)\text{diag}(\mu)(I - B), \tag{5.15}$$

and B is the $K \times K$ matrix defined by

$$B_{k\ell} = \begin{cases} 1 & \text{if } \ell = k+, \\ 0 & \text{otherwise,} \end{cases} \tag{5.16}$$

for classes $\ell, k \in \mathcal{K}$, where we call that $k+$ is the lowest class in $H(k) \setminus \{k\}$ (see (3.9)). Define matrices A_L, A_{LH}, A_{HL}, and A_H as

(i) A_L and A_H are the principal submatrices corresponding the index set $\mathcal{L} \subset \mathcal{K}$ and $\mathcal{H} \subset \mathcal{K}$, respectively.
(ii) A_{LH} and A_{HL} are corresponding other blocks of A.

Thus, if the indexes of customer classes are appropriately chosen, then we can write A as

$$A = \begin{pmatrix} A_L & A_{LH} \\ A_{HL} & A_H \end{pmatrix}. \tag{5.17}$$

Assumption 5.5. *The matrix A_H is assumed to be invertible. Furthermore, define $\mathcal{L} \times \mathcal{L}$ matrix R via*

$$R = A_L - A_{LH} A_H^{-1} A_{HL}; \tag{5.18}$$

then R is completely $\mathcal{S}$, and (R, b) is a tight system as defined in Definition 4.2.

To define the $\mathcal{L} \times \mathcal{L}$ matrix Σ, let

$$\begin{aligned} q(\theta) = {} & \frac{1}{2} \sum_{k \in \mathcal{E}} \lambda_k c_{e,k}^2 \theta_k^2 \\ & + \frac{1}{2} \sum_{k \in \mathcal{K}} \alpha_k \left[\sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell^2 - \left(\sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell \right)^2 + c_{s,k}^2 \left(\theta_k - \sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell \right)^2 \right], \quad \theta \in \mathbb{R}^K. \end{aligned} \tag{5.19}$$

By Jensen's inequality, $\sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell^2 - (\sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell)^2 \geq 0$; thus, $q(\theta)$ is a nonnegative quadratic function of $\theta \in \mathbb{R}^K$. For each (column) vector $\theta_L \in \mathbb{R}^{\mathcal{L}}$, define vector

$$\theta_H = -(A_H^{-1})^T (A_{LH})^T \theta_L. \tag{5.20}$$

With θ_H defined through Equation (5.20), it is clear that $q(\theta_L, \theta_H)$ is a nonnegative quadratic function of $\theta_L \in \mathbb{R}^{\mathcal{L}}$, where (θ_L, θ_H) is the K-vector θ following Convention (5.11). Therefore, there is a nonnegative definite $\mathcal{L} \times \mathcal{L}$ symmetric matrix Σ such that

$$q(\theta_L, \theta_H) = \langle \theta_L, \Sigma \theta_L \rangle, \qquad \theta_L \in \mathbb{R}^{\mathcal{L}}. \tag{5.21}$$

Theorem 5.1. *Assume Assumptions 5.1–5.5 hold. Assume further that Σ defined in (5.21) is positive definite. Then the heavy-traffic steady-state convergence (5.10) holds, where Z_L^* is a random vector whose distribution is uniquely determined from (4.2) in Lemma 4.1 with parameters (R, Σ, b) defined in (5.18), (5.21), and (5.8).*

We used Assumption 5.1 in this theorem for simplicity in the exposition. This assumption can be replaced by the following weaker one.

Assumption 5.1A. *All interarrival distributions and the service-time distributions of classes $k \in \mathcal{K}_1$ have finite second moments, whereas the service-time distributions of class $k \in \mathcal{K} \setminus \mathcal{K}_1$ have finite $(2 + \delta_0)$th moments for some $\delta_0 > 0$, where the set of highest classes $\mathcal{K}_1$ is defined in (3.6).*

We explain in Appendix B the reason that Assumption 5.1 in Theorem 5.1 can be replaced by Assumption 5.1A. One may wonder whether Assumption 5.1A can be replaced by a further weaker assumption.

Assumption 5.1B. *All the interarrival and service-time distributions have finite second moments.*

At this point, we could not verify this replacement, and we leave it as a future research topic.

Specializing Theorem 5.1 to reentrant lines, we have the following corollary. For a reentrant line, without loss of generality, we assume $\mathcal{E} = \{1\}$.

Corollary 5.1. *Consider a sequence of reentrant lines in the setting of this section that satisfies Assumption 5.2. Assume that*

$$\mathbb{E}(T_{e,1})^{2+\delta_0} < \infty, \quad \mathbb{E}(T_{s,k})^{2+\delta_0} < \infty, \quad k \in \mathcal{K} \tag{5.22}$$

for some $\delta_0 > 0$. Then, matrix Σ is well defined through (5.21). Assume Σ is positive definite. Assume further that $T_{e,1}$ is "unbounded" and "spread out" as defined in (1.4) *and* (1.5) *of Dai (1995). Then,*

(a) *The heavy-traffic steady-state convergence (5.10) holds for reentrant lines under LBFS service discipline.*

(b) *The heavy-traffic steady-state convergence (5.10) holds for the two-station, five-class reentrant line in Section 2 operating under SBP Discipline (2.1) if Condition (2.11) is satisfied.*

Proof. Clearly, Condition (5.22) implies Assumption 5.1. For both cases, Assumption 5.3 is verified using theorem 4.1 of Dai (1995). To verify that this assumption is satisfied for each of these two cases, we need to prove the stability of the corresponding fluid model. The stability of the LBFS reentrant fluid model is proved in theorem 4.4 of Dai and Weiss (1996). Under Condition (2.11), the fluid model stability of the two-station, five-class reentrant line is proved in theorem 8.25 of Dai and Harrison (2020). In both cases, state space collapse Condition

(5.14) is satisfied by Cao et al. (2022), which in turn implies Assumption 5.4. In both cases, Assumption 5.5 is verified in Dai et al. (2024). □

For the two-station, five-class reentrant line, Assumptions 5.4 (SSC) and 5.5 (matrix R) are equivalent to Condition (2.11). However, Assumption 5.3 (stability) is weaker than (2.11). Understanding the relationship among these three assumptions in a general network is a future research direction.

In Section 7.1, we give an outline of the proof of Theorem 5.1. The main tool is the basic adjoint relationship (BAR) that characterizes the stationary distribution $\pi^{(r)}$. For this BAR, we extend the approach that was developed in Braverman et al. (2017) for single-class networks (generalized Jackson networks), which we call the BAR approach. For a special case, we have done this for the two-station, five-class reentrant line in Section 2, starting from the case that interarrival and service-time distributions are exponential.

Compared with Braverman et al. (2017), the novelty of the present BAR approach, to be developed in detail in Sections 6–8, is to explicitly define Palm distributions carefully and demonstrate the intricate interplay between the stationary measure and Palm measures in the setting of multiclass queueing networks. This interplay has recently been explored in Guang et al. (2024).

Readers who are not familiar with our setting may be puzzled by our reason for introducing a *sequence* of networks and imposing the heavy-traffic conditions (5.4)–(5.6). As a motivation, one can consider the following situation. In a production system, it is up to the manager to decide how quickly jobs are to be released into the system. In particular, one needs to decide how heavily the system should be loaded to effectively use its resources. Ideally, one would like to choose each ρ_j close to one. A sequence corresponding to such a network arises by varying the load condition imposed by the manager; one envisions the network as a member of the sequence, with r chosen small since ρ is close to e. The heavy-traffic limit corresponding to this sequence of networks should then provide insight on the behavior of the original network.

Finally, for later usage, we state the following lemma, whose proof is straightforward, where we recall that $\beta_k^{(r)}$ is the probability that class k can be served.

Lemma 5.1. *Equations (5.4) and (5.5) imply that as* $r \longrightarrow 0$,

$$\mu_k^{(r)} = \mu_k + r\mu_k^* + o(r), \qquad k \in \mathcal{K}, \tag{5.23}$$

$$\beta_\ell^{(r)} = 1 - \rho_{s(\ell)}^{(r)} = rb_\ell + o(r), \qquad \ell \in \mathcal{L}, \tag{5.24}$$

$$\beta_k^{(r)} = \beta_k + o(1) \text{ with } \beta_k > 0, \qquad k \in \mathcal{H}, \tag{5.25}$$

where $\mu_k^* = -m_k^*/(m_k^2)$, $\beta_k = 1 - \sum_{\ell \in H(k)} \gamma_k$, *and* $\gamma_k = \alpha_k m_k > 0$ *for* $k \in \mathcal{K}$. *Here and later, we adopt the convention that* $f(r) = o(g(r))$ *and* $f(r) = O(g(r))$ *mean, respectively,*

$$\lim_{r\downarrow 0} \frac{f(r)}{g(r)} = 0, \qquad \limsup_{r\downarrow 0} \left|\frac{f(r)}{g(r)}\right| < \infty.$$

6. BAR Approach for Queueing Networks

Our final goal is to prove Theorem 5.1, in which the interarrival and service times are generally distributed. Under this distributional assumption, $Z(\cdot) \equiv \{Z(t); t \geq 0\}$ is not a Markov process. Because of this, we first consider continuous-time Markov process $X(\cdot) \equiv \{X(t); t \geq 0\}$, which was introduced in Section 3. A prominent feature of this Markov process is that it has finitely many jumps in each finite interval and partially differentiable deterministic sample paths between adjacent jump instants of them. This class of Markov processes is called a piecewise deterministic Markov process and was studied in Davis (1984).

Our starting point is the stationary distribution of this Markov process $X(\cdot)$, assuming its existence. To consider this distribution, we derive its basic adjoint relationship (BAR), also known in the literature as the stationary equation (Miyazawa 1994). In Davis (1984), such a stationary equation is derived, but it requires a boundary condition as an additional condition, which is hard to handle. Here, we do not use such an additional condition. Namely, we recapitulate the BAR approach that was first developed in Miyazawa (2017) and later expanded in Braverman et al. (2017) for generalized Jackson networks. Most of the foundational results are carefully developed here again for the purpose of completeness and easy reference.

BAR has been used to characterize the stationary distributions for various Markov processes. See, for example, Ethier and Kurtz (1986) for diffusion processes, Harrison and Williams (1987) for SRBMs, and Glynn and Zeevi

(2008) for Markov chains. The present BAR approach is in the same line, but a crucially different feature is included to handle well the discontinuous state changes of $X(\cdot)$, for which Palm distributions are used. We will fully detail these Palm distributions, which are slightly different from those in Miyazawa (2017) and not explicitly considered in Braverman et al. (2017).

6.1. Framework for Deriving BAR for $X(\cdot)$

In this section, we present a framework of our BAR approach for the piecewise deterministic Markov process $X(\cdot)$ introduced in Section 3. Recall that this Markov process has state space $S = \mathbb{Z}_+^K \times \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}_+^K$, and $X(t) = (Z(t), R_e(t), R_s(t))$ at time $t \geq 0$.

Define the set $\mathcal{D}$ as the set of functions $f : S \to \mathbb{R}$ satisfying the following conditions.

(6.a) $f(x)$ is bounded in $x \equiv (z, u, v) \in S$ and $f(z, u, v)$ is continuous in $(u, v) \in \mathbb{R}_+^{E+K}$ for each fixed $z \in \mathbb{Z}_+^K$.

(6.b) For each fixed $z \in \mathbb{Z}_+^K$, $f(z, u, v)$ has partial derivatives from the right in u_ℓ and v_k for each $\ell \in \mathcal{E}$ and $k \in \mathcal{K}$, and these partial derivatives are bounded, where $u \equiv (u_k)_{k\in\mathcal{E}}$ and $v \equiv (v_k)_{k\in\mathcal{K}}$.

For $f \in \mathcal{D}$, we introduce the following notations for describing the dynamics of $X(\cdot)$.

$$\mathcal{A}f(x) = -\sum_{k\in\mathcal{E}} \frac{\partial f}{\partial u_k}(x) - \sum_{k\in\mathcal{K}} \frac{\partial f}{\partial v_k}(x) 1(z_k > 0, z_{H_+(k)} = 0), \tag{6.1}$$

$$\Delta f(X)(t) = f(X(t)) - f(X(t-)), \tag{6.2}$$

where again $H_+(k)$ is defined in (3.5), and $X(t-) \equiv \lim_{\varepsilon\downarrow 0} X(t-\varepsilon)$ for $t > 0$. It follows from the fundamental theorem of calculus that for $t > 0$,

$$f(X(t)) - f(X(0)) = \int_0^t \mathcal{A}f(X(s))ds + \sum_{m=1}^{\infty} \Delta f(X)(\tau_m) 1(0 < \tau_m \leq t), \tag{6.3}$$

where τ_m is the mth jump instant of $X(\cdot)$ for $m \geq 1$. It is not hard to see that τ_m is finite for each $m \geq 1$ and $\tau_m \to \infty$ as $m \to \infty$ since $T_{e,k}$ and $T_{s,k}$ are finite with probability one.

To facilitate the introduction of our version of Palm probability measures, for each class $\ell \in \mathcal{E}$, we use $t_{e,\ell}^n$ to denote the arrival time of the nth class ℓ external arrival. Similarly, for each $k \in \mathcal{K}$, we use $t_{s,k}^n$ to denote the service completion time of the nth class k job. We call each entry in $\{t_{e,\ell}^n, n \geq 1\}$ for $\ell \in \mathcal{E}$ and $\{t_{s,k}^n, n \geq 1\}$ for $k \in \mathcal{K}$ an *event time* of the queueing network. It is possible that multiple event times are identical, corresponding to a single jump instant of $X(\cdot)$. In the following, we first assume that

(6.c) multiple events *cannot* occur simultaneously.

At the end of this subsection, we will argue that all results in this section continue to hold when this assumption is removed.

For each $\ell \in \mathcal{E}$ and $k \in \mathcal{K}$, let $N_{e,\ell}(\cdot) = \{N_{e,\ell}(t), t \geq 0\}$ and $N_{s,k}(\cdot) = \{N_{s,k}(t), t \geq 0\}$ be the counting processes associated with $\{t_{e,\ell}^n, n \geq 1\}$ and $\{t_{s,k}^n, n \geq 1\}$, respectively. In general, $N(\cdot) = \{N(t); t \geq 0\}$ is called a counting process if it is a nonnegative integer-valued process on $[0, \infty)$ that is nondecreasing and right-continuous and that has limits from the left. Note that $N(\cdot)$ must have finitely many jump instants in each finite interval. Define the integration of a function $g : \mathbb{R}_+ \to \mathbb{R}$ by counting process $N(\cdot)$ by

$$\int_{(0,t]} g(s) dN(0,s] = \sum_{m=1}^{\infty} 1(s_m \leq t) g(s_m) \Delta N(s_m)$$

where $0 < s_1 < s_2 < \cdots < s_m < \cdots$ are the jump times of $N(\cdot)$ and $\Delta N(s_m)$ is the jump size at time s_m. It follows from (6.3) that

$$\begin{aligned} f(X(t)) - f(X(0)) &= \int_0^t \mathcal{A}f(X(s))ds \\ &\quad + \sum_{\ell\in\mathcal{E}} \int_{(0,t]} \Delta f(X(s)) dN_{e,\ell}(s) + \sum_{k\in\mathcal{K}} \int_{(0,t]} \Delta f(X(s)) dN_{s,k}(s). \end{aligned} \tag{6.4}$$

Note that (6.4) directly follows from (6.3) when $N_{e,\ell}(t)$ and $N_k(t)$ for $\ell \in \mathcal{E}$ and $k \in \mathcal{K}$ do not have a common jump at any time $t \geq 0$.

Under Assumption 5.3, $X(\cdot) = X^{(r)}(\cdot)$ has the stationary distribution. Taking it as the initial distribution at time 0, then $X(\cdot)$ is a stationary process. In what follows, we always assume that $X(\cdot)$ is a stationary Markov process and denote an S-valued random vector subject to the stationary distribution $\pi = \pi^{(r)}$ of $X(\cdot)$ by $X \equiv (Z, R_e, R_s)$.

For $f \in \mathcal{D}$, because both f and $\mathcal{A}f$ are bounded, taking expectation for both sides of (6.4) with $t=1$ yields

$$\mathbb{E}(\mathcal{A}f(X)) + \mathbb{E}\left(\sum_{\ell\in\mathcal{E}}\int_{(0,1]}\Delta f(X(s))dN_{e,\ell}(s) + \sum_{k\in\mathcal{K}}\int_{(0,1]}\Delta f(X(s))dN_{s,k}(s)\right) = 0. \tag{6.5}$$

This equation exactly corresponds to (2.65) of the two-station, five-class reentrant network in Section 2.5. We there have introduced Palm distributions to handle well the second expectation in (2.65), which corresponds to the last two terms in (6.5). We will take the same approach here. We first introduce a state space for the network states just before its jump instants. For $\ell \in \mathcal{E}$ and $k \in \mathcal{K}$, define

$$\Gamma_{e,\ell} = \{x = (z,u,v) \in S : u_\ell = 0\}, \quad \Gamma_{s,k} = \{x = (z,u,v) \in S : v_k = 0, z_k > 0, z_{H_+(k)} = 0\},$$
$$\Gamma = (\cup_{k\in\mathcal{E}}\Gamma_{e,k}) \cup (\cup_{k\in\mathcal{K}}\Gamma_{s,k}).$$

We call Γ the boundary of state space S. By our convention, the state process $X(\cdot)$ is right continuous. As a consequence, one can verify that $X(t) \in S \setminus \Gamma$ for $t \ge 0$, and $X(t_m-) \in \Gamma$ for each $m \ge 1$.

As we have experienced in the definitions (2.66) and (2.67), it is important to evaluate $\mathbb{E}[N_{e,k}(1)]$ and $\mathbb{E}[N_{s,k}(1)]$ for defining Palm distributions. For this, recall that $\{\alpha_k; k \in \mathcal{K}\}$ is the unique solution of the traffic Equation (3.10). We also note that $\mathbb{E}[N_{e,k}(1)]$ and $\mathbb{E}[N_{s,k}(1)]$ must be finite by the law of large numbers because $T_{e,k}$ and $T_{s,k}$ have finite and positive expectations.

Lemma 6.1. *Assume Assumption 5.3. For each $r \in (0,1]$,*

$$\mathbb{E}[N_{e,k}(1)] = \lambda_k, \quad k \in \mathcal{E}, \tag{6.6}$$

$$\mathbb{E}[N_{s,k}(1)] = \alpha_k, \quad k \in \mathcal{K}. \tag{6.7}$$

Proof. We first prove Equation (6.6). The proof is different from the proof for (A.12) in Braverman et al. (2017). Fix a $k \in \mathcal{E}$. For constant $\kappa > 0$, take $f(x) = u_k \wedge \kappa$ for (6.4). Because$\{R_{e,k}(t); t \ge 0\}$ is a stationary process and

$$\mathcal{A}f(x) = -1(u_k \le \kappa), \qquad \Delta f(X(t^n_{e,k})) = a_k T_{e,k}(n) \wedge \kappa,$$

where $t^n_{e,k}$ is the nth increasing instant of $N_{e,k}(\cdot)$, taking the expectation of (6.4) for $t=1$, we have

$$-\mathbb{P}[R_{e,k} \le \kappa] + \mathbb{E}\left[\sum_{n=1}^{\infty}(a_k T_{e,k}(n) \wedge \kappa)1(n \le N_{e,k}(1))\right] = 0.$$

Because $T_{e,k}(n)$ and $\{n \le N_{e,k}(1)\} = \Omega \setminus \{N_{e,k}(1) < n\}$ are independent for each $n \ge 1$, this yields that

$$\mathbb{P}[R_{e,k} \le \kappa] = \mathbb{E}[a_k T_{e,k}(1) \wedge \kappa]\mathbb{E}[N_{e,k}(1)].$$

Letting $\kappa \to \infty$ in this formula, we have (6.6) because $\mathbb{E}[a_k T_{e,k}(1) \wedge \kappa]$ converges to $a_k = 1/\lambda_k$. (6.7) is similarly proved. In this case, for $k \in \mathcal{K}$, we take $f(x) = z_k$ for (6.4). Then,

$$\Delta f(X(s)) = 1(R_{e,k}(s-) = 0) - 1(\Phi^{(k)} \ne e^{(k)})1(R_{s,k}(s-) = 0)$$
$$+ \sum_{\ell\in\mathcal{K}\setminus\{k\}} 1(\Phi^{(\ell)} = e^{(k)})1(R_{s,\ell}(s-) = 0),$$

and $\mathcal{A}f(x) = 0$. Hence, taking the expectation of (6.4) yields

$$\mathbb{E}[N_{e,k}(1)] - (1 - P_{k,k})\mathbb{E}[N_{s,k}(1)] + \sum_{\ell\in\mathcal{K}\setminus\{k\}} \mathbb{E}[N_{s,\ell}(1)]P_{\ell,k} = 0.$$

Because $\mathbb{E}[N_{e,k}(1)] = \lambda_k$ by (6.6), this equation implies that $\{\mathbb{E}[N_{s,k}(1)]; k \in \mathcal{K}\}$ is the solution of the traffic Equation (3.10). Hence, we have (6.7) because $\{\alpha_k; k \in \mathcal{K}\}$ is the unique solution of (3.10). □

Similar to (2.66) and (2.67), we define probability distributions $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$ on $(S^2, \mathcal{B}(S^2))$ as

$$\mathbb{P}_{e,k}[B] = \frac{1}{\lambda_k}\mathbb{E}\left[\int_{(0,1]} 1((X(t-), X(t)) \in B)dN_{e,k}(t)\right], \qquad B \in \mathcal{B}(S^2), k \in \mathcal{E}, \tag{6.8}$$

$$\mathbb{P}_{s,k}[B] = \frac{1}{\alpha_k}\mathbb{E}\left[\int_{(0,1]} 1((X(t-), X(t)) \in B)dN_{s,k}(t)\right], \qquad B \in \mathcal{B}(S^2), k \in \mathcal{K}, \tag{6.9}$$

where $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$ are indeed probability distributions because Lemma 6.1 implies $\mathbb{P}_{e,k}[S^2] = \mathbb{P}_{s,k}[S^2] = 1$. We call these distributions Palm distributions concerning $N_{e,k}$ and $N_{s,k}$, respectively. Similar to the arguments in Section 2.5, let (X_-, X_+) be a pair of canonical random variables taking values in S^2 on the measurable space $(S^2, \mathcal{B}(S^2))$. Namely,

$$(X_-, X_+)(x_1, x_2) = (x_1, x_2) \quad \text{for each pair} \quad (x_1, x_2) \in S^2.$$

It follows from definitions, on each one of the probability spaces $(S^2, \mathcal{B}(S^2), \mathbb{P}_{e,k})$ and $(S^2, \mathcal{B}(S^2), \mathbb{P}_{s,k})$, with probability one,

$$X_+ \equiv (Z_+, R_{+,e}, R_{+,s}) \in S \setminus \Gamma, \quad X_- \equiv (Z_-, R_{-,e}, R_{-,s}) \in \Gamma. \tag{6.10}$$

Hence, X_- can be considered the prejump state for each jump type caused by either an exogenous arrival or a service completion, and X_+ is the postjump state under the Palm distributions. Denote the expectations under $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$ by $\mathbb{E}_{e,k}$ and $\mathbb{E}_{s,k}$, respectively. Then, for Borel measurable functions $f : \Gamma \to \mathbb{R}_+$ and $g : S \to \mathbb{R}_+$, we have

$$\mathbb{E}_{e,k}[f(X_-)g(X_+)] = \frac{1}{\lambda_k}\mathbb{E}\left[\int_{(0,1]} f(X(t-))g(X(t))dN_{e,k}(t)\right], \quad k \in \mathcal{E}, \tag{6.11}$$

$$\mathbb{E}_{s,k}[f(X_-)g(X_+)] = \frac{1}{\alpha_k}\mathbb{E}\left[\int_{(0,1]} f(X(t-))g(X(t))dN_{s,k}(t)\right], \quad k \in \mathcal{K}. \tag{6.12}$$

Probability distributions $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$ are closely related to Palm measures in the literature; see, for example, Baccelli and Brémaud (2003) and Miyazawa (1994). In this regard, we make the following two remarks. First, Palm measures in Miyazawa (1994) are defined on the same measurable space $(\Omega, \mathcal{F})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is the original probability space on which primitives such as $T_{e,k}(\cdot)$ and $T_{s,k}(\cdot)$ are defined; in our definitions, the measurable space is $(S^2, \mathcal{B}(S^2))$ on which (X_-, X_+) is defined. This paper does not require knowledge of the Palm measures beyond what is defined in (6.8) and (6.9) and thus is self-contained. However, one must be careful because we are dealing with multiple probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, $(S^2, \mathcal{B}(S^2), \mathbb{P}_{e,k})$ and $(S^2, \mathcal{B}(S^2), \mathbb{P}_{s,k})$ at once. Because this is different from the standard stochastic analysis using a single probability space, one may be puzzled at first. However, we will work only through expected values computed on those probability spaces, so there should be no confusion.

For each measurable function f from S to $\mathbb{R}$, let

$$\Delta f(X_+, X_-) = f(X_+) - f(X_-). \tag{6.13}$$

Thus, random variables X_-, X_+, and $\Delta f(X_+, X_-)$ are defined on the common measurable space $(S^2, \mathcal{B}(S^2))$. The notation $\Delta f(X_+, X_-)$ is inconsistent with $\Delta f(X(t))$, but they can be distinguished by their arguments. Immediately from (6.11) and (6.12), we have the following lemma.

Lemma 6.2. *For each bounded Borel measurable function $f : S \to \mathbb{R}$,*

$$\mathbb{E}\left[\int_{(0,1]} \Delta f(X(s))dN_{e,\ell}(s)\right] = \lambda_\ell \mathbb{E}_{e,\ell}[\Delta f(X_+, X_-)], \quad \ell \in \mathcal{E}, \tag{6.14}$$

$$\mathbb{E}\left[\int_{(0,1]} \Delta f(X(s))dN_{s,k}(s)\right] = \alpha_k \mathbb{E}_{s,k}[\Delta f(X_+, X_-)], \quad k \in \mathcal{K}. \tag{6.15}$$

With the new notational system, BAR (6.5) becomes

$$\mathbb{E}[\mathcal{A}f(X)] + \sum_{\ell\in\mathcal{E}} \lambda_\ell \mathbb{E}_{e,\ell}[\Delta f(X_+, X_-)] + \sum_{k\in\mathcal{K}} \alpha_k \mathbb{E}_{s,k}[\Delta f(X_+, X_-)] = 0, \quad f \in \mathcal{D}. \tag{6.16}$$

At this point, (6.5) and (6.16) differ only in symbols and not in mathematical substance. Our next lemma and its corollary allow a practical way to compute $\mathbb{E}_{e,\ell}[\Delta f(X_+, X_-)]$ and $\mathbb{E}_{s,k}[\Delta f(X_+, X_-)]$. They also show that $X_+ - X_-$ exactly corresponds to $X(t) - X(t-)$ at jump instants.

Lemma 6.3. *The prejump state X_- and the postjump state X_+ have the following representation,*

$$X_+ = X_- + (e^{(k)}, a_k T_{e,k} e^{(k)}, 0)), \qquad \text{under } \mathbb{P}_{e,k}, k \in \mathcal{E}, \tag{6.17}$$

$$X_+ = X_- + (-e^{(k)} + \Phi^{(k)}, 0, m_k T_{s,k} e^{(k)}), \qquad \text{under } \mathbb{P}_{s,k}, k \in \mathcal{K}, \tag{6.18}$$

where $T_{e,\ell}$ for $\ell \in \mathcal{E}$ and $T_{s,k}, \Phi^{(k)}$ for $k \in \mathcal{K}$ are random variables defined on the measurable space $(S^2, \mathcal{B}(S^2))$ such that, under Palm distribution $\mathbb{P}_{e,\ell}$, $T_{e,\ell}$ is independent of X_- and has the same distribution as that of $T_{e,\ell}(1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$, and, under Palm distribution $\mathbb{P}_{s,k}$, $(T_{s,k}, \Phi^{(k)})$ is independent of X_- and has the same distribution as that of $(T_{s,k}(1), \Phi^{(k)}(1))$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Let $t^m_{e,k}$ be the mth increasing instant of $N_{e,k}(\cdot)$. From (6.11), for bounded Borel measurable functions $f : \Gamma \to \mathbb{R}_+$ and $h : S \to \mathbb{R}_+$

$$\begin{aligned}
\mathbb{E}_{e,k}[f(X_-)h(X_+ - X_-)] &= \frac{1}{\lambda_k}\mathbb{E}\left[\int_{(0,1]} f(X(t-))h(X(t) - X(t-))dN_{e,k}(t)\right] \\
&= \frac{1}{\lambda_k}\mathbb{E}\left[\sum_{m=1}^{\infty} 1(t^m_{e,k} \le 1) f(X(t^m_{e,k}-))h(X(t^m_{e,k}) - X(t^m_{e,k}-))\right] \\
&= \frac{1}{\lambda_k}\sum_{m=1}^{\infty}\mathbb{E}[1(t^m_{e,k} \le 1) f(X(t^m_{e,k}-))]\mathbb{E}[h((e^k, a_k T_{e,k}(1), 0))] \\
&= \mathbb{E}_{e,k}[f(X_-)]\mathbb{E}[h((e^k, a_k T_{e,k}(1), 0))],
\end{aligned} \tag{6.19}$$

where in obtaining the third equality, we have used the following three facts on probability space $(\Omega, \mathcal{F}, \mathbb{P})$:
(a)

$$X(t^m_{e,k}) = X(t^m_{e,k}-) + (e^{(k)}, a_k T_{e,k}(m) e^{(k)}, 0)),$$

(b) $T_{e,k}(m)$ is independent of $X(t^m_{e,k}-)$ and $t^m_{e,k}$, and (c) $\{T_{e,k}(m), m \ge 1\}$ is an *i.i.d.* sequence. Because (6.19) implies that $\mathbb{E}_{e,k}[h(X_+ - X_-)] = \mathbb{E}[h((e^k, a_k T_{e,k}(1), 0))]$ for $f(x) \equiv 1$, we have

$$\mathbb{E}_{e,k}[f(X_-)h(X_+ - X_-)] = \mathbb{E}_{e,k}[f(X_-)]\mathbb{E}_{e,k}[h(X_+ - X_-)].$$

That is, X_- and $X_+ - X_-$ are independent under $\mathbb{P}_{e,k}$, and $X_+ - X_-$ under $\mathbb{P}_{e,k}$ has the same distribution as $(e^k, a_k T_{e,k}(1), 0)$ under $\mathbb{P}$. Thus, there exists a random variable $T_{e,k}$ on $(S^2, \mathcal{B}(S^2))$ that has the same distribution as $T_{e,k}(1)$ under that of $\mathbb{P}$ such that

$$X_+ = X_- + (e^{(k)}, a_k T_{e,k} e^{(k)}, 0), \quad \text{under} \quad \mathbb{P}_{e,k},$$

where $T_{e,k}$ is independent of X_-. This proves all the claims on $X_+ - X_-$. Similar results are also obtained for $\mathbb{P}_{s,k}$. Thus, the lemma is proved. □

The following lemma is immediate from this lemma.

Corollary 6.1. *For each $f \in \mathcal{D}$, define $\overline{f}_{e,k}, \overline{f}_{s,k}$ and $\overline{f}$ as*

$$\begin{aligned}
\overline{f}_{e,k}(x) &= \mathbb{E}_{e,k}[f(z + e^{(k)}, u + a_k T_{e,k}(1) e^{(k)}, v)] \quad x \in \Gamma_{e,k}, \quad k \in \mathcal{E}, \\
\overline{f}_{s,k}(x) &= \mathbb{E}_{s,k}[f(z - e^{(k)} + \Phi^{(k)}(1), u, v + m_k T_{s,k} e^{(k)})] \quad x \in \Gamma_{s,k}. \quad k \in \mathcal{K}, \\
\overline{f}(x) &= \sum_{k \in \mathcal{E}} \overline{f}_{e,k}(x) 1(u_k = 0) + \sum_{k \in \mathcal{K}} \overline{f}_{s,k}(x) 1(v_k = 0), \quad x = (z, u, v) \in S,
\end{aligned}$$

then

$$\mathbb{E}_{e,k}[\Delta f(X_+, X_-)] = \mathbb{E}_{e,k}[\Delta \overline{f}(X_-)], \quad k \in \mathcal{E}, \tag{6.20}$$

$$\mathbb{E}_{s,k}[\Delta f(X_+, X_-)] = \mathbb{E}_{s,k}[\Delta \overline{f}(X_-)], \quad k \in \mathcal{K}, \tag{6.21}$$

where $\Delta \overline{f}(x) = \overline{f}(x) - f(x), \quad x \in S$.

It can be proved that (6.16) fully characterizes the stationary distribution of $X(\cdot)$ in the following sense (e.g., see Miyazawa (1991) for its proof). The stationary distribution exists if and only if there are distributions ν on S and $\nu_{e,k}, \nu_{s,k}$ on Γ such that

$$\int_S \mathcal{A}f(x)\nu(dx) + \sum_{k\in\mathcal{E}} \lambda_k \int_{\Gamma_{e,k}} \Delta\bar{f}(dx)\nu_{e,k}(dx) + \sum_{k\in\mathcal{K}} \alpha_k \int_{\Gamma_{s,k}} \Delta\bar{f}(dx)\nu_{s,k}(dx) = 0, \quad f \in \mathcal{D}.$$

We now use (6.16) to prove the following lemmas. For each of our proofs, we construct a particular test function $f \in \mathcal{D}$ to be used in BAR (6.16). For $f \in \mathcal{D}$, both f and $\mathcal{A}f$ need to be bounded. In the following, our f's are not always bounded. To overcome this difficulty, we apply (6.16) to test function $f \wedge \kappa$ for each fixed $\kappa > 0$. Then, we take the limit in each of the terms in (6.16) as $\kappa \to \infty$. Since this limit procedure is standard (see, for example, the proof of (A.12) in Braverman et al. (2017)), we omit it in our proofs.

We next state and prove a lemma that evaluates the tail of expectations.

Lemma 6.4. *For $n \ge 0$,*

$$\mathbb{E}[R_{e,k}^n 1(R_{e,k} \ge c)] = \frac{1}{n+1} a_k^n \mathbb{E}[(T_{e,k}^{n+1} - c^{n+1}/a_k^{n+1}) 1(a_k T_{e,k} \ge c)], \qquad k \in \mathcal{E},\ c \in \mathbb{R}_+ \tag{6.22}$$

$$\begin{aligned}&\mathbb{E}[R_{s,k}^n 1(R_{s,k} \ge c, Z_k > 0, Z_{H_+(k)} = 0)]\\ &\quad = \frac{1}{n+1}\gamma_k m_k^n \mathbb{E}[(T_{s,k}^{n+1} - c^{n+1}/m_k^{n+1}) 1(m_k T_{s,k} \ge c)], \qquad k \in \mathcal{K},\ c \in \mathbb{R}_+,\end{aligned} \tag{6.23}$$

where we recall $\gamma_k = \alpha_k m_k$.

Proof. Fix $n \ge 0$, $k \in \mathcal{E}$, and $c \in \mathbb{R}_+$. We first prove (6.22). Let $f(x) = [\max(u_k, c)]^{n+1}$ for $x = (z, u, v) \in S$. It follows that

$$\begin{aligned}\mathcal{A}f(X) &= -(n+1)R_{e,k}^n 1(R_{e,k} \ge c),\\ \Delta f(X_+, X_-) &= a_k^{n+1}(T_{e,k}^{n+1} - c^{n+1}/a_k^{n+1}) 1(a_k T_{e,k} \ge c) 1(R_{-,e,k} = 0).\end{aligned}$$

By (6.16),

$$\begin{aligned}(n+1)\mathbb{E}[R_{e,k}^n 1(R_{e,k} \ge c)] &= \lambda_k a_k^{n+1}(\mathbb{E}[T_{e,k}^{n+1} - c^{n+1}/a^{n+1}) 1(a_k T_{e,k} \ge c)]\mathbb{E}_{e,k}[1]\\ &= a_k^n \mathbb{E}[(T_{e,k}^{n+1} - c^{n+1}/a^{n+1}) 1(a_k T_{e,k} \ge c)],\end{aligned}$$

proving (6.22). Next we prove (6.23). Let $f(x) = [\max(v_k, c)]^{n+1}$, then

$$\begin{aligned}\mathcal{A}f(X) &= -(n+1)R_{s,k}^n (R_{s,k} \ge c, Z_k > 0, Z_{H_+(k)} = 0),\\ \Delta f(X_+, X_-) &= m_k^{n+1}(T_{s,k}^{n+1} - c^{n+1}/m_k^{n+1}) 1(m_k T_{s,k} \ge c) 1(R_{-,s,k} = 0).\end{aligned}$$

Hence, similarly to (6.22), BAR (6.16) implies (6.23). □

This lemma will be used in the proofs of Lemmas 6.6 and 8.2 and Appendix B. We now make a connection with Braverman et al. (2017). The following lemma appeared in Braverman et al. (2017). Its proof follows from (6.16) immediately.

Lemma 6.5. *Assume $f \in \mathcal{D}$ satisfies*

$$\bar{f}_{e,k}(x) = f(x), \qquad k \in \mathcal{E}, \quad x \in \Gamma_{e,k}, \tag{6.24}$$

$$\bar{f}_{s,k}(x) = f(x), \qquad k \in \mathcal{K}, \quad x \in \Gamma_{s,k}. \tag{6.25}$$

Then

$$\mathbb{E}(\mathcal{A}f(X)) = 0. \tag{6.26}$$

BAR (6.26) is the main tool used in Braverman et al. (2017). We can still rely on (6.26) to prove some cases of Theorem 5.1 in this paper. For example, assume that $T_{e,k}$ for $k \in \mathcal{E}$ and $T_{s,k}$ for $k \in \mathcal{K}$ have general distributions but have bounded supports. Then, similar to (2.61), redefine f_θ as

$$f_\theta(x) = g_\theta(z)\exp(-\langle\eta(\theta), \lambda u\rangle - \langle\xi(\theta), \mu v\rangle), \quad x = (z, u, v) \in S, \quad \theta \in \mathbb{R}^K, \tag{6.27}$$

where $\lambda u = (\lambda_k u_k, k \in \mathcal{E})$, $\mu v = (\mu_k v_k, k \in \mathcal{K})$, and

$$g_\theta(z) = \exp(\langle\theta, z\rangle), \qquad z \in \mathbb{Z}_+^K, \tag{6.28}$$

and, similar to (2.58)–(2.60), functions $\eta_k(\theta_k)$ and $\xi_k(\theta)$ are defined through the following equations:

$$e^{\theta_k}\mathbb{E}(e^{-\eta_k(\theta_k)T_{e,k}}) = 1, \qquad k \in \mathcal{E}, \tag{6.29}$$

$$\sum_{\ell \in \overline{\mathcal{K}}} P_{k,\ell} e^{-\theta_k + \theta_\ell}\mathbb{E}(e^{-\xi_k(\theta)T_{s,k}}) = 1, \qquad k \in \mathcal{K}. \tag{6.30}$$

Then we can derive a BAR of $X^{(r)}$ because Conditions (6.24) and (6.25) are satisfied, respectively. Indeed, it follows from (6.28) and (6.1) that

$$\mathcal{A}f_\theta(X) = \sum_{k \in \mathcal{E}} \lambda_k \eta_k(\theta_k) f_\theta(X) + \sum_{k \in \mathcal{K}} \mu_k \xi_k(\theta) f_\theta(X) 1(Z_k > 0, Z_{H_+(k)} = 0).$$

Therefore, (6.26) implies that for each $\theta \in \mathbb{R}_-^K$

$$\sum_{k \in \mathcal{E}} \lambda_k \eta_k(\theta_k)\mathbb{E}[f_\theta(X)] + \sum_{k \in \mathcal{K}} \mu_k \xi_k(\theta)\mathbb{E}[f_\theta(X) 1(Z_k > 0, Z_{H_+(k)} = 0)] = 0. \tag{6.31}$$

Furthermore, if a set $\Theta_L \subset \mathbb{R}_-^{\mathcal{L}}$ similar to the one defined in (2.44) is nonempty, all the arguments in the case of the two-station, five-class network in Section 2 similarly work for the present network, and Theorem 5.1 can be proved.

However, it is too strong to assume that $T_{e,k}$ for $k \in \mathcal{E}$ and $T_{s,k}$ for $k \in \mathcal{K}$ have bounded supports, and we are not able to prove Θ_L nonempty in general. To prevent these extra assumptions, we truncate u, v, and z_H in the test function $f_\theta(z, u, v)$ in (6.27), where $z = (z_L, z_H)$. This kind of truncation was done for u, v in Braverman et al. (2017). However, the truncation of z_H causes a serious problem in deriving a BAR because (6.24) and (6.25) are no longer satisfied after the truncation. This is a challenge that did not arise in Braverman et al. (2017). We will attack this problem using a so-called asymptotic BAR in the next section.

We end this section by outlining an approach to remove the no-simultaneous-events assumption (6.c). First we sort all event times $t_{e,\ell}^n$ and $t_{e,\ell}^n$, $n \geq 1$, $\ell \in \mathcal{E}$, and $k \in \mathcal{K}$. The sorted sequence in nondecreasing order is denoted by $\{\tau_m, m \geq 1\}$. We assume there is a rule to break ties when multiple event times are equal. For example, one may adopt a rule that arrival events precede service completion events, and low-class events precede high-class events. The event sequence $\{\tau_m, m \geq 1\}$ here is different from the jump instant sequence in (6.3). Here, when $\tau_m = \tau_{m+1}$ by definition

$$X(\tau_m) = X(\tau_{m+1}) \quad \text{and} \quad X(\tau_m-) = X(\tau_{m+1}-). \tag{6.32}$$

We now define what we call intermediate states Y_m and Y_{m-} for $m \geq 1$. In general, $Y_m \neq Y_{m+1}$ in contrast to $X(\tau_m) = X(\tau_{m+1})$ in (6.32). To define intermediate states, we call

$$\tau_{n-1} < \tau_n = \cdots = \tau_{n-1+\delta} < \tau_{n+\delta} \tag{6.33}$$

an event block of size δ starting from n. We now define Y_m and Y_{m-} for $m = n, \ldots, n-1+\delta$ as follows:

$$Y_{n-} = X(\tau_n-),$$

for $m = n, \ldots, n-1+\delta$,

$$Y_m = Y_{m-} + \begin{cases} (e^{(\ell)}, a_\ell T_{e,\ell}(i) e^{(\ell)}, 0)) & \text{if } \tau_m = t_{e,\ell}^i \\ (-e^{(k)} + \Phi^{(k)}(j), 0, m_k T_{s,k}(j) e^{(k)})) & \text{if } \tau_m = t_{s,k}^j \end{cases},$$

$$Y_{(m+1)-} = Y_m.$$

One can verify that all the exposition and proofs in this section continue to be valid as long as for each $m \geq 1$, $X(\tau_m)$ and $X(\tau_m-)$ are replaced by Y_m and Y_{m-}, respectively. The paragraph below (3.13) of Braverman et al. (2017)

provides a similar approach to dealing with simultaneous events in generalized Jackson networks. See also (M4) of Miyazawa (2024) for a similar treatment.

6.2. Test Functions for BAR of $X^{(r)}$

Recall that $X^{(r)} = (Z^{(r)}, R_e^{(r)}, R_s^{(r)})$ is subject to the stationary distribution of Markov process $X^{(r)}(\cdot)$. To prove Theorem 5.1, we need to find an equation to characterize the limit of the distributions $Z^{(r)}$ as $r \downarrow 0$. To this end, we first derive a BAR for $X^{(r)}$, then derive a BAR for $Z^{(r)}$. For the BAR of $X^{(r)}$, we take a test function from the state space $S = \mathbb{Z}_+^K \times \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}_+^K$ to $\mathbb{R}$. The choice of this test function is crucial to our approach.

As discussed at the end of Section 6.1, we truncate z_H, u, v in the test function $f_\theta^{(r)}(x)$. This is done in the following way. We first truncate the queue length vector, and define test function $g_{\theta,s}$ for $z \in \mathbb{Z}_+^K$ as

$$g_{\theta,s}(z) = \exp(\langle \theta_L, z_L \rangle + \langle \theta_H, z_H \wedge 1/s \rangle), \qquad z \in \mathbb{Z}_+^K, \tag{6.34}$$

where $z_H \wedge 1/s$ is the H-dimensional vector whose kth entry is $\min(z_k, 1/s)$ for $k \in \mathcal{H}$. Then, incorporating those two truncations on u, v, we define the test function $f_{\theta,s,t}^{(r)}$ for $r, s, t \in (0,1]$ and $\theta \in \Theta$ for $X^{(r)}$ as

$$f_{\theta,s,t}^{(r)}(x) = g_{\theta,s}(z)\exp(-\langle \eta(\theta,t), \lambda^{(r)} u \wedge t^{-1} \rangle - \langle \xi(\theta,t), \mu^{(r)} v \wedge t^{-1} \rangle). \tag{6.35}$$

For this test function, we have to change (6.29) and (6.30) to

$$e^{\theta_k}\mathbb{E}(e^{-\eta_k(\theta_k,t)(T_{e,k} \wedge t^{-1})}) = 1, \qquad k \in \mathcal{E}, \tag{6.36}$$

$$\sum_{\ell \in \overline{\mathcal{K}}} P_{k,\ell} e^{-\theta_k + \theta_\ell}\mathbb{E}(e^{-\xi_k(\theta,t)(T_{s,k} \wedge t^{-1})}) = 1, \qquad k \in \mathcal{K}. \tag{6.37}$$

These $\eta_k(\theta_k, t)$ and $\xi_k(\theta, t)$ are uniquely determined by (6.36) and (6.37) as shown in Braverman et al. (2017), but the proof there is a bit complicated. Therefore, we will verify these facts in a simpler way by Lemma C.1 in Appendix C.

Define

$$\Theta = \mathbb{R}_-^{\mathcal{L}} \times \mathbb{R}^{\mathcal{H}}. \tag{6.38}$$

We are now ready to define two sets of moment generating functions (MGFs) for $Z^{(r)}$ and $X^{(r)}$. Recall that $H(k) \subset \mathcal{K}$ is the set of classes at station $s(k)$ with priority at least as high as k (see (3.4)). For $Z^{(r)}$, define, for each $\theta \in \Theta$ and $r \in (0,1]$,

$$\phi^{(r)}(\theta) = \mathbb{E}[g_{\theta,r}(Z^{(r)})], \quad \phi_k^{(r)}(\theta) = \mathbb{E}[g_{\theta,r}(Z^{(r)}) | Z_{H(k)}^{(r)} = 0], \quad k \in \mathcal{K}, \quad \theta \in \mathbb{R}_-^K, \tag{6.39}$$

where $z \in \mathbb{R}^K$, z_A is defined to be $(z_k; k \in A)$. Note that $\phi^{(r)}(r\theta)$ is the Laplace transform of $rZ^{(r)}$ for $\theta \in \mathbb{R}_-^{\mathcal{K}} \subset \Theta$. Because $\mathbb{P}\{Z_{H(k)}^{(r)} = 0\}$ is the probability that there is no customer whose priority is at least k at station $s(k)$, the following lemma is intuitively clear.

Lemma 6.6. *Under Assumption 5.3,*

$$\mathbb{P}\{Z_{H(k)}^{(r)} = 0\} = \beta_k^{(r)}, \qquad k \in \mathcal{K}, \quad r \in (0,1], \tag{6.40}$$

where $\beta_k^{(r)}$ is defined in (5.2).

Proof. Note that (6.40) is a special case of (6.23) (by setting $n = 0$ and $c = 0$ there) in Lemma 6.4 of Section 6.1 because $\mathbb{P}(Z_k^{(r)} > 0, Z_{H_+(k)}^{(r)} = 0) = \mathbb{P}(Z_{H_+(k)}^{(r)} = 0) - \mathbb{P}(Z_{H(k)}^{(r)} = 0)$. □

For $X^{(r)}$, we let $s = r$ and $t = r^{1-\varepsilon_0}$, where $\varepsilon_0 \in (0, \delta_0/(1+\delta_0)]$, and define truncated MGFs $\psi^{(r)}(\theta)$ and $\psi_k^{(r)}(\theta)$ for $k \in \mathcal{K}$ as

$$\psi^{(r)}(\theta) = \mathbb{E}[f_{\theta,r,r^{1-\varepsilon_0}}^{(r)}(X^{(r)})], \qquad \psi_k^{(r)}(\theta) = \mathbb{E}[f_{\theta,r,r^{1-\varepsilon_0}}^{(r)}(X^{(r)}) | Z_{H(k)}^{(r)} = 0], \tag{6.41}$$

which are well defined for each $\theta \in \Theta$ because $f_{\theta,s,t}^{(r)}(x)$ is bounded in x for each $\theta \in \Theta$, $r, s, t \in (0,1]$. These MGFs cannot be called Laplace transforms because their domains are not limited to $\mathbb{R}_-^K$.

7. Proof of Theorem 5.1

The aim of this section is to prove Theorem 5.1. Throughout this section, we assume Assumptions 5.1–5.5 and use $X^{(r)} \equiv (Z^{(r)}, R_e^{(r)}, R_s^{(r)})$ to denote an S-valued random variable subject to the stationary distribution of the Markov process $X^{(r)}(\cdot)$. This proof of Theorem 5.1 requires several steps. We first outline the proof in six steps in Section 7.1. All steps are fully detailed in the subsequent sections.

7.1. Outline of the Proof

Recall that, once Lemma 4.1 is proved, the proof of Theorem 5.1 is completed with help of Lemma 4.2. Thus, we aim to prove the BAR (4.2) in Lemma 4.1. This BAR will be obtained from prelimit BARs, which takes six steps.

(Step 1) Using the MGFs $\psi^{(r)}$ and $\psi_k^{(r)}$ of (6.41), we derive an asymptotic BAR for $X^{(r)}$ in the following proposition.

Proposition 7.1. *Assume the assumptions in Theorem 5.1. Then, for each fixed $\theta \in \Theta$,*

$$q^{(r)}(r\theta, r^{1-\epsilon_0})\psi^{(r)}(r\theta) - \sum_{\ell\in\mathcal{L}} \beta_\ell^{(r)}\mu_\ell^{(r)}\xi_\ell(r\theta, r^{1-\epsilon_0})(\psi_\ell^{(r)}(r\theta) - \psi^{(r)}(r\theta))$$
$$+ \sum_{\ell\in\mathcal{H}} \beta_\ell^{(r)}(\mu_{\ell-}^{(r)}\xi_{\ell-}(r\theta, r^{1-\epsilon_0}) - \mu_\ell^{(r)}\xi_\ell(r\theta, r^{1-\epsilon_0}))(\psi_\ell^{(r)}(r\theta) - \psi^{(r)}(r\theta)) = o(r^2) \tag{7.1}$$

where, for $s \in (0,1)$, $q^{(r)}(\theta, s)$ is defined by

$$q^{(r)}(\theta, s) = \sum_{k\in\mathcal{E}} \lambda_k^{(r)}\eta_k(\theta_k, s) + \sum_{k\in\mathcal{K}} \alpha_k^{(r)}\xi_k(\theta, s), \qquad \theta \in \mathbb{R}^K. \tag{7.2}$$

We call (7.1) an asymptotic BAR of $X^{(r)}$. The proof of this proposition is lengthy and complicated because it requires SSC under the Palm distributions. We defer it to Section 8.

(Step 2) To rewrite (7.1) in more tractable form, we prepare asymptotic expansions of η_k and ξ_k. Namely, uniformly bound $\eta_k(r\theta_k, r^{1-\epsilon_0})$ for $k \in \mathcal{E}$ and $\xi_k(r\theta, r^{1-\epsilon_0})$ for $k \in \mathcal{K}$ by linear functions of θ_k and θ, respectively, and expand them as quadratic functions $\eta_k^*(r\theta_k)$ of θ_k and $\xi_k^*(r\theta)$ of θ, respectively, plus $o(r^2)$, where they are defined as

$$\eta_k^*(\theta_k) = \overline{\eta}_k(\theta_k) + \tilde{\eta}_k(\theta_k), \quad k \in \mathcal{E}, \qquad \xi_k^*(\theta) = \overline{\xi}_k(\theta_k) + \tilde{\xi}_k(\theta_k), \quad k \in \mathcal{K}, \tag{7.3}$$

where

$$\overline{\eta}_k(\theta_k) = \theta_k, \qquad \tilde{\eta}_k(\theta_k) = \frac{1}{2}c_{e,k}^2\theta_k^2, \qquad k \in \mathcal{E}, \tag{7.4}$$

$$\overline{\xi}_k(\theta) = -\theta_k + \sum_{\ell\in\mathcal{K}} P_{k,\ell}\theta_\ell, \qquad k \in \mathcal{K}, \tag{7.5}$$

$$\tilde{\xi}_k(\theta) = \frac{1}{2}\left(\sum_{\ell\in\mathcal{K}} P_{k,\ell}\theta_\ell^2 - \left(\sum_{\ell\in\mathcal{K}} P_{k,\ell}\theta_\ell\right)^2 + c_{s,k}^2\left(-\theta_k + \sum_{\ell\in\mathcal{K}} P_{k,\ell}\theta_\ell\right)^2\right), \qquad k \in \mathcal{K}. \tag{7.6}$$

(Step 3) Using the results obtained in Step 2, we replace $\eta_k(r\theta_k, r^{1-\epsilon_0})$ and $\xi_k(r\theta, r^{1-\epsilon_0})$ in (7.1) by $\eta_k^*(\theta_k)$ and $\xi_k^*(\theta)$ of θ. This yields, for $\theta \in \Theta \equiv \mathbb{R}_-^{\mathcal{L}} \times \mathbb{R}^{\mathcal{H}}$ (see (6.38)),

$$q^*(r\theta, r)\psi^{(r)}(r\theta) - \sum_{\ell\in\mathcal{L}} \mu_\ell^{(r)}\xi_\ell^*(r\theta)\beta_\ell^{(r)}(\psi_\ell^{(r)}(r\theta) - \psi^{(r)}(r\theta))$$
$$+ \sum_{\ell\in\mathcal{H}} (\mu_{\ell-}^{(r)}\xi_{\ell-}^*(r\theta) - \mu_\ell^{(r)}\xi_\ell^*(r\theta))\beta_\ell^{(r)}(\psi_\ell^{(r)}(r\theta) - \psi^{(r)}(r\theta)) = o(r^2), \tag{7.7}$$

where

$$q^*(\theta, r) = \sum_{k\in\mathcal{E}} \lambda_k^{(r)}\eta_k^*(\theta_k) + \sum_{k\in\mathcal{K}} \alpha_k^{(r)}\xi_k^*(\theta), \quad \theta \in \mathbb{R}^K. \tag{7.8}$$

(Step 4) Note that $\phi^{(r)}(r\theta)$ is a Laplace transform for $\theta \in \mathbb{R}_-^K$ and is bounded by one. Hence, by a standard "diagonal argument," we have the following.

Lemma 7.1 (Theorem 5.19 in Kallenberg 2001). *For any sequence in $(0,1]$ that goes to zero, there exists a subsequence $r_n \to 0$, and Laplace transforms $\phi(\theta)$ and $\phi_k(\theta)$, $k \in \mathcal{K}$, of finite measures such that*

$$\lim_{n\to\infty}(\phi^{(r_n)}(r_n\theta), \phi_k^{(r_n)}(r_n\theta), k \in \mathcal{K}) = (\phi(\theta), \phi_k(\theta), k \in \mathcal{K}), \qquad \theta \in \mathbb{R}_-^K. \tag{7.9}$$

We call $(\phi, \phi_k, k \in \mathcal{K})$ in (7.9) a *limit point*. The limit point may depend on the original sequence in $(0,1]$. Thus, there could be multiple limit points. Following the general theory, each component of a limit point $(\phi(\theta), \phi_k(\theta), k \in \mathcal{K})$ for $\theta \in \mathbb{R}_-^K$ is not necessarily the Laplace transform of a probability measure.

(Step 5) Using the moment-SSC Assumption 5.4 and the expansions in Step 2, we prove the following.

Lemma 7.2. *Under heavy-traffic Assumption 5.2 and moment-SSC Assumption 5.4, the following transform SSCs hold. For each $\theta \in \Theta$,*

$$\lim_{r\downarrow 0}(\psi^{(r)}(r\theta) - \phi^{(r)}(\theta_L, 0)) = 0, \quad \lim_{r\downarrow 0}(\psi_k^{(r)}(r\theta) - \phi_k^{(r)}(\theta_L, 0)) = 0, \quad k \in \mathcal{K}. \tag{7.10}$$

In the remainder of this step and Step 6, we use $(\phi, \phi_k, k \in \mathcal{K})$ to denote a fixed limit point with corresponding subsequence $\{r_n\}$. For notational simplicity, we omit index n and simply write r_n as r in both $r \downarrow 0$ and $o(r^2)$. This does not cause any problems because the subsequence $\{r_n; n \geq 1\}$ is fixed once it is chosen.

By Lemmas 7.1 and 7.2, we have

$$\lim_{r\downarrow 0}\psi^{(r)}(r\theta) = \phi(\theta_L, 0), \qquad \lim_{r\downarrow 0}\psi_k^{(r)}(r\theta) = \phi_k(\theta_L, 0). \tag{7.11}$$

Hence, we can replace $\psi^{(r)}(r\theta)$ and $\psi_\ell^{(r)}(r\theta)$ in (7.7) by $\phi(\theta_L, 0)$ and $\phi_k(\theta_L, 0)$, which yields

$$\begin{aligned} &q^*(r\theta, r)\phi(\theta) - \sum_{\ell\in\mathcal{L}} \mu_\ell^{(r)} \xi_\ell^*(r\theta)\beta_\ell^{(r)}(\phi_\ell(\theta_L, 0_H) - \phi(\theta_L, 0_H)) \\ &+ \sum_{\ell\in\mathcal{H}}(\mu_{\ell-}^{(r)}\xi_{\ell-}^*(r\theta) - \mu_\ell^{(r)}\xi_\ell^*(r\theta))\beta_\ell^{(r)}(\phi_\ell(\theta_L, 0_H) - \phi(\theta_L, 0_H)) = o(r^2), \ \theta \in \Theta. \end{aligned} \tag{7.12}$$

(Step 6) Suitably choosing θ_H, we can remove the second summation in (7.12) as $r \downarrow 0$. In this way, we prove the next lemma.

Lemma 7.3. *Assume Assumptions 5.1–5.5. Then each limit point $(\phi, \phi_k, k \in \mathcal{K})$ in (7.9) satisfies*

$$q(\theta_L, \theta_H)\phi(\theta_L, 0) + \sum_{\ell\in\mathcal{L}} b_\ell\langle\theta_L, R^{(\ell)}\rangle(\phi_\ell(\theta_L, 0) - \phi(\theta_L, 0)) = 0, \quad \theta_L \in \mathbb{R}_-^{\mathcal{L}}, \tag{7.13}$$

where q is defined by (5.19) and $\theta_H = -(A_H^{-1})'(A_{LH})'\theta_L$.

Obviously, this lemma yields the BAR (4.2) in Lemma 4.1, where Σ is determined through (5.21). Furthermore, $\phi(\theta_L, 0_H)$ and $\phi_k(\theta_L.0_H)$ are the Laplace transforms of unique probability measures by Lemma 4.2. Denote those probability measures by ν and ν_ℓ, $\ell \in \mathcal{L}$, respectively. Then, they are uniquely stationary distributions of SRBM with (R, Σ, b) because R is assumed to be completely $\mathcal{S}$ and Σ is assumed to be nondegenerate. This completes the proof of Theorem 5.1.

In what follows, we detail Steps 2–6, including the proofs of Lemmas 7.2 and 7.3, whereas Step 1 is proved in Section 8.

7.2. Expansions of η_k, ξ_k and Bounds for MGFs (Step 2)

In moving from MGFs $\psi^{(r)}(r\theta), \psi_k^{(r)}(r\theta)$ to MGFs $\phi^{(r)}(r\theta), \phi_k^{(r)}(r\theta)$, we need to well control the extra terms involving $R_e^{(r)}, R_s^{(r)}$ in $\psi^{(r)}(r\theta), \psi_k^{(r)}(r\theta)$ so that they are ignorable as $r \downarrow 0$. For this, we bound and expand $\eta_k(\theta_k, r^{1-\varepsilon_0})$ and $\xi_k(\theta, r^{1-\varepsilon_0})$.

Lemma 7.4. *For each fixed $a > 0$, there are positive constants $d_{e,a}$ and $d_{s,a}$ such that for any $r \in (0,1]$ and any $\theta \in \mathbb{R}^K$ with $|\theta| < a$*

$$|\eta_k(\theta_k, r^{1-\varepsilon_0})| \leq d_{e,a}|\theta_k|, \quad k \in \mathcal{E}, |\xi_k(\theta_k, r^{1-\varepsilon_0})| \leq d_{s,a}|\theta|, \quad k \in \mathcal{K}, \tag{7.14}$$

where $|\theta| = \sum_{k\in\mathcal{K}}|\theta_k|$.

For the expansions, recall the definitions (7.3) in Step 2. Then, Taylor expansions for $\eta(r\theta, r^{1-\varepsilon_0})$ and $\xi(r\theta, r^{1-\varepsilon_0})$ as $r \longrightarrow 0$ are obtained as follows.

Lemma 7.5. *For each fixed $\theta \in \mathbb{R}^K$, as $r \downarrow 0$,*

$$\eta_k(r\theta_k, r^{1-\varepsilon_0}) = \eta_k^*(r\theta_k) + o(r^2), \qquad k \in \mathcal{E}, \tag{7.15}$$

$$\xi_k(r\theta, r^{1-\varepsilon_0}) = \xi_k^*(r\theta) + o(r^2), \qquad k \in \mathcal{K}. \tag{7.16}$$

These two lemmas are essentially the same as lemmas 4.2 and 4.3 in Braverman et al. (2017), but the results are notationally much simplified. For completeness and easy reference, these two lemmas will be proved in Appendix C. We observe that when all distributions are exponential, Taylor expansions for $\eta_k(r\theta)$ and $\xi_k(r\theta)$ (2.35)–(2.36) are identical to the ones given in (7.15) and (7.16), respectively.

To bound $f^{(r)}_{\theta,r,r^{1-\varepsilon_0}}$ of (6.35), we rewrite it as

$$f^{(r)}_{\theta,r,r^{1-\varepsilon_0}}(x) = g_{\theta,r}(z)e^{-\Lambda^{(r)}_{\theta,r^{1-\varepsilon_0}}(u,v)}, \quad \text{for } x=(z,u,v) \text{ and } r\in(0,1], \tag{7.17}$$

where

$$\Lambda^{(r)}_{\theta,r^{1-\varepsilon_0}}(u,v) = \langle \eta(\theta, r^{1-\varepsilon_0}), \lambda^{(r)}u \wedge r^{\varepsilon_0-1}\rangle + \langle \xi(\theta, r^{1-\varepsilon_0}), \mu^{(r)}v \wedge r^{\varepsilon_0-1}\rangle. \tag{7.18}$$

Then, we have the following facts.

Lemma 7.6. *For each $\theta\in\Theta$, we have*

$$g_{r\theta,r}(z) \le e^{|\theta_H|}, \qquad \forall z\in\mathbb{Z}^K_+ \text{ and } \forall r\in(0,1], \tag{7.19}$$

and for each $a>0$,

$$|\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)| \le |\theta|\left(d_{e,a}\sum_{\ell\in\mathcal{E}}(r\lambda^{(r)}_\ell u_\ell \wedge r^{\varepsilon_0}) + d_{s,a}\sum_{\ell\in\mathcal{K}}(r\mu^{(r)}_\ell v_\ell\wedge r^{\varepsilon_0})\right) \tag{7.20}$$

$$\le r^{\varepsilon_0}|\theta|(d_{e,a}E + d_{s,a}K), \quad \forall x\in S,\ \forall r\in(0,1] \text{ with } r|\theta| < a, \tag{7.21}$$

where we recall that $d_{e,a}$ and $d_{s,a}$ are constants in Lemma 7.4, $E=|\mathcal{E}|$, and $K=|\mathcal{K}|$. Hence, for each fixed $\theta\in\Theta$ and $a>0$,

$$f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(x) \le e^{|\theta_H|+|\theta|(d_{e,a}E+d_{s,a}K)}$$
$$\text{for } x=(z,u,v)\in S \text{ and } r\in(0,1] \text{ with } |\theta|\le a. \tag{7.22}$$

Proof. Note that (7.19) is immediate from the definition (6.34) of $g_{\theta,s}(x)$ for $s=r$. To prove (7.20), we apply Lemma 7.4 to the definition (7.18), and then for any $r\in(0,1]$,

$$|\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)| \le r|\theta|\left(d_{e,a}\sum_{\ell\in\mathcal{E}}(\lambda^{(r)}_\ell u_\ell\wedge r^{\varepsilon_0-1}) + d_{s,a}\sum_{\ell\in\mathcal{K}}(\mu^{(r)}_\ell v_\ell \wedge r^{\varepsilon_0-1})\right),$$

which proves (7.20) and (7.21). Finally, the bound (7.22) on $f^{(r)}_{r\theta,r,r}$ immediately follows from (7.19) and (7.21). □

To prove Proposition 7.1, we also use the following lemma, which is a direct consequence of (6.22) and (6.23) in Lemma 6.4 of Section 6.1.

Lemma 7.7. *Assume Assumption 5.1 and Assumption 5.3. For each $\ell\in\mathcal{E}$ and $k\in\mathcal{K}$, $\{R^{(r)}_{e,\ell}, r\in(0,1]\}$ and $\{R^{(r)}_{s,k}1(Z^{(r)}_k>0, Z^{(r)}_{H_+(k)}=0), r\in(0,1]\}$ are uniformly integrable.*

7.3. Tractable BAR (Step 3) and Limit Points (Step 4)

For Step 3, we first bound $\psi^{(r)}(r\theta)$ and $\psi^{(r)}_\ell(r\theta)$ for each $\theta\in\Theta$ by the deterministic bound in (7.22). Namely,

$$\max(\psi^{(r)}(r\theta), \psi^{(r)}_k(r\theta), k\in\mathcal{K}) \le e^{|\theta_H|+|\theta|(d_{e,a}E+d_{s,a}K)},\ r\in(0,1], \theta\in\{\zeta\in\Theta; |\zeta|\le a\}. \tag{7.23}$$

Then, (7.7) is obtained from the asymptotic BAR (7.1) by applying Lemmas 7.5 and 7.6. In Step 4, Lemma 7.1 shows the existence of the limit points $\phi(\theta)$ and $\phi_k(\theta)$ for $\phi^{(r)}(r\theta)$ and $\phi^{(r)}_k(r\theta)$ for $\theta\in\mathbb{R}^K_-$. Note that $\phi(\theta)$ and $\phi_k(\theta)$ are the Laplace transforms of finite measures but may not be the Laplace transform of probability measures.

7.4. SSC (Step 5)

Using the moment SSC Assumption 5.4, we prove a version of transform SSC.

Lemma 7.8. *Under Assumption 5.4, for each $\theta=(\theta_L,\theta_H)\in\mathbb{R}^K_-$,*

$$\lim_{r\downarrow 0}(\phi^{(r)}(r\theta) - \phi^{(r)}(r\theta_L,0)) = 0, \qquad \lim_{r\downarrow 0}(\phi^{(r)}_k(r\theta)-\phi^{(r)}_k(r\theta_L,0)) = 0, \quad k\in\mathcal{K}. \tag{7.24}$$

Furthermore, each limit point $(\phi, \phi_k, k \in \mathcal{K})$ *satisfies*

$$\phi(\theta) = \phi(\theta_L, 0_H), \qquad \phi_k(\theta) = \phi_k(\theta_L, 0_H), \quad k \in \mathcal{K}, \qquad \theta \in \mathbb{R}^K_-. \tag{7.25}$$

Proof. We prove (7.24). Fix a $\theta = (\theta_L, \theta_H) \in \Theta$. One can verify that

$$\begin{aligned}|g_{r(\theta_L,0),r}(z) - g_{r\theta,r}(z)| &= \exp(\langle \theta_L, rz_L\rangle)|\exp(\langle \theta_H, rz_H \wedge 1\rangle) - 1|\\ &\le e^{|\langle \theta_H, rz_H \wedge 1\rangle|}|\langle \theta_H, rz_H \wedge 1\rangle| \le e^{|\theta_H|}|\theta_H||rz_H \wedge 1|\\ &\le re^{|\theta_H|}|\theta_H||z_H|,\end{aligned}$$

where the first equation follows from $\theta_L \le 0$ and

$$|e^x - 1| \le e^{|x|}|x| \text{for } x \in \mathbb{R}. \tag{7.26}$$

Hence, it follows by Assumption 5.4 that

$$|\phi^{(r)}(r\theta_L, 0) - \phi^{(r)}(r\theta))| \le re^{|\theta_H|}|\theta_H|\sum_{\ell \in \mathcal{H}} \mathbb{E}[Z^{(r)}_\ell] \to 0, \qquad r \downarrow 0.$$

Thus, the first equation of (7.24) is obtained. For the second equation, we first consider it for $k \in \mathcal{H}$. Similarly to the case of the first equation, we have

$$\begin{aligned}|\phi^{(r)}_k(r\theta_L, 0) - \phi^{(r)}_k(r\theta))| &\le re^{|\theta_H|}|\theta_H|\sum_{\ell \in \mathcal{H}} \mathbb{E}[Z^{(r)}_\ell | Z^{(r)}_{H(k)} = 0]\\ &\le \frac{re^{|\theta_H|}|\theta_H|}{\mathbb{P}(Z^{(r)}_{H(k)} = 0)}\sum_{\ell \in \mathcal{H}} \mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_{H(k)} = 0)] \to 0,\end{aligned}$$

because $\mathbb{E}[Z^{(r)}_\ell]$ is uniformly bounded for each $\ell \in \mathcal{H}$ and $\mathbb{P}(Z^{(r)}_{H(k)} = 0) = \beta^{(r)}_k \to \beta_k > 0$ as $r \downarrow 0$ for $k \in \mathcal{H}$ by Lemma 6.6. We next consider the case for $k \in \mathcal{L}$. Because $\beta^{(r)}_k = rb_k + o(r)$ by Lemma 5.1, we need to show that

$$\lim_{r \downarrow 0} \mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_{H(k)} = 0)] = 0, \qquad \ell \in \mathcal{H}, k \in \mathcal{L}. \tag{7.27}$$

For this, we use the following inequality:

$$\begin{aligned}\mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_{H(k)} = 0)] &= \mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_\ell > r^{-1/2} Z^{(r)}_{H(k)} = 0)] + \mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_\ell \le r^{-1/2} Z^{(r)}_{H(k)} = 0)]\\ &\le \mathbb{E}[Z^{(r)}_\ell 1(Z^{(r)}_\ell > r^{-1/2})] + r^{-1/2}\mathbb{P}[Z^{(r)}_{H(k)} = 0].\end{aligned}$$

This proves (7.27) because of Assumption 5.4. Thus, the second equation in (7.24) is obtained for all $k \in \mathcal{K}$. Finally, (7.25) is immediate from (7.24). □

Lemma 7.8 is the first version of transform-SSC. We extend it to the MGF $\psi^{(r)}(r\theta) \equiv \mathbb{E}[f^{(r)}_\theta(X^{(r)})]$, which is Lemma 7.2. Recall that $\Theta \equiv \mathbb{R}^L_- \times \mathbb{R}^H$ and we use $\theta = (\theta_L, \theta_H)$ to denote an element in $\mathbb{R}^K = \mathbb{R}^L \times \mathbb{R}^H$ following convention (5.11).

Proof of Lemma 7.2. We first prove that

$$\lim_{r \downarrow 0}(\psi^{(r)}(r\theta) - \phi^{(r)}(r\theta)) = 0, \qquad \lim_{r \downarrow 0}(\psi^{(r)}_k(r\theta) - \phi^{(r)}_k(r\theta)) = 0, \qquad \theta \in \Theta. \tag{7.28}$$

Fix a $\theta = (\theta_L, \theta_H) \in \Theta$. Applying Lemma 7.6 to Expression (7.17) of $f_{r\theta,r,r^{1-\varepsilon_0}}$, we have

$$\begin{aligned}|f_{r\theta,r,r^{1-\varepsilon_0}}(x) - g_{r\theta,r}(z)| &= g_{r\theta,r}(z)|e^{-\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)} - 1| \le e^{|\theta_H|}|\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)|e^{|\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)|}\\ &\le r^{\varepsilon_0}|\theta|(d_{e,a}E + d_{s,a}K)e^{|\theta_H| + (d_{e,a}E + d_{s,a}K)r^{\varepsilon_0}|\theta|}\end{aligned}$$

for $r \in (0, 1]$ satisfying $r|\theta| \le a$, where the first inequality follows from (7.19) and (7.26) and the second from (7.21). Because $r|\theta| \le a$ is satisfied for any $\theta \in \Theta$ and any $a > 0$ for sufficiently small $r > 0$, the first equation of (7.28) is immediate from the previous inequality. Similarly, the second equation of (7.28) is obtained from

$$
\begin{aligned}
|\psi_k^{(r)}(r\theta) - \phi_k^{(r)}(r\theta)| &\le \mathbb{E}[|f_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)}) - g_{r\theta,r}(Z^{(r)})|1(Z_{H(k)}^{(r)} = 0)]/\mathbb{P}[Z_{H(k)}^{(r)} = 0] \\
&\le r^{\varepsilon_0}|\theta|(d_{e,a}E + d_{s,a}K)e^{|\theta_H| + (d_{e,a}E + d_{s,a}K)r^{\varepsilon_0}|\theta|}.
\end{aligned}
$$

Combining (7.28) with (7.24) of Lemma 7.8 proves (7.10) of Lemma 7.2. □

7.5. BAR for Class $\mathcal{L}$ (Step 6)

In this section, we prove Lemma 7.3, which is composed of two parts. We first derive a limit BAR for classes in $\mathcal{H}$, which is Lemma 7.9. We then complete the proof of Lemma 7.3.

Lemma 7.9. *The limit point $(\phi, \phi_k, k \in \mathcal{K})$ in (7.9) satisfies the following equations:*

$$
\sum_{\ell \in \mathcal{H}} (\mu_{\ell-}\overline{\xi}_{\ell-}(\theta) - \mu_\ell \overline{\xi}_\ell(\theta))\beta_\ell(\phi_\ell(\theta_L, 0_H) - \phi(\theta_L, 0_H)) = 0, \qquad \theta \in \Theta, \tag{7.29}
$$

$$
\phi_k(\theta_L, 0_H) = \phi(\theta_L, 0_H), \qquad k \in \mathcal{H}, \quad \theta_L \in \mathbb{R}_-^{\mathcal{L}}. \tag{7.30}
$$

Proof. From Definition (7.8) of $q^*(\theta, r)$ and Lemma 7.5, for each fixed $\theta \in \mathbb{R}^K$,

$$
\begin{aligned}
q^*(r\theta, r) &= \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \overline{\eta}_k(r\theta_k) + \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \overline{\xi}_k(r\theta) + \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \tilde{\eta}_k(r\theta_k) + \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \tilde{\xi}_k(r\theta) \\
&= r^2 \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \tilde{\eta}_k(\theta_k) + r^2 \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \tilde{\xi}_k(\theta) = o(r),
\end{aligned} \tag{7.31}
$$

because $\tilde{\eta}_k(\theta_k)$ and $\tilde{\xi}_k(\theta)$ are quadratic in θ and

$$
\begin{aligned}
&\sum_{k \in \mathcal{E}} \lambda_k^{(r)} \overline{\eta}_k(\theta_k) + \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \overline{\xi}_k(\theta) = \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \theta_k - \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \left(\theta_k - \sum_{\ell \in \mathcal{K}} P_{k,\ell} \theta_\ell \right) \\
&= \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \theta_k - \sum_{\ell \in \mathcal{K}} \alpha_k^{(r)} \theta_k + \sum_{\ell \in \mathcal{K}} \sum_{k \in \mathcal{K}} \alpha_k^{(r)} P_{k,\ell} \theta_\ell \\
&= \sum_{k \in \mathcal{E}} \lambda_k^{(r)} \theta_k - \sum_{k \in \mathcal{K}} \alpha_k^{(r)} \theta_k + \sum_{\ell \in \mathcal{K}} (\alpha_\ell^{(r)} - \lambda_\ell^{(r)}) \theta_\ell = 0
\end{aligned}
$$

by the definition of $\alpha^{(r)}$ in (5.1). Because $\mu_\ell^{(r)} \xi_\ell^*(r\theta) \beta_\ell^{(r)} = o(r)$ for $\ell \in \mathcal{L}$ by (5.24) and (7.16), we also have

$$
\sum_{\ell \in \mathcal{L}} \mu_\ell^{(r)} \xi_\ell^*(r\theta) \beta_\ell^{(r)} (\psi_\ell^{(r)}(r\theta) - \psi^{(r)}(r\theta)) = o(r).
$$

Hence, dividing (7.7) by r and letting $r \downarrow 0$, Lemma 7.2 yields (7.29) because $\mu_\ell^{(r)} \to \mu_\ell$ and $\beta_\ell^{(r)} \to \beta_\ell$ as $r \downarrow 0$ by (5.25).

We next use (7.29) to prove (7.30). For the proof, we fix a $\theta_L \in \mathbb{R}_-^{\mathcal{L}}$ and a class $k \in \mathcal{H}$. In the proof, we will use θ_L to construct a special $\theta = (\theta_L, \theta_H) \in \Theta$ that can be plugged into (7.29). Recall that θ_L, θ_H, and θ are all envisioned as column vectors, even though we have adopted Convention (5.11) in writing $\theta = (\theta_L, \theta_H)$. For fixed $k \in \mathcal{H}$, define

$$
\theta_H = -(A_H^{-1})'(A_{LH})'\theta_L + (A_H^{-1})' e_H^{(k)}, \tag{7.32}
$$

where $e_H^{(k)}$ is the H-vector with component k being one and all other components zero. Clearly $\theta = (\theta_L, \theta_H) \in \Theta$. We claim that

$$
\mu_{\ell-}\overline{\xi}_{\ell-}(\theta) - \mu_\ell \overline{\xi}_\ell(\theta) = 0, \quad \ell \in \mathcal{H} \setminus \{k\}, \tag{7.33}
$$

$$
\mu_{k-}\overline{\xi}_{k-}(\theta) - \mu_k \overline{\xi}_k(\theta) = 1, \tag{7.34}
$$

where recall that $k-$ is the highest class in $\{\ell \in \mathcal{K}; s(\ell) = s(k)\} \setminus H(k)$. Once this claim is verified, (7.30) is immediate from (7.29).

To verify (7.33) and (7.34), recall $\overline{\xi}_k(y)$ defined in (7.5) for $k \in \mathcal{K}$ and $y \in \mathbb{R}^K$. In vector form,

$$\overline{\xi}(y) = (P - I)y \quad \text{and} \quad (\mu_1\overline{\xi}_1(y), \dots, \mu_K\overline{\xi}_K(y))' = \text{diag}(\mu)(P - I)y. \tag{7.35}$$

Following from the definition of B in (5.16),

$$(\mu_{1-}\overline{\xi}_{1-}(y), \dots, \mu_{K-}\overline{\xi}_{K-}(y))' = B'(\mu_1\overline{\xi}_1(y), \dots, \mu_K\overline{\xi}_K(y))' = B' \ \text{diag}(\mu)(P - I)y.$$

Therefore,

$$(\mu_{1-}\overline{\xi}_{1-}(y), \dots, \mu_{K-}\overline{\xi}_{K-}(y))' - (\mu_1\overline{\xi}_1(y), \dots, \mu_K\overline{\xi}_K(y))' = (B' - I)\text{diag}(\mu)(P - I)y$$
$$= A'y = \begin{pmatrix} (A_L)'y_L + (A_{HL})'y_H \\ (A_{LH})'y_L + (A_H)'y_H \end{pmatrix}, \tag{7.36}$$

where A is defined in (5.15) with its block decomposition defined in (5.17). Fix $y_L = \theta_L$ and solve

$$(A_{LH})'y_L + (A_H)'y_H = e_H^{(k)} \tag{7.37}$$

yields unique solution $y_H = \theta_H$ in (7.32). One can verify that with $y_L = \theta_L$, (7.37) is equivalent to (7.33) and (7.34). In solving (7.37), we assume $(A_H)'$ is invertible, which is assumed in Assumption 5.5. □

We are now in the final step to prove Lemma 7.3.

Proof of Lemma 7.3. Fix a limit point $(\phi, \phi_k, k \in \mathcal{K})$ in (7.9) with the corresponding subsequence $r_n \downarrow 0$ and a point $\theta_L \in \mathbb{R}_-^{\mathcal{L}}$. We would like to prove that (7.13) holds for this θ_L. Recall the definition of θ_H in (5.20). Set $\theta = (\theta_L, \theta_H)$. Clearly $\theta \in \Theta$ because Θ has no sign restriction in θ_H. We now prove that the left side of (7.7), divided by r_n^2, goes to the left side of (7.13), proving the lemma. The left side has three terms. We now study the limit for each term.

We start with the third term. Similar to the derivation of (7.37), one can check that

$$(A_{LH})'\theta_L + (A_H)'\theta_H = 0,$$

which is equivalent to

$$\mu_{k-}\overline{\xi}_{k-}(\theta) - \mu_k\overline{\xi}_k(\theta) = 0, \quad k \in \mathcal{H}. \tag{7.38}$$

Fix a $k \in \mathcal{H}$. Using Definition (7.3), for each $r \in (0, 1]$

$$\begin{aligned}
&\mu_{k-}^{(r)}\xi_{k-}^*(r\theta) - \mu_k^{(r)}\xi_k^*(r\theta) \\
&\quad = \mu_{k-}^{(r)}(r\overline{\xi}_{k-}(\theta) + r^2\tilde{\xi}_{k-}(\theta)) - \mu_k^{(r)}(r\overline{\xi}_k(\theta) + r^2\tilde{\xi}_k(\theta)) \\
&\quad = (\mu_{k-} + r\mu_{k-}^*)(r\overline{\xi}_{k-}(\theta) + r^2\tilde{\xi}_{k-}(\theta)) \\
&\qquad - (\mu_k + r\mu_k^*)(r\overline{\xi}_k(\theta) + r^2\tilde{\xi}_k(\theta)) + o(r^2) \\
&\quad = r^2\mu_{k-}^*\overline{\xi}_{k-}(\theta) + r^2\mu_{k-}\tilde{\xi}_{k-}(\theta) - r^2\mu_k^*\overline{\xi}_k(\theta) + r^2\mu_k\tilde{\xi}_k(\theta) + o(r^2),
\end{aligned}$$

where the second equality follows from (5.23) and the third equality follows from (7.38). It follows from Lemma 7.2 and (7.30) that the third term in the left side of (7.7), divided by r^2, goes to zero as $r \downarrow 0$.

Next we study the first term. From (7.31) and (5.19),

$$\lim_{r\downarrow 0} \ r^{-2}q^{(r)}(r\theta, r) = \sum_{k\in\mathcal{E}} \lambda_k\tilde{\eta}_k(\theta) + \sum_{k\in\mathcal{K}} \alpha_k\tilde{\xi}_k(\theta) = q(\theta_L, \ -(A_H^{-1})'(A_{LH})'\theta_L).$$

This, together with Lemma 7.2, proves that the first term in the left side of (7.7), divided by r^2, goes to

$$q(\theta_L, \theta_H)\phi(\theta_L, 0)$$

as $r \downarrow 0$, recalling that $\theta_H = -(A_H^{-1})'A_{LH}'\theta_L$ in (5.20).

Finally, we study the second term. For $\ell \in \mathcal{L}$, we have

$$-\lim_{r\downarrow 0} \ r^{-2}\mu_\ell^{(r)}\xi_\ell^*(r\theta)\beta_\ell^{(r)} = -\lim_{r\downarrow 0} \ r^{-2}\mu_\ell^{(r)}\overline{\xi}_\ell(r\theta)\beta_\ell^{(r)} = \mu_\ell\overline{\xi}_\ell(\theta)b_\ell. \tag{7.39}$$

Conversely, substituting $y = (\theta_L, \theta_H)$ into (7.36) and choosing the entry $\ell \in \mathcal{L}$, we have

$$\mu_\ell \overline{\xi}_\ell(\theta) = -\langle \theta_L, R^{(\ell)} \rangle,$$

because $\mu_{\ell-} \overline{\xi}_{\ell-}(\theta) = 0$ for $\ell \in \mathcal{L}$ by our convention and $R = A_L - A_{LH} A_H^{-1} A_{HL}$. Clearly, this and (7.39), together with Lemma 7.2, proves that the second term in the left side of (7.7), divided by r^2, goes to

$$\sum_{\ell \in \mathcal{L}} b_\ell \langle \theta_L, R^{(\ell)} \rangle (\phi_\ell(\theta_L, 0) - \phi(\theta_L, 0))$$

as $r \downarrow 0$. The study of these three terms leads to the conclusion that the left side of the asymptotic BAR (7.7), divided by r^2, converges to the left side of SRBM BAR (7.13), which proves Lemma 7.3. □

8. Deriving BARs

As we said in Section 7.1, the proof of Theorem 5.1 is completed once the asymptotic BAR (7.1) in Proposition 7.1 is obtained. In this section, we prove Proposition 7.1. This will be done step by step. We first derive a BAR of $X^{(r)}$ using the test function $f^{(r)}_{\theta,s,t}$ in Section 8.1. From this BAR, we derive the asymptotic BAR (7.1), which proves Proposition 7.1. This will be done in Section 8.2, using three lemmas proved in Sections 8.3 and 8.4.

8.1. Primitive BAR of $X^{(r)}$

We derive a BAR of $X^{(r)}$ for the test function $f^{(r)}_{\theta,s,t}$ of (6.35). Setting $s = r$ and $t = r^{1-\varepsilon_0}$, we recall that this test function can be written as

$$f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(x) = g_{r\theta,r}(z)\exp(-\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v)), \quad x \in S, \tag{8.1}$$

where we recall from (6.34) and (7.18) that

$$g_{r\theta,r}(z) = \exp(\langle r\theta_L, z_L \rangle + \langle r\theta_H, z_H \wedge 1/r \rangle),$$
$$\Lambda^{(r)}_{r\theta,r^{1-\varepsilon_0}}(u,v) = \langle \eta(r\theta, r^{1-\epsilon_0}), \lambda^{(r)} u \wedge r^{\varepsilon_0 - 1} \rangle + \langle \xi(r\theta, r^{1-\epsilon_0}), \mu^{(r)} v \wedge r^{\varepsilon_0 - 1} \rangle.$$

For each $r \in (0,1]$ and each $\theta \in \Theta$, $f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}} \in \mathcal{D}$ by Lemma 7.6. This test function will be used for BAR (6.16). In what follows, $f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}$ is denoted by f for simplicity. To expand (6.16), we compute the following quantities:

$$\begin{aligned}
\frac{\partial f}{\partial u_k}(x) &= \lambda^{(r)}_k \eta_k(r\theta_k, r^{1-\varepsilon_0}) 1(\lambda^{(r)}_k u_k \le r^{\varepsilon_0 - 1}) f(x) \\
&= \lambda^{(r)}_k \eta_k(r\theta_k, r^{1-\varepsilon_0}) f(x) - \lambda^{(r)}_k \eta_k(r\theta_k, r^{1-\varepsilon_0}) 1(\lambda^{(r)}_k u_k > r^{\varepsilon_0 - 1}) f(x), \qquad k \in \mathcal{E}, \\
\frac{\partial f}{\partial v_k}(x) &= \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) 1(\mu^{(r)}_k v_k \le r^{\varepsilon_0 - 1}, z^{(r)}_k > 0, z^{(r)}_{H_+(k)} = 0) \\
&= \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) 1(z^{(r)}_k > 0, z^{(r)}_{H_+(k)} = 0) \\
&\quad - \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) 1(\mu^{(r)}_k v_k > r^{\varepsilon_0 - 1}, z^{(r)}_k > 0, z^{(r)}_{H_+(k)} = 0), \qquad k \in \mathcal{K}.
\end{aligned}$$

Then, it follows from (6.1) and (6.16) that

$$\begin{aligned}
&\sum_{k \in \mathcal{E}} \lambda^{(r)}_k \eta_k(r\theta_k, r^{1-\varepsilon_0}) \mathbb{E}[f(X^{(r)})] \\
&\quad + \sum_{k \in \mathcal{K}} \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) \mathbb{E}[1(Z^{(r)}_k > 0, Z^{(r)}_{H_+(k)} = 0) f(X^{(r)})] + E(r, \theta, r^{1-\epsilon_0}) = 0,
\end{aligned} \tag{8.2}$$

where

$$\begin{aligned}
E(r, \theta, r^{1-\epsilon_0}) = &-\sum_{k \in \mathcal{E}} \lambda^{(r)}_k \eta(r\theta_k, r^{1-\varepsilon_0}) \mathbb{E}[1(\lambda^{(r)}_k R^{(r)}_{e,k} > r^{\varepsilon_0 - 1}) f(X^{(r)})] \\
&- \sum_{k \in \mathcal{K}} \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) \mathbb{E}[1(\mu^{(r)}_k R^{(r)}_{s,k} > r^{\varepsilon_0 - 1}, Z^{(r)}_k > 0, Z^{(r)}_{H_+(k)} = 0) f(X^{(r)})] \\
&+ \sum_{k \in \mathcal{E}} \lambda^{(r)}_k \mathbb{E}_{e,k}[\Delta f(X^{(r)}_+, X^{(r)}_-)] + \sum_{k \in \mathcal{K}} \alpha^{(r)}_k \mathbb{E}_{s,k}[\Delta f(X^{(r)}_+, X^{(r)}_-)].
\end{aligned} \tag{8.3}$$

Recall that $f(x) = f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(x)$, and $\mathbb{E}_{e,k}$ and $\mathbb{E}_{s,k}$ stand for the expectations under the Palm distributions $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$, which, together with random variables $X^{(r)}_+$ and $X^{(r)}_-$, are defined by (6.11) and (6.12) through the counting processes $N^{(r)}_{e,k}(\cdot)$ and $N^{(r)}_{s,k}(\cdot)$, respectively. Equation (8.2) can be considered a BAR. We call it a primitive BAR. This BAR is the starting point of our analysis.

8.2. Proof of Proposition 7.1

We aim to derive (7.1) from (8.2). Recall that $\psi^{(r)}(r\theta) = \mathbb{E}[f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})]$ and $\psi^{(r)}_k(r\theta) = \mathbb{E}[f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)}) \mid Z^{(r)}_{H(k)} = 0]$ for $k \in \mathcal{K}$. If $E(r,\theta,r^{1-\epsilon_0})$ of (8.3) has order $o(r^2)$ as $r \downarrow 0$, then we have

$$\sum_{k\in\mathcal{E}} \lambda^{(r)}_k \eta_k(r\theta_k, r^{1-\varepsilon_0}) \mathbb{E}[f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] + \sum_{k\in\mathcal{K}} \mu^{(r)}_k \xi_k(r\theta, r^{1-\varepsilon_0}) \mathbb{E}[1(Z^{(r)}_k > 0, Z^{(r)}_{H_+(k)} = 0) f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] = o(r^2). \tag{8.4}$$

Then, if (8.4) holds, the proof of Proposition 7.1 is completed by the next lemma.

Lemma 8.1. (8.4) *is equivalent to* (7.1) *in Proposition* 7.1.

Proof. Because

$$\mathbb{E}[1(Z^{(r)}_k > 0, Z^{(r)}_{H_+(k)} = 0) f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] = \mathbb{E}[(1(Z^{(r)}_{H_+(k)} = 0) - 1(Z^{(r)}_{H(k)} = 0)) f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] = \beta^{(r)}_{k+}\psi^{(r)}_{k+}(\theta) - \beta^{(r)}_k \psi^{(r)}_k(\theta)$$

by Lemma 6.6, we can write (8.4) as

$$\sum_{k\in\mathcal{E}} \lambda^{(r)}_k \eta^*_k(r\theta_k)\psi^{(r)}(r\theta) + \sum_{k\in\mathcal{K}} \mu^{(r)}_k \xi^*_k(r\theta)(\beta^{(r)}_{k+}\psi^{(r)}_{k+}(r\theta) - \beta^{(r)}_k \psi^{(r)}_k(r\theta)) = o(r^2), \tag{8.5}$$

because $\eta_k(r\theta_k, r^{1-\varepsilon_0}) = \eta^*_k(r\theta_k) + o(r^2)$ and $\xi_k(r\theta_k, r^{1-\varepsilon_0}) = \xi^*_k(r\theta_k) + o(r^2)$ by Lemma 7.4 and the boundedness of $\psi^{(r)}(r\theta)$ and $\psi^{(r)}_k(r\theta)$.

Because $\alpha^{(r)}_k = \mu^{(r)}_k(\beta^{(r)}_{k+} - \beta^{(r)}_k)$ by (5.2),

$$\sum_{k\in\mathcal{K}} \alpha^{(r)}_k \xi^*_k(\theta)\psi^{(r)}(\theta) = \sum_{k\in\mathcal{K}} \mu^{(r)}_k \xi^*_k(\theta)(\beta^{(r)}_{k+} - \beta^{(r)}_k)\psi^{(r)}(\theta).$$

Hence, (8.5) can be written as

$$\sum_{k\in\mathcal{E}} \lambda^{(r)}_k \eta^*_k(\theta_k)\psi^{(r)}(\theta) + \sum_{k\in\mathcal{K}} \mu^{(r)}_k \xi^*_k(\theta)(\beta^{(r)}_{k+}\psi^{(r)}_{k+}(\theta) - \beta^{(r)}_k \psi^{(r)}_k(\theta)) + \sum_{k\in\mathcal{K}} \alpha^{(r)}_k \xi^*_k(\theta)\psi^{(r)}(\theta) - \sum_{k\in\mathcal{K}} \mu^{(r)}_k(\beta^{(r)}_{k+} - \beta^{(r)}_k)\xi^*_k(\theta)\psi^{(r)}(\theta) = o(r^2).$$

Hence,

$$\left(\sum_{k\in\mathcal{E}} \lambda^{(r)}_k \eta^*_k(\theta_k) + \sum_{k\in\mathcal{K}} \alpha^{(r)}_k \xi^*_k(\theta)\right)\psi^{(r)}(\theta) + \sum_{k\in\mathcal{K}} \mu^{(r)}_k \xi^*_k(\theta)(\beta^{(r)}_{k+}(\psi^{(r)}_{k+}(\theta) - \psi^{(r)}(\theta)) - \beta^{(r)}_k(\psi^{(r)}_k(\theta) - \psi^{(r)}(\theta)))$$
$$= q^{(r)}(\theta)\psi^{(r)}(\theta) + \sum_{k\in\mathcal{K}} (\mu^{(r)}_{k-}\xi^*_{k-}(\theta) - \mu^{(r)}_k \xi^*_k(\theta))\beta^{(r)}_k(\psi^{(r)}_k(\theta) - \psi^{(r)}(\theta)) = o(r^2).$$

This is equivalent to (7.1) because $\mu^{(r)}_{k-} = 0$ for $k \in \mathcal{L}$. □

It remains to prove (8.4) to complete the proof of Proposition 7.1. For this, it is sufficient to show that $E(r,\theta,r^{1-\epsilon_0}) = o(r^2)$, which is proved by the following two lemmas.

Lemma 8.2. *Under Assumptions 5.1 and 5.3, for each* $\theta \in \mathbb{R}^L_- \times \mathbb{R}^H$,

$$\lambda_k^{(r)}\eta_k(r\theta_k, r^{1-\varepsilon_0})\mathbb{E}[1(\lambda_k^{(r)}R_{e,k}^{(r)} > r^{\varepsilon_0-1})f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] = o(r^2), \qquad k \in \mathcal{E}, \tag{8.6}$$

$$\mu_k^{(r)}\xi(r\theta, r^{1-\varepsilon_0})\mathbb{E}[1(\mu_k^{(r)}R_{s,k}^{(r)} > r^{\varepsilon_0-1}, Z_k^{(r)} > 0, Z^{(r)}_{H_+(k)} = 0)f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}(X^{(r)})] = o(r^2), \quad k \in \mathcal{K}. \tag{8.7}$$

Lemma 8.3. *Fix* $\theta \in \Theta$ *and let* $f = f^{(r)}_{r\theta,r,r^{1-\varepsilon_0}}$:

$$\mathbb{E}_{e,k}[\Delta f(X_+^{(r)}, X_-^{(r)})] = o(r^2), \quad k \in \mathcal{E}, \tag{8.8}$$

$$\mathbb{E}_{s,k}[\Delta f(X_+^{(r)}, X_-^{(r)})] = o(r^2), \quad k \in \mathcal{K}. \tag{8.9}$$

Before proving these two lemmas, we prepare one lemma, the SSC of $Z^{(r)}_{-,H}$ under Palm distributions, which will be used to prove Lemma 8.3, where $Z^{(r)}_{-,H}$ is the H-dimensional random vector whose kth entry is $Z^{(r)}_{-,k}$ for $k \in \mathcal{H}$. This SSC is itself interesting because it is not immediate from the SSC of $Z^{(r)}$ under $\mathbb{P}$. Therefore, we verify it in Section 8.3, separate from the proofs of Lemmas 8.2 and 8.3. Finally, these two lemmas are proved in Section 8.4.

8.3. SSC Under Palm Distributions

We prove the SSC of $Z^{(r)}_{-,H}$ under Palm distributions $\mathbb{P}_{e,k}$ and $\mathbb{P}_{s,k}$.

Lemma 8.4. *Under Assumptions 5.1–5.4, for* $\ell \in \mathcal{H}$,

$$\mathbb{P}_{e,k}\{Z^{(r)}_{-,\ell} > r^{-1} - 1\} = o(r), \qquad k \in \mathcal{E}, \tag{8.10}$$

$$\mathbb{P}_{s,k}\{Z^{(r)}_{-,\ell} > r^{-1} - 1\} = o(r), \qquad k \in \mathcal{K}. \tag{8.11}$$

The proof of this lemma requires the next lemma, which relates the tail probabilities of $Z^{(r)}$ under the Palm distributions to those under $\mathbb{P}$.

Lemma 8.5. *For each integer* $n \ge 1$, *each* $r \in (0,]$, *each* $c \in \mathbb{R}_+$ *and* $k, \ell \in \mathcal{K}$,

$$\begin{aligned}\mathbb{P}(Z_k^{(r)} \ge n, Z^{(r)}_{H_+(k)} = 0) &= \gamma_k[(1-P_{kk})\mathbb{P}_{s,k}(Z^{(r)}_{-,k} \ge n+1) + P_{kk}\mathbb{P}_{s,k}(Z^{(r)}_{-,k} \ge n)] \\ &\quad + \lambda_k\mathbb{E}_{e,k}[R^{(r)}_{-,s,k}1(Z^{(r)}_{-,k} = n-1)] + \sum_{\ell \in \mathcal{K}\setminus\{k\}} \alpha_\ell P_{\ell k}\mathbb{E}_{s,\ell}[R^{(r)}_{-,s,k}1(Z^{(r)}_{-,k} = n-1)],\end{aligned} \tag{8.12}$$

$$\begin{aligned}&\mathbb{P}(R^{(r)}_{s,k} \le c, Z_\ell^{(r)} \ge n, Z_k^{(r)} > 0, Z^{(r)}_{H_+(k)} = 0) + \alpha_\ell(1-P_{\ell\ell})\mathbb{E}_{s,\ell}[(R^{(r)}_{s,k} \wedge c)1(Z^{(r)}_{-,\ell} = n)] \\ &= \gamma_k\mathbb{E}[T_{s,k} \wedge (c/m_k)][P_{k\ell}\mathbb{P}_{s,k}(Z^{(r)}_{-,\ell} \ge n-1) + (1-P_{k\ell})\mathbb{P}_{s,k}(Z^{(r)}_{-,\ell} \ge n)] \\ &\quad + \lambda_\ell\mathbb{E}_{e,\ell}[(R^{(r)}_{s,k} \wedge c)1(Z^{(r)}_{-,\ell} = n-1)] \\ &\quad + \sum_{k' \in \mathcal{K}\setminus\{k,\ell\}} \alpha_{k'}P_{k'k}\mathbb{E}_{s,k'}[(R^{(r)}_{s,k} \wedge c)1(Z^{(r)}_{-,\ell} = n-1)],\end{aligned} \tag{8.13}$$

$$\begin{aligned}\mathbb{P}(R^{(r)}_{e,k} \le c, Z_k^{(r)} \ge n) &= \mathbb{E}[T_{e,k} \wedge (c/a_k)]\mathbb{P}_{e,k}(Z^{(r)}_{-,k} \ge n-1) \\ &\quad - \alpha_k(1-P_{kk})\mathbb{E}_{s,k}[(R^{(r)}_{e,k} \wedge c)1(Z^{(r)}_{-,k} = n)] \\ &\quad + \sum_{\ell \in \mathcal{K}\setminus\{k\}} \alpha_\ell P_{\ell k}\mathbb{E}_{s,\ell}[(R^{(r)}_{e,k} \wedge c)1(Z^{(r)}_{-,k} = n-1)],\end{aligned} \tag{8.14}$$

$$\begin{aligned}\mathbb{P}(R^{(r)}_{e,k} \le c, Z_\ell^{(r)} \ge n) &= \mathbb{E}[T_{e,k} \wedge (c/a_k^{(r)})]\mathbb{P}_{e,k}(Z^{(r)}_{-,\ell} \ge n-1) - \alpha_\ell\mathbb{E}_{s,\ell}[(R^{(r)}_{e,k} \wedge c)1(Z_{-,\ell} = n)] \\ &\quad + \sum_{k' \in \mathcal{K}\setminus\{\ell\}} \alpha_{k'}P_{k'\ell}\mathbb{E}_{s,k'}[(R^{(r)}_{e,k} \wedge c)1(Z^{(r)}_{-,\ell} = n-1)],\end{aligned} \tag{8.15}$$

where expectations under $\mathbb{P}_{e,k}$ *and* $\mathbb{P}_{e,\ell}$ *vanish for* $k, \ell \notin \mathcal{E}$.

Because the proof of this lemma is lengthy, we defer it until we prove Lemma 8.4.

Proof of Lemma 8.4. We first prove (8.10) and (8.11) for $\ell = k \in \mathcal{H}$. For this, we use Lemma 8.5 and the SSC property,

$$\mathbb{P}[Z_k^{(r)} \ge n] = o(r), \qquad k \in \mathcal{H}, \tag{8.16}$$

which is immediate from the SSC Assumption 5.4. Set $n = \lfloor r^{-1} \rfloor$. Then, by (8.12),

$$(1 - P_{k,k})\mathbb{P}_{s,k}(Z_{-,k}^{(r)} = n) \le (1 - P_{k,k})\mathbb{P}_{s,k}(Z_{-,k}^{(r)} \ge n) \le \frac{1}{\gamma_k^{(r)}}\mathbb{P}(Z_k^{(r)} \ge n - 1). \tag{8.17}$$

This and (8.16) prove (8.11) for $\ell = k \in \mathcal{H}$ because $1 - P_{k,k} > 0$. We next note that $\mathbb{E}[T_{e,k}] = 1$ and $a_k^{(r)} \to a_k > 0$ as $r \downarrow 0$ for $k \in \mathcal{E}$, so we can find a $r_0 \in (0,1]$ and $c > 0$ such that

$$\mathbb{E}[T_{e,k} \wedge (c/a_k^{(r)})] \ge 1/2 \text{ for } r \in (0, r_0). \tag{8.18}$$

Therefore, it follows from (8.14) that for each $r \in (0, r_0)$,

$$\mathbb{P}_{e,k}(Z_{-,k}^{(r)} \ge n - 1) \le 2\mathbb{P}(Z_k^{(r)} \ge n) + 2c\alpha_k^{(r)}(1 - P_{k,k})\mathbb{P}_{s,k}(Z_{-,k}^{(r)} = n).$$

Because (8.11) is already proved for $\ell = k \in \mathcal{H}$, this and (8.16) prove (8.10) for $\ell = k$.

We next prove (8.10) and (8.11) for $\ell \in \mathcal{H}$ and $k \in \mathcal{K} \setminus \{\ell\}$. This time, we pick $r_0 \in (0,1]$ and $c > 0$ such that

$$\mathbb{E}[T_{s,k} \wedge c/m_k^{(r)}] \ge 1/2, \qquad r \in (0, r_0).$$

Then, it follows from (8.13) and (8.17) for $k = \ell$ that

$$\mathbb{P}_{s,k}(Z_{-,\ell}^{(r)} \ge n) \le \frac{2}{\gamma_k^{(r)}}[\mathbb{P}(Z_\ell^{(r)} \ge n) + c\alpha_\ell^{(r)}(1 - P_{\ell\ell})\mathbb{P}_{s,\ell}(Z_{-,\ell}^{(r)} = n)].$$

Because (8.11) is already proved for $k = \ell \in \mathcal{H}$, this inequality and (8.16) prove (8.11) for $\ell \in \mathcal{H}$ and $k \in \mathcal{K} \setminus \{\ell\}$. Similarly, applying (8.18) to (8.15), we have

$$\begin{aligned}\mathbb{P}_{e,k}(Z_{-,\ell}^{(r)} \ge n - 1) &\le 2\mathbb{P}(R_{e,k}^{(r)} \le c, Z_\ell^{(r)} \ge n) + 2\alpha_\ell \mathbb{E}_{s,\ell}[(R_{e,k}^{(r)} \wedge c)1(Z_{-,\ell} = n)] \\ &\le 2\mathbb{P}(Z_\ell^{(r)} \ge n) + 2\alpha_\ell \mathbb{P}_{s,\ell}[Z_{-,\ell} \ge n].\end{aligned}$$

Hence, (8.16) and (8.11) for $k = \ell \in \mathcal{H}$ prove (8.10) for $\ell \in \mathcal{H}$ and $k \in \mathcal{K} \setminus \{\ell\}$. □

Proof of Lemma 8.5. In this proof, we omit the superscript $^{(r)}$ in $Z_k^{(r)}$, $R_{-,e,k}^{(r)}$ and $R_{-,s,k}^{(r)}$ for simplicity. Therefore, they are written as Z_k, $R_{-,e,k}$ and $R_{-,s,k}$, respectively. To prove (8.12), we fix a $k \in \mathcal{K}$ and an integer $n \ge 1$. Take $f(x) = (v_k \wedge c)1(z_k \ge n)$. Then

$$\begin{aligned}\mathcal{A}f(X) &= -(R_{s,k} \le c, Z_k \ge n, Z_{H_+(k)} = 0), \\ \Delta f(X_+, X_-) &= (R_{-,s,k} \wedge c)1(Z_{-,k} = n - 1)1(R_{-,e,k} = 0) \\ &\quad + \sum_{\ell \in \mathcal{K} \setminus \{k\}} (R_{-,s,k} \wedge c)1(\Phi^{(\ell)} = e^{(k)})1(Z_{-,k} = n - 1, R_{-,s,\ell} = 0) \\ &\quad + (m_k T_{s,k} \wedge c)[1(\Phi^{(k)} \ne e^{(k)})1(Z_{-,k} - 1 \ge n) + 1(\Phi^{(k)} = e^{(k)})1(Z_{-,k} \ge n)]1(R_{-,s,k} = 0).\end{aligned}$$

By (6.16),

$$\begin{aligned}&\mathbb{P}(R_{s,k} \le c, Z_k \ge n, Z_{H_+(k)} = 0) \\ &\quad = \lambda_k \mathbb{E}_{e,k}[(R_{-,s,k} \wedge c)1(Z_{-,k} = n - 1)] + \sum_{\ell \in \mathcal{K} \setminus \{k\}} \alpha_\ell P_{\ell k} \mathbb{E}_{s,\ell}[(R_{-,s,k} \wedge c)1(Z_{-,k} = n - 1)] \\ &\qquad + \gamma_k \mathbb{E}[T_{s,k} \wedge c/m_k]([(1 - P_{kk})\mathbb{P}_{s,k}(Z_{-,k} \ge n + 1) + P_{kk}\mathbb{P}_{s,k}(Z_{-,k} \ge n)].\end{aligned}$$

Letting $c \to \infty$ in this equality proves (8.12) because the left-hand side is bounded by one, and all the terms in the right-hand side are nonnegative.

To prove (8.13), we fix $k, \ell \in \mathcal{K}$ with $k \neq \ell$, an integer $n \geq 1$, and $c \in \mathbb{R}_+$. Take $f(x) = (v_k \wedge c)1(z_\ell \geq n)$. Then

$$\mathcal{A}f(X) = -1(R_{s,k} \leq c, Z_\ell \geq n)1(Z_k > 0, Z_{H_+(k)} = 0),$$

$$\begin{aligned}\Delta f(X_+, X_-) = {}& (R_{s,k} \wedge c)1(Z_{-,\ell} + 1 = n)1(R_{-,e,\ell} = 0),\\ & - 1(\Phi^{(\ell)} \neq e^{(\ell)})(R_{s,k} \wedge c)1(Z_{-,\ell} = n)1(R_{-,s,\ell} = 0)\\ & + (m_k(T_{s,k} \wedge c)[1(\Phi^{(k)} = e^{(\ell)})1(Z_{-,\ell} + 1 \geq n)\\ & + 1(\Phi^{(k)} \neq e^{(\ell)})1(Z_{-,\ell} \geq n)]1(R_{-,s,k} = 0)\\ & + (R_{s,k} \wedge c) \sum_{k' \in \mathcal{K} \setminus \{k,\ell\}} 1(\phi^{(k')} = e^{(\ell)})1(Z_{-,\ell} + 1 = n)1(R_{-,s,k'} = 0).\end{aligned}$$

By (6.16),

$$\begin{aligned}&\mathbb{P}(R_{s,k} \leq c, Z_\ell \geq n, Z_k > 0, Z_{H_+(k)} = 0) = \lambda_\ell \mathbb{E}_{e,\ell}[(R_{s,k} \wedge c)1(Z_{-,\ell} = n-1)]\\ &- (1 - P_{\ell\ell})\alpha_\ell \mathbb{E}_{s,\ell}[(R_{s,k} \wedge c)1(Z_{-,\ell} = n)]\\ &+ \gamma_k \mathbb{E}[T_{s,k} \wedge (c/m_k)][P_{k\ell}\mathbb{P}_{s,k}(Z_{-,\ell} \geq n-1) + (1 - P_{k\ell})\mathbb{P}_{s,k}(Z_{-,\ell} \geq n)]\\ &+ \sum_{k' \in \mathcal{K} \setminus \{k,\ell\}} \alpha_{k'} P_{k'k} \mathbb{E}_{s,k'}[(R_{s,k} \wedge c)1(Z_{-\ell} = n-1)],\end{aligned}$$

which is equivalent to (8.13).

To prove (8.14), we fix $k \in \mathcal{E}$, an integer $n \geq 1$, and $c \in \mathbb{R}_+$. Take $f(x) = (u_k \wedge c)1(z_k \geq n)$. Then

$$\mathcal{A}f(X) = -1(R_{e,k} \leq c)1(Z_k \geq n),$$

$$\begin{aligned}\Delta f(X_+, X_-) = {}& (a_k T_{e,k} \wedge c)1(Z_{-,k} + 1 \geq n)1(R_{-,e,k} = 0)\\ & - (R_{e,k} \wedge c)1(\Phi^{(k)} \neq e^{(k)})1(Z_{-,k} = n)1(R_{-,s,k} = 0)\\ & + (R_{e,k} \wedge c) \sum_{\ell \in \mathcal{K} \setminus \{k\}} 1(\Phi^{(\ell)} = e^{(k)})1(Z_{-,k} = n-1)1(R_{-,s,\ell} = 0).\end{aligned}$$

By (6.16),

$$\begin{aligned}\mathbb{P}(R_{e,k} \leq c, Z_k \geq n) = {}& \mathbb{E}[T_{e,k} \wedge (c/a_k)]\mathbb{P}_{e,k}(Z_{-,k} \geq n-1)\\ & - \alpha_k(1 - P_{kk})\mathbb{E}_{s,k}[(R_{e,k} \wedge c)1(Z_{-,k} = n)]\\ & + \sum_{\ell \in \mathcal{K} \setminus \{k\}} \alpha_\ell P_{\ell k} \mathbb{E}_{s,\ell}[(R_{e,k} \wedge c)1(Z_{-,k} = n-1)],\end{aligned}$$

which proves (8.14). Finally, (8.15) is similarly proved using $f(x) = (u_k \wedge c)1(z_\ell \geq n)$. □

8.4. Proofs of Lemmas 8.2 and 8.3

This is the final step for completing the proof of Proposition 7.1.

Proof of Lemma 8.2. Because $\lambda_k^{(r)} \eta(r\theta, r^{1-\varepsilon_0}) = O(r)$ by (5.4) and (7.14) and $f^{(r)}_{r\theta, r, r^{1-\varepsilon_0}}(z)$ is uniformly bounded by Lemma 7.6, (8.6) is obtained if we show that $\mathbb{P}(R_{e,k}^{(r)} > r^{\varepsilon_0 - 1} a_k^{(r)}) = o(r)$. The latter holds because

$$\begin{aligned}\mathbb{P}(R_{e,k}^{(r)} > r^{\varepsilon_0 - 1} a_k^{(r)}) &\leq r^{1-\varepsilon_0}(a_k^{(r)})^{-1} \mathbb{E}[R_{e,k}^{(r)} 1(R_{e,k}^{(r)} > r^{\varepsilon_0 - 1} a_k^{(r)})]\\ &= \frac{r^{1-\varepsilon_0}}{2} \mathbb{E}[(T_{e,k}^2 - r^{2(\varepsilon_0 - 1)})1(T_{e,k} > r^{\varepsilon_0 - 1})]\\ &\leq \frac{r^{(1-\varepsilon_0)(1+\delta_0)}}{2} \mathbb{E}[T_{e,k}^{2+\delta_0} 1(T_{e,k} > r^{\varepsilon_0 - 1})] = o(r), \qquad r \downarrow 0,\end{aligned}$$

where the first equality follows from (6.22) in Lemma 6.4, and the second is due to $\mathbb{E}[T_{e,k}^{2+\delta_0}] < \infty$ and $(1 - \varepsilon_0)(1 + \delta_0) \geq 1$ for $\varepsilon_0 \in (0, \delta_0/(1 + \delta_0))$.

Similarly, (8.7) can be proved because

$$\begin{aligned}
&\mathbb{P}(R_{s,k}^{(r)} > r^{\varepsilon_0 - 1} m_k^{(r)}, Z_k^{(r)} > 0, Z_{H_+(k)}^{(r)} = 0) \\
&\leq r^{1-\varepsilon_0}(m_k^{(r)})^{-1}\mathbb{E}[R_{s,k}^{(r)} 1(R_{s,k}^{(r)} > r^{\varepsilon_0-1} m_k^{(r)}, Z_k^{(r)} > 0, Z_{H_+(k)}^{(r)} = 0)] \\
&= \frac{r^{1-\varepsilon_0}}{2}\gamma_k^{(r)}\mathbb{E}[(T_{s,k}^2 - r^{2(\varepsilon_0-1)})1(T_{s,k} > r^{\varepsilon_0-1})] \\
&\leq \frac{r^{(1-\varepsilon_0)(1+\delta_0)}}{2}\gamma_k^{(r)}\mathbb{E}[T_{s,k}^{2+\delta_0} 1(T_{s,k} > r^{\varepsilon_0-1})] = o(r),
\end{aligned}$$

where the equality follows from (6.23) in Lemma 6.4. □

We finally prove Lemma 8.3.

Proof of Lemma 8.3. We first prove (8.8). Following Definition (6.24),

$$\begin{aligned}
&\mathbb{E}_{e,k}[f(X_+^{(r)})] \\
&\quad = \mathbb{E}_{e,k}[f(X_-^{(r)}) e^{r\theta_k}\mathbb{E}_{e,k}[e^{-\eta_k(r\theta_k, r^{1-\varepsilon_0})(T_{e,k}\wedge r^{\varepsilon_0-1})} | X_-^{(r)}]] = \mathbb{E}_{e,k}[f(X_-^{(r)})], \quad k \in \mathcal{L}, &(8.19)\\
&\mathbb{E}_{e,k}[f(X_+^{(r)})] = \mathbb{E}_{e,k}[f(X_-^{(r)}) e^{r\theta_k((Z_{-,k}^{(r)}+1)\wedge(1/r) - Z_{-,k}^{(r)}\wedge(1/r)) - r\theta_k}], \qquad k \in \mathcal{H}. &(8.20)
\end{aligned}$$

Thus, (8.8) trivially holds for $k \in \mathcal{E} \cap \mathcal{L}$. Now, fix $k \in \mathcal{E} \cap \mathcal{H}$. For each $\ell \in \mathcal{K}$, $z_\ell \in \mathbb{Z}_+$, and $r \in (0,1]$, define

$$e^+(r, z_k) = (z_k + 1) \wedge (1/r) - z_k \wedge (1/r), \tag{8.21}$$

$$e^-(r, z_k) = (z_k - 1) \wedge (1/r) - z_k \wedge (1/r). \tag{8.22}$$

One can check that

$$e^+(r, z_k) - 1 = \begin{cases} 0 & \text{if } z_k \leq 1/r - 1, \\ r^{-1} - z_k - 1 & \text{if } 1/r - 1 < z_k < 1/r, \\ -1 & \text{if } 1/r \leq z_k. \end{cases}$$

It follows from this and similar expression for $e^-(r, z_k)$ that

$$|e^+(r, z_k) - 1| \leq 1(z_k > 1/r - 1), \ |e^-(r, z_k) + 1| \leq 1(z_k > 1/r). \tag{8.23}$$

Thus, we have

$$|r\theta_k e^+(r, z_k) - r\theta_k| \leq r|\theta_k| 1(z_k > 1/r - 1).$$

Hence, it follows from (7.26) and (8.20) that

$$\begin{aligned}
|\Delta f(X_+^{(r)}, X_-^{(r)})| 1(R_{-,e,k}^{(r)} = 0) &= |e^{r\theta_k e^+(r, Z_{-,k}^{(r)}) - r\theta_k} - 1| f(X_-^{(r)}) 1(R_{-,e,k}^{(r)} = 0) \\
&\leq r|\theta_k| e^{r|\theta_k|} 1(Z_{-,k}^{(r)} + 1 > 1/r) f(X_-^{(r)}) 1(R_{-,e,k}^{(r)} = 0) \\
&\leq r|\theta_k| 1(Z_{-,k}^{(r)} + 1 > 1/r) e^{|\theta_k| + |\theta_H| + (d_{e,a}E + d_{s,a}K)a} 1(R_{-,e,k}^{(r)} = 0),
\end{aligned}$$

where the last inequality follows from Lemma 7.6. Therefore, by Lemma 8.4,

$$\begin{aligned}
&|\mathbb{E}_{e,k}[\Delta f(X_+^{(r)}, X_-^{(r)}) 1(R_{-,e,k}^{(r)} = 0)]| \\
&\quad \leq r|\theta_k| e^{|\theta_k| + |\theta_H| + (d_{e,a}E + d_{s,a}K)a} \mathbb{P}_{e,k}\{Z_{-,k}^{(r)} > 1/r - 1\} = o(r^2).
\end{aligned}$$

We next prove (8.9). We first prove it for $k \in \mathcal{L}$. From Definition (6.10) and Lemma 6.3, under $\mathbb{P}_{s,k}$,

$$\begin{aligned}
&f(X_+^{(r)}) 1(R_{-,s,k}^{(r)} = 0) \\
&= f(X_-^{(r)}) 1(R_{-,s,k}^{(r)} = 0) \Bigg[\sum_{\ell \in \mathcal{L} \cup \{0\}} 1(\Phi^{(k)} = e^{(\ell)}) e^{-\theta_k + \theta_\ell} e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k}\wedge r^{\varepsilon_0-1})} \\
&\qquad + \sum_{\ell \in \mathcal{H}} 1(\Phi^{(k)} = e^{(\ell)}) e^{-\theta_k + \theta_\ell e^+\left(r, Z_{-,\ell}^{(r)}\right)} e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k}\wedge r^{\varepsilon_0-1})} \Bigg].
\end{aligned}$$

Hence, using the definition of $\xi_k(\theta, r)$ in (6.37) for $k \in \mathcal{L}$, which is

$$\sum_{\ell \in \overline{\mathcal{K}}} P_{k\ell} e^{r\theta_\ell - r\theta_k} \mathbb{E}[e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k} \wedge r^{\varepsilon_0 - 1})}] = 1,$$

we have

$$\begin{aligned}\mathbb{E}_{s,k}[\Delta f(X_+^{(r)}, X_-^{(r)})] &= \mathbb{E}_{s,k}[f(X_-^{(r)})] \sum_{\ell \in \mathcal{L} \cup \{0\}} P_{k,\ell} e^{-\theta_k + \theta_\ell} \mathbb{E}(e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k} \wedge r^{\varepsilon_0 - 1})}) \\ &\quad + \mathbb{E}_{s,k}\left[f(X_-^{(r)}) \sum_{\ell \in \mathcal{H}} P_{k,\ell} e^{-\theta_k + \theta_\ell e^+(r, Z_{-,\ell}^{(r)})}\right] \mathbb{E}_{s,k}(e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k} \wedge r^{\varepsilon_0 - 1})}) \\ &= \mathbb{E}_{s,k}\left[f(X_-^{(r)}) \sum_{\ell \in \mathcal{H}} P_{k,\ell}(e^{-\theta_k + \theta_\ell e^+(r, Z_{-,\ell}^{(r)})} - 1)\right] \mathbb{E}_{s,k}(e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k} \wedge r^{\varepsilon_0 - 1})}).\end{aligned}$$

This and Lemma 8.4 prove (8.9) because both $f_{r\theta, r, r^{1-\varepsilon_0}}^{(r)}(x)$ and $\mathbb{E}[e^{-\xi_k(r\theta_k, r^{1-\varepsilon_0})(T_{s,k} \wedge r^{\varepsilon_0 - 1})}]$ are bounded in r and

$$|e^{e^+(r, z_\ell) - r\theta_k} - e^{r\theta_\ell - r\theta_k}| = e^{r\theta_\ell - r\theta_k} |e^{e^+(r, z_\ell)} - 1| \le e^{r\theta_\ell - r\theta_k} e^{r|\theta_\ell|} r|\theta_\ell| 1(z_\ell + 1 > 1/r).$$

It remains to prove (8.9) for $k \in \mathcal{H}$, but we omit this proof because the result is obtained similarly to the case when $k \in \mathcal{L}$. □

Acknowledgments
The authors thank Xinyun Chen and Jin Guang for helpful discussions on Palm measure exposition.

Appendix A. Proof of Lemma 4.2
To prove (i) of Lemma 4.2, We first note that for any subset A of $\mathcal{L}$ and any vector $c_A \in \mathbb{R}_+^{\mathcal{L}}$ whose ℓth entry is $c_\ell 1(\ell \in A)$ with $c_\ell > 0$, $\phi_A(0-)$ and $\phi_{A,\ell}(0-)$ exist in the following sense:

$$\phi_A(0-) = \lim_{\alpha \uparrow 0} \phi(\alpha c_A), \quad \phi_{A,\ell}(0-) = \lim_{\alpha \uparrow 0} \phi_\ell(\alpha c_A), \qquad \alpha \in \mathbb{R},$$

because $\phi(\theta)$ and $\phi_\ell(\theta)$ are nondecreasing continuous functions from $\mathbb{R}_-^L$ to $[0, 1]$. Furthermore, these limits do not depend on c_i's as long as they are positive, as shown in lemma 5.1 of Braverman et al. (2017). Thus, $\phi_A(0-)$ and $\phi_{A,\ell}(0-)$ are well defined and satisfy (4.4).

To prove (ii), we use the tight system Assumption 5.5 to show that (R, b) is a tight system by verifying all the conditions in Definition 4.2 for $(x_A, x_A^{(j)}) \equiv (\phi_A, \phi_{j,A})$. Let us make a few observations about this. Because $\phi(\theta)$ is a monotone function, it follows that $\phi_A(0-) \ge \phi_{A'}(0-)$ if $A \subset A' \subset \mathcal{L}$; the same holds for $\phi_{A,\ell}(0-)$ and $\phi_{A',\ell}(0-)$. Furthermore, if $\ell \in A$, then $\phi_{A,\ell}(0-) = \phi_{A \setminus \{\ell\},\ell}(0-)$ since $\phi_{A,\ell}(\theta)$ does not depend on θ_ℓ. It remains to verify (4.3). This condition is equivalent to

$$\sum_{\ell \in \mathcal{L}} b_\ell R_{i,\ell}(\phi_{A,\ell}(0-) - \phi_A(0-)) = 0, \quad i \in A, A \subset \mathcal{L}. \tag{A.1}$$

We argue that this equation can be obtained from (7.13). To see this, we set $\theta = \alpha c_A$ in (7.13), divide both sides by α, and take $\alpha \uparrow 0$; then we have

$$\sum_{\ell \in \mathcal{L}} b_\ell \langle c_A, R^{(\ell)} \rangle (\phi_{A,\ell}(0-) - \phi_A(0-)) = 0.$$

For $i \in A$, letting $c_A \downarrow c_i e$ in this formula yields (A.1). Hence, all the conditions in Definition 4.2 are satisfied. By Assumption 5.5, (R, b) is a tight system. Thus,

$$\phi(0-) = x_L = 1, \qquad \phi_\ell(0-) = x_L^{(\ell)} = 1, \qquad \ell \in \mathcal{L},$$

which proves (ii) of Lemma 4.2.

Appendix B. Theorem 5.1 under Assumption 5.1A
If we replace Assumption 5.1 in Theorem 5.1 with Assumption 5.1A, then we cannot truncate the remaining times $R_{e,k}^{(r)}$ for $k \in \mathcal{E}$ and $R_{s,k}^{(r)}$ for $k \in \mathcal{K}_1$ in test functions in (7.17) for $X^{(r)}$ by $r^{\varepsilon_0 - 1}$ because (8.6) and (8.7) in Lemma 8.2 require the $(2 + \delta_0)$th moments of $T_{e,k}$ and those of $T_{s,k}$, respectively. Hence, under Assumption 5.1A, we need to truncate $R_{e,\ell}^{(r)}$ for $\ell \in \mathcal{E}$ and $R_{s,k}^{(r)}$ for $k \in \mathcal{K}_1$ in test functions by r. Namely, we need to choose test functions

$$f_\theta^{(r)}(x) = g_{\theta,r}(z) e^{\Lambda_\theta^{(r)}(u,v)}$$

to replace $f^{(r)}_{\theta,r,r^{1-\varepsilon_0}}(x)$ in (7.17) with

$$\Lambda^{(r)}_\theta(u,v) = \langle \eta(\theta,r), \lambda^{(r)}u \wedge r\rangle + \sum_{k\in\mathcal{K}_1} \xi_k(\theta,r)(\mu^{(r)}_k v_k \wedge r) + \sum_{k\in\mathcal{K}\setminus\mathcal{K}_1} \xi_k(\theta, r^{1-\varepsilon_0})(\mu^{(r)}_k v_k \wedge r^{\varepsilon_0-1}),$$

replacing $\Lambda^{(r)}_{\theta,r^{1-\varepsilon_0}}(u,v)$ in (7.18). This causes a problem in verifying (7.28). However, if $\mathbb{E}[R^{(r)}_{e,\ell}]$ for $\ell\in\mathcal{E}$ and $\mathbb{E}[R^{(r)}_{s,k}]$ for $k\in\mathcal{K}_1$ are uniformly bounded in $r\in(0,1]$, then

$$\limsup_{r\downarrow 0}\ \mathbb{E}(rR^{(r)}_{e,\ell}\wedge 1) \le \limsup_{r\downarrow 0}\ r\mathbb{E}(R^{(r)}_{e,\ell}) = 0, \qquad \ell\in\mathcal{E},$$

and similarly $\limsup_{r\downarrow 0} \mathbb{E}(rR^{(r)}_{s,k}\wedge 1) = 0$ for $k\in\mathcal{K}_1$. Hence,

$$\begin{aligned} &|\psi^{(r)}(r\theta) - \phi^{(r)}(r\theta)| \\ &\le |\theta| e^{|\theta_H| + (d_{e,a}E + d_{s,a}K)|\theta|}\Bigg(d_{e,a}\sum_{\ell\in\mathcal{E}} \lambda^{(r)}_\ell \mathbb{E}(rR^{(r)}_{e,\ell}\wedge 1) \\ &\quad + d_{s,a}\sum_{\ell\in\mathcal{K}_1}\mu^{(r)}_\ell \mathbb{E}(rR^{(r)}_{s,\ell}\wedge 1) + d_{s,a}r^{\varepsilon_0}\sum_{\ell\in\mathcal{K}\setminus\mathcal{K}_1}\mu^{(r)}_\ell\mathbb{E}(r^{1-\varepsilon_0}R^{(r)}_{s,\ell}\wedge 1)\Bigg) = o(1), \end{aligned} \tag{B.1}$$

where the inequality follows from an analogous version of (7.20). Thus, we have the first equation of (7.28), and its second equation is similarly proved. Hence, we need to prove only the uniform boundedness of $\mathbb{E}[R^{(r)}_{e,\ell}]$ for $\ell\in\mathcal{E}$ and $\mathbb{E}[R^{(r)}_{s,k}]$ for $k\in\mathcal{K}_1$. This is easily checked by Lemma 6.4. Namely, for $\ell\in\mathcal{E}$, (6.22) yields

$$2\mathbb{E}[R^{(r)}_{e,\ell}] = \lambda^{(r)}_\ell \mathbb{E}_{e,\ell}[T^2_{e,\ell}] < \infty. \tag{B.2}$$

Similarly for $k\in\mathcal{K}_1$, (6.23) yields

$$2\mathbb{E}[R^{(r)}_{s,k}1(Z^{(r)}_k > 0)] = \alpha^{(r)}_k \mathbb{E}(T^2_{s,k}) < \infty,$$

because $H_+(k) = \emptyset$. Because $\mathbb{E}[R^{(r)}_{s,k}1(Z^{(r)}_k = 0)] = m^{(r)}_k \mathbb{E}[T_{s,k}1(Z^{(r)}_k = 0)]$, (6.23) yields

$$\mathbb{E}[R^{(r)}_{s,k}] = \mathbb{E}[R^{(r)}_{s,k}1(Z^{(r)}_k > 0)] + \mathbb{E}[R^{(r)}_{s,k}1(Z_k = 0)] \le \frac{1}{2}\alpha^{(r)}_k \mathbb{E}(T^2_{s,k}) + \mathbb{E}[T_{s,k}] < \infty. \tag{B.3}$$

Thus, Theorem 5.1 can be proved under Assumption 5.1A. This proof also shows that if $\mathbb{E}(R^{(r)}_{s,k})$ is uniformly bounded for $\mathcal{K}\setminus\mathcal{K}_1$ in addition to Assumption 5.1A, then Theorem 5.1 can be proved under Assumption 5.1B. However, we have not been able to prove that $\mathbb{E}(R^{(r)}_{s,k})$ is uniformly bounded for $\mathcal{K}\setminus\mathcal{K}_1$ in this paper.

Appendix C. Taylor Expansions

In this section, we prove Lemmas 7.4 and 7.5 by deriving Taylor expansions for $\eta_k(\theta,w)$ and $\xi_k(\theta,w)$ defined by (6.29) and (6.30). The latter two quantities are defined in Braverman et al. (2017) with a sign difference. The following lemma is a key to this derivation, which also follow immediately from lemma 2.4 of Miyazawa (2017) and lemma 4.1 of Braverman et al. (2017).

Lemma C.1. *Let T be a nonnegative random variable with $\mathbb{E}(T) > 0$. For $t\in(0,1]$, let $T_t = T\wedge 1/t$. Then, there exists a unique function f_t: $\mathbb{R}\to\mathbb{R}$ for each fixed $t\in(0,1]$ such that*

$$\mathbb{E}(e^{-f_t(x)T_t}) = e^{-x}, \qquad x\in(-\infty, -\log\mathbb{P}(T=0)). \tag{C.1}$$

Here $-\log 0 = \infty$. Furthermore,

(i) *For each $x\in(0, -\log\mathbb{P}(T=0))$, $f_{0+}(x)\equiv \lim_{t\downarrow 0} f_t(x)$ exists and is finite.*

(ii) *For $x\in(-\infty,0]$ (resp. $x\in(0,-\log\mathbb{P}(T=0))$), $f_t(x)$ is decreasing (respectively, increasing) in $t\in(0,1]$ as t is decreasing. Hence, $\sup_{t\in(0,1]}|f_t(x)| \le |f_{0+}(x)| + |f_1(x)|$ for $x\in(0,-\log\mathbb{P}(T=0))$.*

(iii) *The function $f_t(x)$ is increasing, convex and infinitely differentiable in $x\in(0,-\log\mathbb{P}(T=0))$.*

(iv) *For any $t\in(0,1]$ and any $a\in(0,-\log\mathbb{P}(T=0))$,*

$$\left|f_t(x) - \lambda_{T_t}x - \frac{1}{2}\lambda^3_{T_t}\sigma^2_{T_t}x^2\right| \le \frac{x^2}{2}\sup_{|y|\le|x|}|f''_t(y) - f''_t(0)|, \tag{C.2}$$

$$|f_t(x)| \le \max(\lambda_{T_1}, (|f_{0+}(a)| + |f_1(a)|)/a)|x|, \quad |x| < a, \tag{C.3}$$

where $\lambda_{T_t} = 1/\mathbb{E}(T_t)$ and $\sigma^2_{T_t}$ is the variance of T_t, and

$$f''_t(y) = \frac{e^{-y}}{\mathbb{E}(T_t e^{-f_t(y)T_t})}\left(e^{-2y}\frac{\mathbb{E}(T^2_t e^{-f_t(y)T_t})}{[\mathbb{E}(T_t e^{-f_t(y)T_t})]^2} - 1\right), \qquad y\in\mathbb{R}. \tag{C.4}$$

Proof of Lemma 7.4. To prove the first inequality of (7.14), fix a $k \in \mathcal{E}$ and we apply Lemma C.1 for $T_{e,k}$ and $x = \theta_k \in \mathbb{R}$. Following definitions in (6.36) and (C.1), one has

$$\eta_k(\theta_k, t) = f_t(\theta_k).$$

Therefore, the inequality follows immediately from (C.3) of Lemma C.1.

To prove the second inequality of (7.14), fix a $k \in \mathcal{K}$, and we apply Lemma C.1 for $T_{s,k}$ and $x = x_s(\theta)$, where $\theta \in \mathbb{R}^K$ and

$$x_s(\theta) = \log\left(e^{-\theta_k}\sum_{\ell\in\overline{\mathcal{K}}} P_{k,\ell}e^{\theta_\ell}\right).$$

Following definitions in (6.37) and (C.1), one has

$$\xi_k(\theta, t) = f_t(x_s(\theta)).$$

The second inequality follows similarly from (C.3) of Lemma C.1 because

$$|x_s(\theta)| \le |\theta_k| + \log\left(\sum_{\ell\in\overline{\mathcal{K}}} P_{k,\ell}e^{|\theta_\ell|}\right) \le |\theta|(1+e^a), \quad |\theta| \le a,$$

by

$$\log\left(\sum_{\ell\in\overline{\mathcal{K}}} P_{k,\ell}e^{|\theta_\ell|}\right) \le \sum_{\ell\in\overline{\mathcal{K}}} P_{k,\ell}(e^{|\theta_\ell|}-1) \le \sum_{\ell\in\overline{\mathcal{K}}} P_{k,\ell}|\theta_\ell|e^{|\theta_\ell|} \le |\theta|e^{|\theta|},$$

where the first inequality from $\log y \le y-1$ for $y>0$, and the second inequality from $e^y - 1 \le ye^y$ for $y \ge 0$. □

To prove Lemma 7.5, we take $\eta_k(x,t)$ or $\xi_k(x,t)$ for $f_t(x)$, and put $x = r\theta$ or $x = x_s(r\theta)$, respectively, and $t = r^{1-\varepsilon_0}$, then let $r \downarrow 0$. Hence, it is sufficient to consider the convergence of $f_{r^{1-\varepsilon_0}}(x)$ as $r \downarrow 0$ for $0 < |x| \le ar$ for each positive constant a. By (C.3),

$$|f_{r^{1-\varepsilon_0}}(x)(T \wedge r^{\varepsilon_0 - 1})| \le a(rT \wedge r^{\varepsilon_0})\max(\lambda_{T_1}, (|f_{0+}(a)| + |f_1(a)|)/a), \quad |x| \le a.$$

Hence, for $0 < |x| \le ar$ and $t = r^{1-\varepsilon_0}$, $|f_t(x)T_t|$ is bounded by $a \times \max(\lambda_{T_1}, (|f_{0+}(a)| + |f_1(a)|)/a)$, which is independent of r, and vanishes as $|x| \downarrow 0$, and therefore $\mathbb{E}(T_t e^{-f_t(y)T_t})$ and $\mathbb{E}(T_t^2 e^{-f_t(y)T_t})$ converge to $\mathbb{E}(T_t)$ and $\mathbb{E}(T_t^2)$, respectively, as $|y| \downarrow 0$ uniformly in r for $|y| \le |x| \le ar$. Hence, by (C.4), $\sup_{|y|\le|x|} |f_t''(y) - f_t''(0)|$ in (C.2) vanishes uniformly in $r \in (0,1]$ as $|x| \downarrow 0$ if $|x| \le ar$. Thus, we have the following corollary of Lemma C.1.

Corollary C.1. *Under the same assumptions as Lemma C.1, for $t = r^{1-\varepsilon_0}$ and any constant $a > 0$, as $|x| \downarrow 0$,*

$$\sup_{r\in(a^{-1}|x|,1]}\left|f_t(x) - \lambda_{T_t}x - \frac{1}{2}\lambda_{T_t}^3\sigma_{T_t}^2x^2\right| = o(|x|^2). \tag{C.5}$$

Proof of Lemma 7.5. We first prove (7.15). Let $f_{r^{1-\varepsilon_0}}(r\theta_k) = \eta_k(r\theta_k, r^{1-\varepsilon_0})$. Then it follows from (C.5) with $T = T_{e,k}$ and $x = r\theta_k$ that

$$\eta_k(r\theta_k, r^{1-\varepsilon_0}) = r\theta_k\frac{1}{\mathbb{E}(T_{e,k}\wedge r^{\varepsilon_0-1})} + \frac{1}{2}r^2\theta_k^2\frac{\mathrm{Var}(T_{e,k}\wedge r^{\varepsilon_0-1})}{(\mathbb{E}(T_{e,k}\wedge r^{\varepsilon_0-1}))^3} + o(r^2), \qquad k \in \mathcal{E}.$$

This implies (7.15) if we can show that

$$\frac{1}{\mathbb{E}(T_{e,k}\wedge r^{\varepsilon_0-1})} - 1 = o(r), \qquad \frac{\mathrm{Var}(T_{e,k}\wedge r^{\varepsilon_0-1})}{(\mathbb{E}(T_{e,k}\wedge r^{\varepsilon_0-1}))^3} - c_{e,k}^2 = o(1). \tag{C.6}$$

We now choose $\varepsilon_0 > 0$ such that $\varepsilon_0 \le \delta_0/(1+\delta_0)$, then $(1-\varepsilon_0)(1+\delta_0) \ge 1$, and therefore the first formula is obtained from the fact that, as $r \downarrow 0$,

$$0 \le 1 - \mathbb{E}(T_{e,k}\wedge r^{\varepsilon_0-1}) \le \mathbb{E}(T_{e,k}1(T_{e,k} > r^{\varepsilon_0-1})) \le r\mathbb{E}(T_{e,k}^{2+\delta_0}1(T_{e,k} > r^{\varepsilon_0-1})) = o(r), \tag{C.7}$$

because $\mathbb{E}(T_{e,k}) = 1$ and $\mathbb{E}(T_{e,k}^2) < \infty$. Because $\mathrm{Var}(T_{e,k}) = c_{e,k}^2$, the second formula obviously holds. Thus, we have proved (7.15).

Finally, we prove (7.16). Similarly to (7.15), we apply (C.5) of Corollary C.1 for $T = T_{s,k}$ and $x = x_s(r\theta)$, then

$$\xi_k(r\theta, r^{1-\varepsilon_0}) = \frac{x_s(r\theta)}{\mathbb{E}(T_{s,k}\wedge r^{\varepsilon_0-1})} + \frac{1}{2}x_s^2(r\theta)\frac{\mathrm{Var}(T_{s,k}\wedge r^{\varepsilon_0-1})}{(\mathbb{E}(T_{s,k}\wedge r^{\varepsilon_0-1}))^3} + o(x_s^2(r\theta)), \qquad k \in \mathcal{E}.$$

Hence, using Taylor expansion of $x_s(r\theta)$ concerning r around the origin,

$$x_s(r\theta) = r\left(-\theta_k + \sum_{\ell\in K} P_{k,\ell}\theta_\ell\right) + \frac{1}{2}r^2\left(\sum_{\ell\in K} P_{k,\ell}\theta_\ell^2 - \left(\sum_{\ell\in K} P_{k,\ell}\theta_\ell\right)^2\right) + o(r^2), \tag{C.8}$$

and similar asymptotic behaviors to (C.6) for $T_{s,k}$, we have (7.16). □

References

Baccelli F, Brémaud P (2003) *Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences*, Applications of Mathematics, 2nd ed., vol. 26 (Springer, Berlin).

Baskett F, Chandy KM, Muntz RR, Palacios FG (1975) Open, closed, and mixed networks of queues with different classes of customers. *J. Assoc. Comput. Machine* 22:248–260.

Bramson M (1998) State space collapse for queueing networks. Rehmann U, Fischer G, eds. *Proc. Internat. Congress Mathematicians*, vol. III (Geronimo GmbH, Rosenheim, Germany), 213–222.

Bramson M, Dai JG (2001) Heavy traffic limits for some queueing networks. *Ann. Appl. Probability* 11(1):49–90.

Braverman A, Dai J, Miyazawa M (2017) Heavy traffic approximation for the stationary distribution of a generalized Jackson network: The BAR approach. *Stochastic Systems* 7(1):143–196.

Cao C, Dai J, Zhang X (2022) Steady-state state space collapse in muliclass queueing networks. *Queueing Systems* 102:87–122.

Chen H, Ye HQ (2001) Existence condition for the diffusion approximations of multiclass priority queueing networks. *Queueing Systems* 38(4):435–470.

Chen H, Zhang H (1997) Stability of multiclass queueing networks under FIFO service discipline. *Math. Oper. Res.* 22:691–725.

Chen H, Zhang H (2000a) Diffusion approximations for some multiclass queueing networks with FIFO service disciplines. *Math. Oper. Res.* 25:679–707.

Chen H, Zhang H (2000b) A sufficient condition and a necessary condition for the diffusion approximations of multiclass queueing networks under priority service disciplines. *Queueing Systems* 34(1–4):237–268.

Dai JG (1995) On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Probability* 5(1):49–77.

Dai J, Harrison JM (2020) *Processing Networks: Fluid Models and Stability* (Cambridge University Press, Cambridge, UK).

Dai JG, Kurtz TG (1994) Characterization of the stationary distribution for a semimartingale reflecting Brownian motion in a convex polyhedron. Preprint.

Dai JG, Vande Vate JH (2000) The stability of two-station multitype fluid networks. *Oper. Res.* 48(5):721–744.

Dai JG, Weiss G (1996) Stability and instability of fluid models for reentrant lines. *Math. Oper. Res.* 21(1):115–134.

Dai J, Glynn P, Xu Y (2023) Asymptotic product-form steady-state for generalized Jackson networks in multi-scale heavy traffic. Preprint, submitted April 4, https://arxiv.org/abs/2304.01499.

Dai J, Ji Y, Miyazawa M (2024) Tight matrices and heavy traffic steady-state convergence in queueing networks. Preprint, submitted April 21, https://arxiv.org/abs/2404.13651.

Dai JG, Miyazawa M, Wu J (2014) A multi-dimensional SRBM: Geometric views of its product form stationary distribution. *Queueing Systems* 78(4):313–335.

Davis MHA (1984) Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* 46(3):353–388.

Eryilmaz A, Srikant R (2012) Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems* 72(3–4):311–359.

Ethier SN, Kurtz TG (1986) *Markov Processes: Characterization and Convergence,* Probability and Mathematical Statistics (John Wiley & Sons, New York).

Gamarnik D, Zeevi A (2006) Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Ann. Appl. Probability* 16(1):56–90.

Glynn PW, Zeevi A (2008) Bounding stationary expectations of Markov processes. Ethier S, Feng J, Stockbridge R, eds. *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*, vol. 4 (Institute of Math and Statistics, Beachwood, OH), 195–214.

Guang J, Chen X, Dai J (2024) Uniform moment bounds for generalized Jackson networks in multi-scale heavy traffic. Preprint, submitted January 29, https://arxiv.org/abs/2401.14647.

Gurvich I (2014) Validity of heavy-traffic steady-state approximations in multiclass queueing networks: The case of queue-ratio disciplines. *Math. Oper. Res.* 39(1):121–162.

Harrison JM (1988) Brownian models of queueing networks with heterogeneous customer populations. Fleming W, Lions PL, eds. *Stochastic Differential Systems, Stochastic Control Theory and Applications*, The IMA Volumes in Mathematics and Its Applications, vol. 10 (Springer, New York), 147–186.

Harrison JM, Williams RJ (1987) Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* 22(2): 77–115.

Hurtado-Lange D, Maguluri ST (2020) Transform methods for heavy-traffic analysis. *Stochastic Systems* 10(4):275–309.

Johnson DP (1983) Diffusion approximations for optimal filtering of jump processes and for queueing networks. PhD thesis, University of Wisconsin-Madison, Madison, WI.

Kallenberg O (2001) *Foundations of Modern Probability,* Statistics, Probability and Its Applications (Springer, New York).

Kelly FP (1975) Networks of queues with customers of different types. *J. Appl. Probability* 12:542–554.

Maguluri ST, Srikant R (2016) Heavy traffic queue length behavior in a switch under the maxweight algorithm. *Stochastic Systems* 6(1): 211–250.

Miyazawa M (1991) The characterization of the stationary distribution of the supplemented self-clicking jump process. *Math. Oper. Res.* 16(3):547–565.

Miyazawa M (1994) Rate conservation laws: A survey. *Queueing Systems* 15:1–58.

Miyazawa M (2017) A unified approach for large queue asymptotics in a heterogeneous multiserver queue. *Adv. Appl. Probability* 49(1):182–220.

Miyazawa M (2024) Palm problems arising in BAR approach and its applications. Preprint, submitted March 14, https://arxiv.org/abs/2308.03553.

Reiman MI (1984) Open queueing networks in heavy traffic. *Math. Oper. Res.* 9:441–458.

Serfozo R (1999) *Introduction to Stochastic Networks*, vol. 44 of Applications of Mathematics (Springer-Verlag, New York).

Stolyar AL (2004) Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *Ann. Appl. Probability* 14(1):1–53.

Wang W, Maguluri ST, Srikant R, Ying L (2022) Heavy-traffic insensitive bounds for weighted proportionally fair bandwidth sharing policies. *Math. Oper. Res.* 47(4):2691–2720.

Williams RJ (1998) Diffusion approximations for open multiclass queueing networks: Sufficient conditions involving state space collapse. *Queueing Systems* 30:27–88.

Ye HQ, Yao DD (2016) Diffusion limit of fair resource control—Stationarity and interchange of limits. *Math. Oper. Res.* 41(4):1161–1207.

Ye HQ, Yao DD (2018) Justifying diffusion approximations for multiclass queueing networks under a moment condition. *Annals Appl. Probability* 28(6):3652–3697.