Check for updates

Adv. Appl. Probab. **57,** 1167–1205 (2025)
*doi:*10.1017/apr.2025.5

DIFFUSION APPROXIMATION OF THE STATIONARY DISTRIBUTION OF A TWO-LEVEL SINGLE SERVER QUEUE

MASAKIYO MIYAZAWA,* *Tokyo University of Science*

Abstract

We consider a single server queue that has a threshold to change its arrival process and service speed by its queue length, which is referred to as a two-level *GI/G/1* queue. This model is motivated by an energy saving problem for a single server queue whose arrival process and service speed are controlled. To obtain its performance in tractable form, we study the limit of the stationary distribution of the queue length in this two-level queue under scaling in heavy traffic. Except for a special case, this limit corresponds to its diffusion approximation. It is shown that this limiting distribution is truncated exponential (or uniform if the drift is null) below the threshold level and exponential above it under suitably chosen system parameters and generally distributed interarrival times and workloads brought by customers. This result is proved under a mild limitation on arrival parameters using the so-called basic adjoint relationship (BAR) approach studied in Braverman, Dai, and Miyazawa (2017, 2024) and Miyazawa (2017, 2024). We also intuitively discuss about a diffusion process corresponding to the limit of the stationary distribution under scaling.

Keywords: single server queue; level dependent arrival and service; stationary distribution; heavy traffic; diffusion approximation; Palm distribution; stochastic integral equation; multi-level SRBM.

2010 Mathematics Subject Classification: Primary 60K25; 60J70
Secondary 60J25; 60G55; 60M20

1. Introduction

We consider a single server queue that has a threshold to change its arrival process and service rate by its queue length. Namely, customers arrive subject to a renewal process as long as its queue length (including a customer being served) is less than threshold ℓ_1, while it is changed to another renewal process otherwise. The service speed is changed similarly but in a slightly different way. That is, the service speed is constant when the queue length is not greater than the threshold, while it is another constant otherwise. We assume that customers are served in the first-come–first-served manner, and the amounts of work that are brought by them are independent and identically distributed (i.i.d.) and independent of the arrival process. We refer to this model as a two-level $GI/G/1$ queue.

Received 5 December 2024; accepted 16 January 2025.
* Postal address: Yamazaki 2641, Noda, Chiba, Japan, 278-8051. Email address: miyazawa@rs.tus.ac.jp

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This two-level $GI/G/1$ queue is motivated by an energy saving problem to control the arrival process and service speed. Here, the system is an internet server, and its energy consumption depends on service speed. Then, the problem is to find an optimal threshold to minimize the total cost of the operating energy of the server and customer's waiting.

So far, we are interested in the performance of this two-level $GI/G/1$ queue over a long time, for example, the mean queue length. However, it is hard to derive such a characteristic in an analytically tractable form in general, while, for the special case that all the distributions are exponential, its analytic expression is available, but the variability of the arrival times and service amounts are not well captured in this exponential case. Because of these facts, an approximation by a reflected diffusion is helpful, where a reflected diffusion means a one-dimensional diffusion reflected at the origin, in which the term 'diffusion' is used for a one-dimensional continuous Markov process on the real line that has a finite and positive second moment at each time.

So we study the diffusion approximation through the limit of the stationary distributions of the scaled queue length of this two-level $GI/G/1$ queue in heavy traffic, where the scaling factor and threshold level are chosen so that the limit is well obtained as a probability distribution. It is expected that this limiting distribution corresponds to the stationary distribution of a reflected diffusion whose drift and variance are discontinuously changed at the level corresponding to the threshold (see Section 6.1). The stationary distribution of this reflected diffusion is recently studied in [11, 13]. Thus, we will indirectly consider a diffusion approximation for the two-level $GI/G/1$ queue.

There is another motivation of this study. In general, it is hard to theoretically study diffusion approximation for state-dependent queues. For example, there are some theoretical studies about them, but they assume that the arrival and service processes are of exponential types, and their rates are smoothly changed (see, e.g., [8, 15]), while the arrival process and service speed discontinuously change in the present model. Thus, a theoretical study for its diffusion approximation is challenging.

As is well known, the scaling limit of the stationary distribution is exponential for the standard $GI/G/1$ queue in heavy traffic (see, e.g., [7]). Furthermore, [12] shows that the limiting distribution is truncated exponential (or uniform for the null drift case) for the $GI/G/1$ queue with a finite waiting room. Both distributions agree with the stationary distributions of the corresponding reflected diffusions (see [6]). Thus, one may conjecture that the limiting distribution for the two-level $GI/G/1$ queue is truncated exponential or uniform below the threshold level and exponential above it. This conjecture can be directly verified if all the distributions are exponential (see Proposition 1), but this is exceptional. Under the generally distributed assumptions, we verify it, employing the basic adjoint relationship (BAR) approach studied in [3, 4, 10, 12]. This verification, Theorem 1, requires a mild limitation on the parameters of the arrival process (see (f) in Section 3.1). Its proof needs some knowledge about Palm distributions and a truncation technique in computations, but does not require any deeper theory. However, the proof is complicated, so we split it into several components so that they are orthogonal as long as possible. This is the reason why we have devoted two sections to the proof.

In this paper we only study the limit of the scaled stationary distribution, but, as our results are related to the diffusion approximation, it is interesting to identify a reflected diffusion as a process limit for the two-level $GI/G/1$ queue. This is a hard question because this diffusion discontinuously changes its drift and variance at the fixed state. We intuitively discuss what can be expected about this process limit in Section 6.1.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This paper is made up by six sections. We first introduce a general framework for a Markov process describing the two-level $GI/G/1$ queue in Section 2, then provide notations and assumptions for a sequence of the two-level $GI/G/1$ queues in heavy traffic in Section 3.1. A main result, Theorem 1, and the exponentially distributed case, Proposition 1, are presented in Section 3.2. In Section 4 we introduce the so-called BAR approach, and apply it to the two-level $GI/G/1$ queue. Using these results, Theorem 1 is proved in Section 5. Finally, in Section 6 we intuitively discuss two topics. In Section 6.1 we consider a process limit for the two-level $GI/G/1$ queue in heavy traffic. In Section 6.2 we intuitively verify the conjecture in Remark 2.

2. Framework for the two-level *GI/G/1* queue

We take the stochastic framework introduced in [12] to study the two-level $GI/G/1$ queue. Let $L(t)$, $R_e(t)$, and $R_s(t)$ be the queue length, the remaining arrive time, and the remaining workload of a customer being served, respectively, at time t. Note that the remaining service time cannot be used for $R_s(t)$ here because the service speed may change in the future. Define

$$X(\cdot) = \{(L(t), R_e(t), R_s(t)); t \geq 0\}.$$

Then, it is easy to see that $X(\cdot)$ is a Markov process with state space $\mathbb{Z}_+ \times \mathbb{R}_+^2$, and the two-level $GI/G/1$ queue is described by $X(\cdot)$, where $\mathbb{Z}_+$ is the set of nonnegative integers and $\mathbb{R}_+ = [0, \infty)$. It can be assumed that $X(t)$ is right continuous and has left limits at any point of time. Let $N_e(\cdot) \equiv \{N_e(t); t \geq 0\}$ and $N_s(\cdot) \equiv \{N_s(t); t \geq 0\}$ be the counting processes generated by the vanishing instants of $R_e(t)$ and $R_s(t)$, respectively. Note that those counting processes can be viewed as point processes that are integer-valued random measures on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ whose atoms are the counting instants of $N_e(\cdot)$ and $N_s(\cdot)$, where $\mathcal{B}(\mathbb{R}_+)$ is the Borel field on $\mathbb{R}_+$. We denote these point processes by N_e and N_s, respectively.

According to [12], we take stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with time-shift operator semi-group $\Theta_\bullet \equiv \{\Theta_t; t \geq 0\}$ to jointly consider the Markov process $X(\cdot)$ and the counting processes $N_e(\cdot)$ and $N_s(\cdot)$, where $\mathbb{F} \equiv \{\mathcal{F}_t; t \geq 0\}$ is a filtration generated by $X(\cdot)$, that is, $\mathcal{F}_t = \sigma(\{X(u); 0 \leq u \leq t\})$, and $\Theta_\bullet$ is an operator semi-group (namely, $\Theta_{s+t} = \Theta_s \circ \Theta_t$ for $s, t \geq 0$) on Ω satisfying the following two conditions.

(i) Measurability:

$$A \in \mathcal{F}_t \text{ implies } \Theta_s^{-1} A \in \mathcal{F}_{s+t}, \qquad s, t \geq 0,$$

(ii) Consistency:

$$X(t) \circ \Theta_s(\omega) = X(s+t)(\omega), \qquad s, t \geq 0, \omega \in \Omega,$$
$$N(t) \circ \Theta_s(\omega) = N(s+t)(\omega) - N(s)(\omega), \quad s, t \geq 0, \omega \in \Omega,$$
$$\text{for } N(\cdot) = N_e(\cdot) \quad \text{and} \quad N(\cdot) = N_s(\cdot),$$

where $W \circ \Theta_s(\omega) = W(\Theta_s(\omega))$ for random variable W, and $\Theta_s^{-1} A = \{\omega \in \Omega; \Theta_s(\omega) \in A\}$.

3. A sequence of the two-level $GI/G/1$ queues

We consider the scaling limit of the stationary distribution of the two-level $GI/G/1$ queue in heavy traffic. To this end, we introduce a sequence of this queueing model indexed by $r \in (0, 1]$. Since this sequence can be countable, one may take $r = 1/\sqrt{n}$ for positive integer n so that the scaling is identical with the standard diffusion approximation, but we keep r just

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

for convenience. Let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$ for $a, b \in \mathbb{R}$, which will be used throughout the paper.

3.1. Assumptions

For the rth model, let $\ell_1^{(r)} \geq 1$ be the threshold level, and let $L^{(r)}(t)$ be the number of customers in the system at time t. We refer to this number as the queue length. Let $t_{e,n}^{(r)}$ be the nth arrival time of a customer, and let $T_s^{(r)}(n)$ be the workload of the nth arriving customer. We assume that $T_e^{(r)}(n) \equiv t_{e,n}^{(r)} - t_{e,n-1}^{(r)}$ is independent of $\{t_{e,n'}^{(r)}; 0 \leq n' \leq n-1\}$ and everything else for all $n \geq 1$, where $t_{e,0}^{(r)} = 0$. Thus, the interarrival times are independent, but their distributions depend on the queue length. Namely, $T_e^{(r)}(n)$ has the same distribution as $T_{e,1}^{(r)}$ if $L(t_{e,n-1}^{(r)} -) < \ell_1^{(r)}$ while its distribution agrees with that of $T_{e,2}^{(r)}$ if $L(t_{e,n-1}^{(r)} -) \geq \ell_1^{(r)}$, where $T_{e,1}^{(r)}$ and $T_{e,2}^{(r)}$ are independent random variables that may have different distributions. As for service, let $T_s^{(r)}(n)$ be the workload of the nth arriving customers. We assume that $\{T_s^{(r)}(n); n \geq 1\}$ is a sequence of i.i.d. random variables. We denote a random variable subject to the common distribution of $T_s^{(r)}(n)$ by $T_s^{(r)}$. Let $c_1^{(r)}$ and $c_2^{(r)}$ be the service speeds when the queue length is not greater than $\ell_1^{(r)}$ and otherwise, respectively. Namely, as long as $R_s^{(r)}(t)$ is differentiable at time t,

$$\frac{d}{dt} R_s^{(r)}(t) = -c_1^{(r)} 1(1 \leq L^{(r)}(t) \leq \ell_1^{(r)}) - c_2^{(r)} 1(L^{(r)}(t) > \ell_1^{(r)}),$$

where $1(\cdot)$ is the indicator function of the statement '$\cdot$'. These dynamics is depicted in the sample path of $L^{(r)}(t)$ in Figure 1.

We will use the following notations:

$$m_{e,i}^{(r)} = \mathbb{E}(T_{e,i}^{(r)}), \qquad \lambda_i^{(r)} = \frac{1}{m_{e,i}^{(r)}}, \qquad \sigma_{e,i}^{(r)} = (\mathbb{E}[(T_{e,i}^{(r)} - m_{e,i}^{(r)})^2])^{\frac{1}{2}}, \quad i = 1, 2,$$

$$m_s^{(r)} = \mathbb{E}(T_s^{(r)}), \qquad \mu^{(r)} = \frac{1}{m_s^{(r)}}, \qquad \sigma_s^{(r)} = (\mathbb{E}[(T_s^{(r)} - m_s^{(r)})^2])^{\frac{1}{2}},$$

$$\rho_i^{(r)} = \frac{\lambda_i^{(r)}}{c_i^{(r)} \mu^{(r)}}, \quad i = 1, 2.$$

For the rth model, let $R_e^{(r)}(t)$ and $R_s^{(r)}(t)$ be the remaining arrival time and the remaining workload of a customer being served, respectively, at time t. Define $X^{(r)}(\cdot) \equiv \{X^{(r)}(t);\ t \geq 0\}$ by

$$X^{(r)}(t) = (L^{(r)}(t), R_e^{(r)}(t), R_s^{(r)}(t)),$$

and denote the counting processes of the arrivals and service completions by $N_e^{(r)}(\cdot)$ and $N_s^{(r)}(\cdot)$, respectively. Note that $X^{(r)}(\cdot)$ is a piecewise deterministic Markov process, so it is strong Markov (see, e.g., [5]). Obviously, $X^{(r)}(\cdot)$, $N_e^{(r)}(\cdot)$, and $N_s^{(r)}(\cdot)$ are consistent with the time-shift operator $\Theta_\bullet$. We assume the following assumptions.

(A1) $\{\mathbb{E}[(T_{e,i}^{(r)})^3]; r \in (0, 1]\}$ for $i = 1, 2$ and $\{\mathbb{E}[(T_s^{(r)})^3]; r \in (0, 1]\}$ are bounded.

(A2) For $i = 1, 2$, $m_{e,i}^{(r)}$, $\lambda_i^{(r)}$, $\sigma_{e,i}^{(r)}$, $c_i^{(r)}$, and $m_s^{(r)}$, $\mu^{(r)}$, $\sigma_s^{(r)}$ are finite and positive, and converge to $m_{e,i}$, λ_i, $\sigma_{e,i}$, $c_i > 0$ and m_s, μ, $\sigma_s > 0$, respectively, as $r \downarrow 0$.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

(A3) There are constants $b_1 \in \mathbb{R}$ and $b_2 > 0$ such that $c_i^{(r)}\mu^{(r)} - \lambda_i^{(r)} = r\mu^{(r)}b_i + o(r)$ as $r \downarrow 0$.

(A4) There is a constant ℓ_1 such that $r\ell_1^{(r)} = \ell_1 + o(r)$ as $r \downarrow 0$.

(A5) $\rho_2^{(r)} < 1$.

Here, for functions $f, g\colon (0, 1] \to (0, \infty)$, we write $f(r) = o(g(r))$ as $r \downarrow 0$ if $\lim_{r\downarrow 0} f(r)/g(r) = 0$. Similarly, $f(r) = O(g(r))$ if $\limsup_{r\downarrow 0} f(r)/g(r) < \infty$. Note that (A1) implies that

(A1) $\{(T_{e,i}^{(r)})^2; r \in (0, 1]\}$ for $i = 1, 2$ and $\{(T_s^{(r)})^2; r \in (0, 1]\}$ are uniformly integrable.

Furthermore, define $\rho_i = \lambda_i/(c_i\mu)$, then (A2) and (A3) imply that $1 - \rho_i^{(r)} = rb_i/c_i + o(r)$, $\rho_i = 1$ and $\lambda_i = c_i\mu$. Hence, $\lambda_1 = \lambda_2$ if and only if $c_1 = c_2$.

Most of our arguments are valid under the assumptions (A1)–(A5), but we require the following condition for their final step to determine the limiting distribution of the scaled queue length in heavy traffic:

(A6) $\sigma_{e,1}^2 = \sigma_{e,2}^2$ if $\lambda_1 = \lambda_2$.

Remark 1. This is a mild limitation because no condition is needed for $\lambda_1 \neq \lambda_2$ and it is automatically satisfied when $T_{e,1}$ and $T_{e,2}$ are identically distributed. This condition is likely not needed but hard to remove. See Remark 6 for its details.

We take countably many rs for a limiting operation. For such a sequence, there exists a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with time-shift operator $\Theta_\bullet$ on which $X^{(r)}(\cdot)$ is a stationary process. In what follows, $\mathbb{P}$ is always this probability measure. By the condition (A5), the Markov process $X^{(r)}(\cdot)$ has the stationary distribution. Let $X^{(r)} \equiv (L^{(r)}, R_e^{(r)}, R_s^{(r)})$ be a random vector subject to the stationary distribution of $X^{(r)}(t)$. For convenience, let $R^{(r)} = (R_e^{(r)}, R_s^{(r)})$. For $i = 1, 2$, let $N_{e,i}^{(r)}$ be the stationary point process obtained from the renewal process whose inter-counting times have the same distribution as that of $T_{e,i}^{(r)}$. Similarly, let $N_s^{(r)}$ be the stationary point process for the service completion instants.

Define $\alpha_e^{(r)} = \mathbb{E}[N_e^{(r)}(1)]$ and $\alpha_s^{(r)} = \mathbb{E}[N_s^{(r)}(1)]$. At present, we cannot compute $\alpha_e^{(r)}$ and $\alpha_s^{(r)}$, but the following lemma is intuitively clear because $X^{(r)}(\cdot)$ is a stationary process by (A5). Since this lemma is a key for our analysis, we formally prove it in Appendix A.1.

Lemma 1. *Assume the assumptions (A2) and (A5) hold, then $\alpha_e^{(r)}$ is finite and positive, and $\alpha_s^{(r)} = \alpha_e^{(r)}$.*

3.2. Main results

We now present a main result, which will be proved in Section 5. Recall that $L^{(r)}$ is a random variable subject to the stationary distribution of the rth two-level $GI/G/1$ queue. Define probability distribution $\nu^{(r)}$ under the assumptions (A1)–(A5) as

$$\nu^{(r)}(B) = \mathbb{P}(rL^{(r)} \in B), \quad B \in \mathcal{B}(\mathbb{R}_+).$$

We refer to this $\nu^{(r)}$ as a scaled stationary distribution. Define finite measures $\nu_i^{(r)}$ and conditional probability distributions $\nu_{i|cd}^{(r)}$ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ for $i = 1, 2$ as

$$\nu_1^{(r)}(B) = \mathbb{P}(rL^{(r)} \in B; L^{(r)} \le \ell_1^{(r)}), \qquad \nu_2^{(r)}(B) = \mathbb{P}(rL^{(r)} \in B; \ell_1^{(r)} < L^{(r)}),$$
$$\nu_{1|cd}^{(r)}(B) = \mathbb{P}(rL^{(r)} \in B | L^{(r)} \le \ell_1^{(r)}), \qquad \nu_{2|cd}^{(r)}(B) = \mathbb{P}(rL^{(r)} \in B | \ell_1^{(r)} < L^{(r)}),$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

for $B \in \mathcal{B}(\mathbb{R}_+)$. Obviously, $\nu^{(r)}(B) = \nu_1^{(r)}(B) + \nu_2^{(r)}(B)$.

Theorem 1. *Assume the assumptions (A1)–(A6) hold. Let*

$$\sigma_i^2 = \lambda_i^2 \sigma_{e,i}^2 + \mu^2 \sigma_s^2, \qquad \beta_i = \frac{2b_i}{c_i \sigma_i^2}, \quad i = 1, 2. \tag{3.1}$$

For each $r \in (0, 1]$, define $\widetilde{\nu}_{i|cd}$ as the probability measures on $\mathcal{B}(\mathbb{R}_+)$ that have the density functions $h_{i|cd}$ for $i = 1, 2$ such that

$$h_{1|cd}(x) = 1(0 \le x \le \ell_1) \times \begin{cases} \dfrac{1}{\ell_1}, & b_1 = 0, \\ \dfrac{e^{\beta_1 \ell_1} \beta_1}{e^{\beta_1 \ell_1} - 1} e^{-\beta_1 x}, & b_1 \ne 0, \end{cases} \tag{3.2}$$

$$h_{2|cd}(x) = 1(x > \ell_1) \beta_2 e^{-\beta_2 (x - \ell_1)}. \tag{3.3}$$

(i) For each $i = 1, 2$, if $\nu_i^{(r)}(\mathbb{R}_+)$ is bounded away from 0 as $r \downarrow 0$, that is,

$$\liminf_{r \downarrow 0} \varphi_i^{(r)}(0) > 0, \quad i = 1, 2, \tag{3.4}$$

then $\nu_{i|cd}^{(r)}$ weakly converges to $\widetilde{\nu}_{i|cd}$, namely,

$$\lim_{r \downarrow 0} \nu_{i|cd}^{(r)}(B) = \widetilde{\nu}_{i|cd}(B), \quad B \in \mathcal{B}(\mathbb{R}_+). \tag{3.5}$$

(ii) Define the probability measure $\widetilde{\nu}$ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ and constant $\widetilde{\alpha}_e$ as

$$\widetilde{\nu}(B) = \widetilde{\nu}_{1|cd}(B) \widetilde{A}_1 + \widetilde{\nu}_{2|cd}(B) \widetilde{A}_2, \qquad B \in \mathcal{B}(\mathbb{R}_+), \tag{3.6}$$

$$\widetilde{\alpha}_e = \lambda_1 \widetilde{A}_1 + \lambda_2 \widetilde{A}_2, \tag{3.7}$$

where $\widetilde{A}_i$ for $i = 1, 2$ are given by

$$\widetilde{A}_1 = \begin{cases} \dfrac{2b_2 \ell_1}{c_2 \sigma_1^2 + 2b_2 \ell_1}, & b_1 = 0, \\ \dfrac{c_1 b_2 (e^{\beta_1 \ell_1} - 1)}{c_2 b_1 + c_1 b_2 (e^{\beta_1 \ell_1} - 1)}, & b_1 \ne 0, \end{cases} \qquad \widetilde{A}_2 = 1 - \widetilde{A}_1. \tag{3.8}$$

If either $\lambda_1^{(r)} - \lambda_2^{(r)} = O(r)$, equivalently, $c_1^{(r)} - c_2^{(r)} = O(r)$, or $T_s^{(r)}$ is exponentially distributed for all $r \in (0, 1]$, then

$$\lim_{r \downarrow 0} \nu^{(r)}(B) = \widetilde{\nu}(B), \qquad B \in \mathcal{B}(\mathbb{R}_+), \tag{3.9}$$

$$\lim_{r \downarrow 0} \alpha_e^{(r)} = \widetilde{\alpha}_e. \tag{3.10}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

(iii) If $\lambda_1 > \lambda_2$ ($\lambda_1 < \lambda_2$) then $\nu_{1|cd}^{(r)}$ ($\nu_{2|cd}^{(r)}$) weakly converges to $\widetilde{\nu}_{1|cd}$ ($\widetilde{\nu}_{2|cd}$, respectively), as $r \downarrow 0$.

Remark 2. Note that, under one of the conditions of part (ii) of Theorem 1, we have

$$\lim_{r\downarrow 0} \nu_i^{(r)}(\mathbb{R}_+) = \widetilde{A}_i, \quad i = 1, 2. \tag{3.11}$$

Conversely, if (3.11) holds then we have the weak convergence (3.9) from (i) because $\widetilde{A}_i > 0$. Thus, (3.11) is equivalent to (3.9). We conjecture that (3.9) holds without any condition in (ii). This is supported by the intuitive verification of (3.11), which is detailed in Section 6.2.

Let $S^{(r)}(n)$ be the sojourn time of the nth arriving customer in service, then the traffic intensity $\rho^{(r)}$ can be defined under the stationary setting as

$$\rho^{(r)} = \alpha_e^{(r)} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} S^{(r)}(k). \tag{3.12}$$

Hence, by Little's law, we have $\rho^{(r)} = \mathbb{P}(L^{(r)} > 0) = 1 - \mathbb{P}(L^{(r)} = 0)$. In diffusion applications of a single server queue, the mean queue length is estimated through $(1 - \rho^{(r)})\mathbb{E}[L^{(r)}]$ as $r \downarrow 0$. This limit is computed in the following corollary, which is proved in Appendix A.2.

Corollary 1. *Under the assumptions (A1)–(A6) and either one of the assumptions in part (ii) of Theorem 1,*

$$\begin{aligned}\lim_{r\downarrow 0} (1 - \rho^{(r)})\mathbb{E}[L^{(r)}] &= \lim_{r\downarrow 0} \mathbb{P}(L^{(r)} = 0)\mathbb{E}[L^{(r)}] \\ &= \begin{cases} \dfrac{\sigma_1^2(2b_2\ell_1(b_2\ell_1 + c_2\sigma_1^2) + c_2^2\sigma_1^2\sigma_2^2)}{2(2b_2\ell_1 + c_2\sigma_1^2)^2}, & b_1 = 0, \\ \dfrac{c_1^2 b_2^2\sigma_1^2(e^{\beta_1\ell_1} - 1) + 2b_1b_2\ell_1(c_2b_1 - c_1b_2) + c_2^2b_1^2\sigma_2^2}{2b_1b_2(c_2b_1 + c_1b_2(e^{\beta_1\ell_1} - 1))}, & b_1 \neq 0. \end{cases}\end{aligned} \tag{3.13}$$

Let us consider Theorem 1 when $T_{e,1}^{(r)}$, $T_{e,2}^{(r)}$, and $T_s^{(r)}$ are all exponentially distributed. We call this model the rth two-level $M/M/1$ queue. Note that (A6) is automatically satisfied in this exponential model.

Proposition 1. *Under the assumptions (A1)–(A5), the scaled stationary distribution $\nu_{\exp}^{(r)}$ of the rth two-level $M/M/1$ queue weakly converges to $\widetilde{\nu}_{\exp}$ as $r \downarrow 0$, where $\widetilde{\nu}_{\exp}$ is defined as $\widetilde{\nu}_{\exp}(B) = \widetilde{\nu}_{1|cd}(B)\widetilde{A}_{\exp,1} + \widetilde{\nu}_{2|cd}(B)\widetilde{A}_{\exp,2}$ for $B \in \mathcal{B}(\mathbb{R}_+)$ in which*

$$\widetilde{A}_{\exp,1} = \begin{cases} \dfrac{\beta_2\ell_1}{1 + \beta_2\ell_1}, & b_1 = 0, \\ \dfrac{\beta_2(e^{\beta_1\ell_1} - 1)}{\beta_1 + \beta_2(e^{\beta_1\ell_1} - 1)}, & b_1 \neq 0, \end{cases} \tag{3.14}$$

and $\widetilde{A}_{\exp,2} = 1 - \widetilde{A}_{\exp,1}$.

Remark 3. Since $\sigma_i^2 = 2$ and $\beta_i = b_i/c_i$ for $i = 1, 2$ in the exponential case, $\widetilde{\nu}_{\exp}$ and $\widetilde{A}_{\exp,i}$ agree with $\widetilde{\nu}$ and $\widetilde{A}_{\exp,i}$ in Theorem 1, respectively, for $i = 1, 2$.

Of course, Proposition 1 is the special case of part (ii) of Theorem 1, so no proof is needed. However, for a double check, we give its direct proof in Appendix A.3, which is helpful to

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

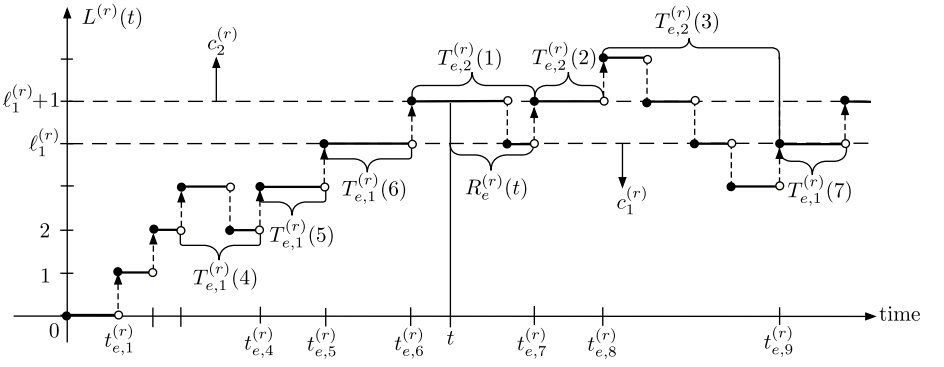


FIGURE 1. Sample path of the queue length process of the rth model.

convince (ii) of Theorem 1. This proof is not obvious because the exponential model requires a Markov chain with a two-dimensional state space to correctly describe the case that an arrival finds the queue length less than ℓ_1 when the previous arrival found the queue length greater than ℓ_1 but multiple departures occurred (see the arrival at time $t_e^{(r)}(8)$ in Figure 1). This Markov chain will be detailed in Appendix A.3.

4. The BAR's for the two-level *GI/G/*1 queue

In this section we prepare auxiliary notations and results for proving Theorem 1. A major difficulty for this proof comes from the fact that $L^{(r)}(\cdot)$ is not a Markov chain, and we have to work with $X^{(r)}(\cdot)$. Instead of directly computing the stationary distribution, we study it through a stationary equation, called a BAR, using Palm distributions, which will be defined below. This method is called a BAR approach and has been used in [3, 4, 10]. Recently, this approach has been further studied in [12].

Throughout this section, we assume that (A1)–(A5) hold unless stated otherwise, but (A6) is not assumed. The proof of Theorem 1 will be done in Section 5.

4.1. Palm distribution and BAR

Recall that we are working on stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with time-shift operator semigroup $\Theta_\bullet$. Furthermore, by the stability assumption (A5), we can assume that this framework is stationary, that is,

$$\mathbb{P}(\Theta_t^{-1}A) = \mathbb{P}(A), \qquad A \in \mathcal{F}.$$

Then, for each $r \in (0, 1]$, $X^{(r)}(\cdot)$ and the increments of $N_e^{(r)}$ and $N_s^{(r)}$ are jointly stationary. Since $\alpha_e^{(r)}$ and $\alpha_s^{(r)}$ are finite and positive by Lemma 1, we can define probability distributions $\mathbb{P}_e^{(r)}$ and $\mathbb{P}_s^{(r)}$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{P}_v^{(r)}(A) = (\alpha_v^{(r)})^{-1}\mathbb{E}\left[\int_{(0,1]} 1_{\Theta_u^{-1}A} N_v^{(r)}(du)\right], \qquad A \in \mathcal{F},\ v = e, s, \tag{4.15}$$

where 1_A is the indicator function of the event $A \in \mathcal{F}$. For $v = e, s$, $\mathbb{P}_v^{(r)}$ is called a Palm distribution concerning the counting process $N_v^{(r)}$.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

The Palm distributions play a major role for the derivation of the stationary equations in our analysis (see, e.g., (4.17)). Intuitively, $\mathbb{P}_v^{(r)}$ is considered as the conditional probability distribution given the event that $N_v^{(r)}$ counts arrivals for $v = e$ (service completion for $v = s$) at time 0. This event occurs with probability 0 under the stationary framework, so such an event is often handled as a boundary condition in the stationary equation (see, e.g., [5]). On the other hand, Palm distributions are defined through time averages under the stationary framework. This enables us to derive the stationary equation without any boundary condition, where the boundary conditions are handled by Palm distributions in the stationary equation. This is quite convenient because the boundary conditions may be only asymptotically satisfied. This is exactly our case.

We now derive a BAR for $X^{(r)}(\cdot)$ using the framework proposed in [12]. For this, we introduce notations for describing the dynamics of $X^{(r)}(\cdot)$. Let S be its state space, that is, $S = \mathbb{Z}_+ \times \mathbb{R}_+^2$, and let $C_b^p(S)$ be the set of all bounded and continuous functions from S to $\mathbb{R}_+$ that are partially differentiable from the right concerning the second and third variables and in which $(\partial^+/\partial x_i)f$ is the ith partial derivative from the right. Define operator $\mathcal{H}$ on $C_b^p(S)$ that maps $f \in C_b^p(S)$ to $\mathcal{H}f \in C_b^p(S)$ by

$$\mathcal{H}f(x) = -\frac{\partial^+}{\partial x_2} f(x) - (c_1 1(1 \le x_1 \le \ell_1) + c_2 1(x_1 > \ell_1)) \frac{\partial^+}{\partial x_3} f(x), \qquad x \in S,$$

where $x = (x_1, x_2, x_3) \in \mathbb{R}_+^2$.

In what follows, we often use a difference operator Δ, which is defined as $\Delta g(t) = g(t) - g(t-)$ for function g from $\mathbb{R}_+$ to $\mathbb{R}$, which is right continuous with the left limit. For $t \ge 0$, define $\mathbb{F}$-adapted process $X_1^{(r)}(\cdot)$ and $L_1^{(r)}(\cdot)$ as

$$X_1^{(r)}(t) = \begin{cases} X^{(r)}(t-), & \Delta N_e^{(r)}(t) = 0, \\ (L^{(r)}(t-) + 1, T_{e,1}^{(r)}(n_{e,1}(t)), R_s^{(r)}(t-)), & \Delta N_e^{(r)}(t) = 1, L^{(r)}(t-) < \ell_1^{(r)}, \\ (L^{(r)}(t-) + 1, T_{e,2}^{(r)}(n_{e,2}(t)), R_s^{(r)}(t-)), & \Delta N_e^{(r)}(t) = 1, L^{(r)}(t-) \ge \ell_1^{(r)}, \end{cases}$$

$$L_1^{(r)}(t) = L^{(r)}(t-) + 1(\Delta N_e^{(r)}(t) = 1), \qquad R_1^{(r)}(t) = (R_e^{(r)}(t), R_s^{(r)}(t-)),$$

where $n_{e,1}(t) = \sum_{n=1}^{N_e^{(r)}(t)} 1(L^{(r)}(t_{e,n}^{(r)}-) < \ell_1^{(r)})$ and $n_{e,2}(t) = \sum_{n=1}^{N_e^{(r)}(t)} 1(L^{(r)}(t_{e,n}^{(r)}-) \ge \ell_1^{(r)})$. We call $X_1^{(r)}(t)$ an intermediate state. Note that

$$X^{(r)}(t) = (L_1^{(r)}(t) - 1, R_e^{(r)}(t), T_s^{(r)}(N_s^{(r)}(t)))$$

when $\Delta N_s^{(r)}(t) = 1$, and $X^{(r)}(t) = X_1^{(r)}(t)$ when $\Delta N_s^{(r)}(t) = 0$.

In these definitions, $X^{(r)}(0-)$ is required, but formally not defined. However, this does not cause any problem; see part (i) of Remark 4 for its details. For convenience, we define the difference operators Δ_e and Δ_s as

$$\Delta_e f(X^{(r)})(t) = f(X_1^{(r)}(t)) - f(X^{(r)}(t-)),$$
$$\Delta_s f(X^{(r)})(t) = f(X^{(r)}(t)) - f(X_1^{(r)}(t)).$$

Furthermore, let $f(X)(t) = f(X(t))$. Then, from the time evolution of $X^{(r)}(t)$, we have

$$f(X^{(r)}(t)) = f(X^{(r)}(0)) + \int_0^t \mathcal{H}f(X^{(r)}(u))\,du + \sum_{u \in (0,t]} [\Delta_e f(X^{(r)})(u) + \Delta_s f(X^{(r)})(u)], \qquad (4.16)$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

where the summation is well defined because $\Delta_e f(X^{(r)})(u)$ and $\Delta_s f(X^{(r)})(u)$ vanish except for finitely many $u \in (0, t]$. Take the expectation of (4.16) for $t = 1$ by $\mathbb{P}$, then the stationarity of $X(\cdot)$ and the definition of the Palm distribution yield

$$\mathbb{E}\left[\int_0^1 \mathcal{H}f(X^{(r)}(u))\,du\right] = \int_0^1 \mathbb{E}\left[\mathcal{H}f(X^{(r)}(u))\right]du = \mathbb{E}\left[\mathcal{H}f(X^{(r)}(0))\right],$$

$$\mathbb{E}\left[\sum_{u\in(0,1]} \Delta_v f(X^{(r)})(u)]\right] = \alpha_v^{(r)}\mathbb{E}_v^{(r)}\left[\Delta_v f(X^{(r)})(0)\right], \quad v = e, s.$$

Hence, we have a stationary equation:

$$\mathbb{E}\left[\mathcal{H}f(X^{(r)}(0))\right] + \alpha_e^{(r)}\mathbb{E}_e^{(r)}\left[\Delta_e f(X^{(r)})(0)\right] + \alpha_s^{(r)}\mathbb{E}_s^{(r)}\left[\Delta_s f(X^{(r)})(0)\right] = 0, \qquad f \in C_b^p. \quad (4.17)$$

Equation (4.17) is a special case of the rate conservation law (see, e.g., [9]) and referred to as a BAR, we then call the analysis based on this BAR a BAR approach. A major difficulty in this BAR approach is to evaluate the expectations under the Palm distributions (the last two terms in (4.17)). Their evaluations will be done for the two-level $GI/G/1$ queue in Section 4.3.

Remark 4. We give the following two remarks on Palm distributions.

(i) $X^{(r)}(0-)$ is not defined under $\mathbb{P}$, but its distribution is well defined under the Palm distributions $\mathbb{P}_e^{(r)}$ and $\mathbb{P}_e^{(r)}$ because the definition (4.15) only uses $X^{(r)}(t)$ for $t > 0$ under $\mathbb{P}$. So, $X^{(r)}(0-)$ is a random variable that can be defined under $\mathbb{P}_e^{(r)}$ and $\mathbb{P}_s^{(r)}$. Hence, $\Delta_e f(X^{(r)})(0)$ and $\Delta_s f(X^{(r)})(0)$ in (4.17) are well defined under the Palm distributions.

(ii) Under the Palm distributions, $X_1^{(r)}(0)$ is an intermediate state between just after $N_e^{(r)}$ is counted at time 0 and just before $N_s^{(r)}$ is counted at time 0. This means that we take a so-called arrival first framework. As for the intermediate state, see [12] for its details. We consider $L^{(r)}(0-)$ under $\mathbb{P}_e^{(r)}$, while $L^{(r)}(0)$ under $\mathbb{P}_s^{(r)}$. This prevents us directly referring to the intermediate state $X_1^{(r)}(0)$ in our arguments. Of course, the intermediate state is not needed if $N_e^{(r)}$ and $N_s^{(r)}$ do not have a simultaneous count, but such a restriction is avoided by introducing $X_1^{(r)}(0)$.

We note basic facts on $\alpha_e^{(r)}$ and $L^{(r)}(\cdot)$ under the Palm distributions.

Lemma 2. *Under conditions (A2) and (A5),*

$$\mathbb{P}_e^{(r)}[(L^{(r)}(0-) = k] = \mathbb{P}_s^{(r)}[L^{(r)}(0) = k], \quad k \geq 0, \quad (4.18)$$

$$\alpha_e^{(r)} = \frac{\lambda_1^{(r)}\lambda_2^{(r)}}{\lambda_2^{(r)}\mathbb{P}_e^{(r)}(L^{(r)}(0-) < \ell_1^{(r)}) + \lambda_1^{(r)}\mathbb{P}_e^{(r)}(L^{(r)}(0-) \geq \ell_1^{(r)})}, \quad (4.19)$$

$$\alpha_e^{(r)} = c_1^{(r)}\mu^{(r)}\mathbb{P}(1 \leq L^{(r)} \leq \ell_1^{(r)}) + c_2^{(r)}\mu^{(r)}\mathbb{P}(L^{(r)} > \ell_1^{(r)}). \quad (4.20)$$

Proof. For each integer $k \geq 0$, let $f(z, x) = z \wedge (k+1)$ for $(z, x) \in S$, then obviously $f \in C_b^p(S)$. Since $\Delta_e f(X^{(r)})(u) = 1(L^{(r)}(u-) \leq k, N_e^{(r)}(u) - N_e^{(r)}(u-) = 1)$ and $\Delta_s f(X^{(r)})(u) =$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$-1(L^{(r)}(u) \le k, N_s^{(r)}(u) - N_s^{(r)}(u-) = 1)$, we have

$$\alpha_e^{(r)}\mathbb{E}_e^{(r)}\left[\Delta_e f(X^{(r)})(0)\right] + \alpha_s^{(r)}\mathbb{E}_s^{(r)}\left[\Delta_s f(X^{(r)})(0)\right] = \alpha_e^{(r)}\mathbb{P}_e^{(r)}\left[L^{(r)}(0-) \le k\right] - \alpha_s^{(r)}\mathbb{P}_s^{(r)}\left[L^{(r)}(0) \le k\right].$$

Applying this to (4.17) yields

$$\alpha_e^{(r)}\mathbb{P}_e^{(r)}\left[L^{(r)}(0-) \le k\right] - \alpha_s^{(r)}\mathbb{P}_s^{(r)}\left[L^{(r)}(0) \le k\right] = 0, \quad k \ge 0.$$

Since $\alpha_e^{(r)} = \alpha_s^{(r)}$ by Lemma 1, the above equation yields (4.18). We next let $f(z, x) = x_1$ and apply (4.17), then

$$-1 + \alpha_e^{(r)}\mathbb{E}_e^{(r)}\left[T_{e,1}^{(r)}1(L^{(r)}(0-) < \ell_1^{(r)}) + T_{e,2}^{(r)}1(L^{(r)}(0-) \ge \ell_1^{(r)})\right] = 0$$

because $(R_e^{(r)})'(t) = -1$. This proves (4.19). Similarly, let $f(z, x) = x_2$, then

$$(R_s^{(r)})'(t) = -c_1^{(r)}1(1 \le L^{(r)}(0) \le \ell_1^{(r)}) - c_2^{(r)}1(L^{(r)}(0) > \ell_1^{(r)}), \tag{4.21}$$

and $\mathbb{E}_s^{(r)}[\Delta_s R_s^{(r)}(0)] = \mathbb{E}[T_s^{(r)}]$. Hence, (4.17) yields (4.20). □

Since it follows from (4.19) and (A2) that

$$0 < \lambda_1^{(r)} \wedge \lambda_2^{(r)} \le \alpha_e^{(r)} \le \lambda_1^{(r)} \vee \lambda_2^{(r)} < \infty, \quad r \in (0, 1], \tag{4.22}$$

we have the following corollary.

Corollary 2. *Under conditions (A2) and (A5), $\alpha_e^{(r)}$ is uniformly bounded away from 0 and uniformly finite for all $r \in (0, 1]$.*

To prove Theorem 1, we fully use the BAR (4.17) replacing $\alpha_s^{(r)}$ by $\alpha_e^{(r)}$ because $\alpha_s^{(r)} = \alpha_e^{(r)}$ by Lemma 1. We first apply an exponential type of a test function to this BAR, and derive various properties on it under the stationary distribution $\mathbb{P}$ and the Palm distributions $\mathbb{P}_e^{(r)}$ and $\mathbb{P}_s^{(r)}$ in Section 4.2, then Theorem 1 is proved in Section 5.

4.2. Asymptotic BAR for the prelimit process $X^{(r)}(\cdot)$

We prove Theorem 1 using the moment generating functions $\varphi_i^{(r)}(\theta_i)$ and $\varphi_{i|cd}^{(r)}(\theta_i)$ of $\nu_i^{(r)}$ and $\nu_{i|cd}^{(r)}$, respectively, for $i = 1, 2$. Namely, for $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$,

$$\varphi_1^{(r)}(\theta_1) = \mathbb{E}[e^{\theta_1 r L^{(r)}}1(L^{(r)} \le \ell_1^{(r)})], \qquad \varphi_2^{(r)}(\theta_2) = \mathbb{E}[e^{\theta_2 r L^{(r)}}1(\ell_1^{(r)} < L^{(r)})],$$
$$\varphi_{1|cd}^{(r)}(\theta_1) = \mathbb{E}[e^{\theta_1 r L^{(r)}}|L^{(r)} \le \ell_1^{(r)}], \qquad \varphi_{2|cd}^{(r)}(\theta_2) = \mathbb{E}[e^{\theta_2 r L^{(r)}}|\ell_1^{(r)} < L^{(r)}].$$

To this end, we introduce test functions for the BAR (4.17). We first define functions $w_i^{(r)}$ and $g_{i,\theta}^{(r)}$ for $i = 1, 2$ and $\theta \in \mathbb{R}$ as

$$w_1^{(r)}(z) = 1(z \le \ell_1^{(r)}), \qquad w_2^{(r)}(z) = 1(z > \ell_1^{(r)}), \qquad z \in \mathbb{Z}_+,$$
$$g_{i,\theta}^{(r)}(x_1, x_2) = \exp\left(-\eta_i^{(r)}(\theta)\left(x_1 \wedge \frac{1}{r}\right) - \zeta^{(r)}(\theta)\left(x_2 \wedge \frac{1}{r}\right)\right), \qquad (x_1, x_2) \in \mathbb{R}_+^2,$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

where $\eta_i^{(r)}(\theta)$, $\zeta^{(r)}(\theta)$ are the unique solutions of the following equations for each $\theta \in \mathbb{R}$:

$$e^{\theta}\mathbb{E}\left[e^{-\eta_i^{(r)}(\theta)(T_{e,i}^{(r)}\wedge 1/r)}\right]=1, \quad i=1,2, \qquad e^{-\theta}\mathbb{E}\left[e^{-\zeta^{(r)}(\theta)(T_s^{(r)}\wedge 1/r)}\right]=1, \tag{4.23}$$

because $\eta_i^{(r)}(\theta)$ is the inverse function of the Laplace transform of the finite positive random variable $T_{e,i}^{(r)} \wedge 1/r$, taking the value $e^{-\theta}$ and $\zeta^{(r)}(\theta)$ is similarly determined. Then, for each $\boldsymbol{\theta} \equiv (\theta_1, \theta_2) \in \mathbb{R}^2$, define the test function $f_{\boldsymbol{\theta}}^{(r)}$ for the BAR as

$$f_{\boldsymbol{\theta}}^{(r)}(z, x_1, x_2) = \sum_{i=1,2} e^{\theta_i z} g_{i,\theta_i}^{(r)}(x_1, x_2) w_i^{(r)}(z), \qquad (z, x_1, x_2) \in S. \tag{4.24}$$

Functions $\eta_i^{(r)}(\theta)$ and $\zeta^{(r)}(\theta)$ play key roles in our analysis. Consider their properties. First, it is easy to see that they are infinitely many differentiable functions of θ. Hence, by twice differentiating the equations in (4.23), the following lemma can be obtained by their Taylor expansions around the origin. It is formally proved in [4, Lemma 5.8].

Lemma 3. *Under the assumptions (A1)–(A4), we have the following expansions as $r \downarrow 0$ for each fixed $\theta \in \mathbb{R}$:*

$$\eta_i^{(r)}(r\theta) = \lambda_i^{(r)} r\theta + \tfrac{1}{2}(\lambda_i^{(r)})^3(\sigma_{e,i}^{(r)})^2 r^2\theta^2 + o((\theta r)^2), \qquad i=1,2, \tag{4.25}$$

$$\zeta^{(r)}(r\theta) = -\mu^{(r)} r\theta + \tfrac{1}{2}(\mu^{(r)})^3(\sigma_s^{(r)})^2 r^2\theta^2 + o((\theta r)^2). \tag{4.26}$$

Furthermore, for $i=1,2$, there are constant $d_{e,i}, d_s, a > 0$ such that, for $r \in (0,1]$ and $\theta \in \mathbb{R}$ satisfying $r|\theta| < a$,

$$\left|\eta_i^{(r)}(r\theta)(u_1 \wedge 1/r)\right| \le |r\theta| d_{e,i}(u_1 \wedge 1/r), \quad i=1,2, \tag{4.27}$$

$$\left|\zeta^{(r)}(r\theta)(u_2 \wedge 1/r)\right| \le |r\theta| d_s(u_2 \wedge 1/r). \tag{4.28}$$

Remark 5. In (4.25) and (4.26), $o((\theta r)^2)$ can be replaced by $o(r^2)$ because θ is fixed, but $o((\theta r)^2)$ is convenient for our computation because we need to divide both sides of them by θ in some cases. We take this convention throughout the paper as long as it is possible.

Recall that (A1)–(A5) are always assumed in Section 4 if particular assumptions among them are specified. By Lemma 3, $g_{i,r\theta_i}^{(r)}(R^{(r)})$ is bounded by $e^{|\theta_i|(d_{e,i}+d_s)}$ for sufficiently small r for each $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$, so we have the following facts.

Corollary 3. *Let $R^{(r)} = (R_e^{(r)}, R_s^{(r)})$, then $g_{i,r\theta_i}^{(r)}(R^{(r)})$ converges to 1 as $r \downarrow 0$, and*

$$\lim_{r\downarrow 0}\mathbb{E}\left[g_{i,r\theta_i}^{(r)}(R^{(r)})\right]=1, \qquad \lim_{r\downarrow 0}\mathbb{E}_u\left[g_{i,r\theta_i}^{(r)}(R^{(r)})\right]=1, \quad u=e,s. \tag{4.29}$$

We next introduce auxiliary functions $\psi_i^{(r)}(\theta_i)$ for $i=1,2$ as

$$\psi_i^{(r)}(\theta_i) = \mathbb{E}\left[e^{r\theta_i L^{(r)}} g_{i,r\theta_i}^{(r)}(R^{(r)}) w_i^{(r)}(L^{(r)})\right], \quad \theta_1 \in \mathbb{R}, \theta_2 \le 0. \tag{4.30}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

Then, by Lemma 3 and Corollary 3, we have

$$\psi_i^{(r)}(\theta_i) < \infty, \qquad \mathbb{E}_u\left[f_{r\boldsymbol{\theta}}^{(r)}(L^{(r)}, R^{(r)})\right] < \infty, \quad u = e, s, \quad \text{uniformly in } r \in (0, 1], \tag{4.31}$$

because $r\theta_1 L^{(r)} 1(L^{(r)} \le \ell_1^{(r)}) \le \theta_1(\ell + o(r))$ for $\theta_1 > 0$. These finiteness guarantees our computations for $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$. We also note the following fact.

Corollary 4. *For $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$,*

$$\lim_{r \downarrow 0} \left|\psi_i^{(r)}(\theta_i) - \varphi_i^{(r)}(\theta_i)\right| = 0, \quad i = 1, 2. \tag{4.32}$$

Proof. By Corollary 3, we have

$$\begin{aligned} |\psi_i^{(r)}(\theta_i) - \varphi_i^{(r)}(\theta_i)| &= |\mathbb{E}[e^{r\theta_i L^{(r)}}(g_{1,\theta_i}(R_e^{(r)}) - 1)1(L^{(r)} \le \ell_1^{(r)})]| \\ &\le e^{|\theta_1| r\ell_1^{(r)}} |\mathbb{E}[g_{i,\theta_i}(R_e^{(r)}) - 1]| \\ &\to 0, \quad r \downarrow 0. \end{aligned}$$

This proves (4.32). □

Based on these facts, we compute $\mathbb{E}\left[\mathcal{H}f(X^{(r)}(0))\right]$ in (4.17) for $f_{r\boldsymbol{\theta}}^{(r)}$. Then,

$$\begin{aligned} \mathbb{E}\left[\mathcal{H}f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)}(0))\right] &= \sum_{i=1,2} \eta_i^{(r)}(r\theta_i)\psi_i^{(r)}(\theta_i) + \sum_{i=1,2} c_i^{(r)}\zeta^{(r)}(r\theta_i)\psi_i^{(r)}(\theta_i) \\ &\quad - c_1^{(r)}\zeta^{(r)}(r\theta_1)\mathbb{E}\left[1(L^{(r)} = 0)g_{1,r\theta_1}^{(r)}(R^{(r)}))\right]. \end{aligned} \tag{4.33}$$

Since $\lambda_i^{(r)} - c_i^{(r)}\mu^{(r)} = -b_i\mu^{(r)}r + o(r)$, it follows from Lemma 3, $\beta_i = 2b_i/(c_i\sigma_i^2)$, and $\lambda_i = c_i\mu$ that, for each fixed $\theta_i \in \mathbb{R}$,

$$\begin{aligned} \eta_i^{(r)}(r\theta_i) + c_i^{(r)}\zeta^{(r)}(r\theta_i) &= -b_i r^2\mu\theta_i + \tfrac{1}{2}c_i(\lambda_i^2\sigma_{e,i}^2 + \mu^2\sigma_s^2)r^2\mu\theta_i^2 + o((\theta_i r)^2) \\ &= -\tfrac{1}{2}c_i\sigma_i^2(\beta_i - \theta_i)\mu\theta_i r^2 + o((\theta_i r)^2), \quad r \downarrow 0. \end{aligned}$$

Hence, from (4.33), we have

$$\begin{aligned} &\mathbb{E}\left[\mathcal{H}f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)}(0))\right] \\ &\quad = \frac{1}{2}\sum_{i=1,2} c_i\sigma_i^2\,(-\beta_i + \theta_i)\,\mu\theta_i r^2\psi_i^{(r)}(\theta_i) \\ &\qquad + c_1^{(r)}\left(1 - \frac{1}{2}\mu^2\sigma_s^2 r\theta_1\right)\mu r\theta_1\mathbb{E}\left[1(L^{(r)} = 0)g_{1,r\theta_1}^{(r)}(R^{(r)})\right] + o((\theta_1 r)^2) + o((\theta_2 r)^2). \end{aligned} \tag{4.34}$$

We next compute the Palm expectation terms in (4.17). In what follows, we use the following abbreviated notations:

$$\widehat{T}_{e,i}^{(r)} \equiv T_{e,i}^{(r)} \wedge 1/r, \quad i = 1, 2, \qquad \widehat{T}_s^{(r)} \equiv T_s^{(r)} \wedge 1/r, \tag{4.35}$$

$$\widehat{R}_e^{(r)} \equiv R_e^{(r)}(0) \wedge 1/r, \qquad \widehat{R}_s^{(r)} \equiv R_s^{(r)}(0) \wedge 1/r, \qquad \widehat{R}_{s-}^{(r)} \equiv R_s^{(r)}(0-) \wedge 1/r. \tag{4.36}$$

Since $g_{i,\theta_i}^{(r)}(R^{(r)}(t))$ changes at $t = 0$ only when $L^{(r)}(0-) = \ell_1^{(r)}$ under $\mathbb{P}_e^{(r)}$ or $L^{(r)}(0) = \ell_1^{(r)}$ under $\mathbb{P}_s^{(r)}$, (4.23) implies that

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$\mathbb{E}_e^{(r)}\left[\Delta_e f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)1(L^{(r)}(0-)\neq \ell_1^{(r)})\right]=0,$$
$$\mathbb{E}_s^{(r)}\left[\Delta_s f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)1(L^{(r)}(0)\neq \ell_1^{(r)})\right]=0.$$

Hence, from the definitions of $X_1^{(r)}(0)$ and $\Delta_e f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)$, we have

$$\begin{aligned}
\mathbb{E}_e^{(r)}[\Delta_e f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)] &= \mathbb{E}_e^{(r)}[(f_{r\boldsymbol{\theta}}^{(r)}(X_1^{(r)}(0))-f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)}(0-)))1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&= \mathbb{E}_e^{(r)}[e^{r\theta_2(\ell_1^{(r)}+1)}g_{2,r\theta_2}^{(r)}(R_1^{(r)}(0))1(L_1^{(r)}(0)=\ell_1^{(r)}+1)\\
&\quad - e^{r\theta_1\ell_1^{(r)}}g_{1,r\theta_1}^{(r)}(R^{(r)}(0-)))1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&= e^{r\theta_2\ell_1^{(r)}}\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad - e^{r\theta_1\ell_1^{(r)}}\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta_1)\widehat{R}_{s-}^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)}))],\\
&= (e^{r\theta_2\ell_1^{(r)}}-e^{r\theta_1\ell_1^{(r)}})\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad + e^{r\theta_1\ell_1^{(r)}}\mathbb{E}_e^{(r)}[(e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}}-e^{-\zeta^{(r)}(r\theta_1)\widehat{R}_{s-}^{(r)}})1(L^{(r)}(0-)=\ell_1^{(r)}))],
\end{aligned}\tag{4.37}$$

where the third equality is obtained by applying (4.23). Similarly, we have

$$\begin{aligned}
\mathbb{E}_s^{(r)}[\Delta_s f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)] &= \mathbb{E}_s^{(r)}[(f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)}(0))-f_{r\boldsymbol{\theta}}^{(r)}(X_1^{(r)}(0)))1(L^{(r)}(0)=\ell_1^{(r)})]\\
&= \mathbb{E}_s^{(r)}[e^{r\theta_1\ell_1^{(r)}}g_{1,r\theta_1}^{(r)}(R^{(r)}(0))1(L^{(r)}(0)=\ell_1^{(r)})\\
&\quad - e^{r\theta_2(\ell_1^{(r)}+1)}g_{2,r\theta_2}^{(r)}(R_1^{(r)}(0)))1(L_1^{(r)}(0-)=\ell_1^{(r)}+1)]\\
&= e^{r\theta_1(\ell_1^{(r)}+1)}\mathbb{E}_s^{(r)}[e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]\\
&\quad - e^{r\theta_2(\ell_1^{(r)}+1)}\mathbb{E}_s^{(r)}[e^{-\eta_2^{(r)}(r\theta_2)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]\\
&= (e^{r\theta_1(\ell_1^{(r)}+1)}-e^{r\theta_2(\ell_1^{(r)}+1)})\mathbb{E}_s^{(r)}[e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]\\
&\quad + e^{r\theta_2(\ell_1^{(r)}+1)}\mathbb{E}_s^{(r)}[(e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}}-e^{-\eta_2^{(r)}(r\theta_2)\widehat{R}_e^{(r)}})1(L^{(r)}(0)=\ell_1^{(r)})].
\end{aligned}\tag{4.38}$$

Define $D^{(r)}(\theta_1,\theta_2)=\alpha_e^{(r)}(\mathbb{E}_e^{(r)}[\Delta_e f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)]+\mathbb{E}_s^{(r)}[\Delta_s f_{r\boldsymbol{\theta}}^{(r)}(X^{(r)})(0)])$. Then, by (4.37) and (4.38),

$$\begin{aligned}
D^{(r)}(\theta_1,\theta_2) &= \alpha_e^{(r)}(e^{r\theta_2\ell_1^{(r)}}-e^{r\theta_1\ell_1^{(r)}})\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}})1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad + \alpha_e^{(r)}e^{r\theta_1\ell_1^{(r)}}\mathbb{E}_e^{(r)}[(e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}}-e^{-\zeta^{(r)}(r\theta_1)\widehat{R}_{s-}^{(r)}})1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad + \alpha_e^{(r)}(e^{r\theta_1(\ell_1^{(r)}+1)}-e^{r\theta_2(\ell_1^{(r)}+1)})\mathbb{E}_s^{(r)}[e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]\\
&\quad + \alpha_e^{(r)}e^{r\theta_2(\ell_1^{(r)}+1)}\mathbb{E}_s^{(r)}[(e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}}-e^{-\eta_2^{(r)}(r\theta_2)\widehat{R}_e^{(r)}})1(L^{(r)}(0)=\ell_1^{(r)})].
\end{aligned}\tag{4.39}$$

Thus, applying (4.34), (4.37), and (4.38) to (4.17), we have a prelimit BAR for the two-level $GI/G/1$ queue.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

Lemma 4. *For each fixed $\theta_1 \in \mathbb{R}$ and $\theta_2 \le 0$, we have, as $r \downarrow 0$,*

$$\begin{aligned}&\frac{1}{2}\sum_{i=1,2} c_i\sigma_i^2\big((\beta_i-\theta_i)\mu\theta_i r^2 + o((\theta_i r)^2)\big)\psi_i^{(r)}(\theta_i)\\ &- c_1\left(\mu r\theta_1 - \frac{\mu^3\sigma_s^2 r^2\theta_1^2}{2} + o((\theta_1 r)^2)\right)\mathbb{E}[1(L^{(r)}=0)g_{1,r\theta_1}^{(r)}(R^{(r)})]\\ &\quad = D^{(r)}(\theta_1,\theta_2).\end{aligned} \tag{4.40}$$

This BAR is called an asymptotic BAR, and a starting point of our proof of Theorem 1.

4.3. Expansion of the asymptotic BAR

We aim to derive the limits of $\psi_1^{(r)}(\theta_1)$ and $\psi_2^{(r)}(\theta_2)$ from the BAR (4.40). For this, we divide it for either $\theta_2 = 0$ or $\theta_1 = 0$ by r^2, and take its limit as $r \downarrow 0$, expanding the exponential functions in (4.40). We perform this computation in the following three steps. Recall that assumptions (A1)–(A5) are assumed but (A6) is not assumed throughout Section 4.

Step 1: In this step we show that the moments of $\widehat{T}_v^{(r)}$ for $v = (e, 1), (e, 2), s$ and $\widehat{R}_{v'}^{(r)}$ for $v' = e, s$ are well behaved as $r \downarrow 0$. This is also examined for $\widehat{R}_{s-}^{(r)}1(L^{(r)} = \ell_1^{(r)})$ and $\widehat{R}_e^{(r)}1(L^{(r)} = \ell_{0-}^{(r)})$ under $\mathbb{P}_e^{(r)}$ and $\mathbb{P}_s^{(r)}$, respectively. These are basis for the expansion of the BAR. We then introduce the set of key quantities, $\Delta_i^{(r)}$ for $i = 1, 2$.

Lemma 5. *For $k = 0, 1, 2$ and $r \in (0, 1]$,*

$$(k+1)\mathbb{E}[(R_e^{(r)})^k] \le \alpha_e^{(r)} \max_{i=1,2} \mathbb{E}[(T_{e,i}^{(r)})^{k+1}], \tag{4.41}$$

$$(k+1)(\min_{i=1,2} c_i^{(r)})\mathbb{E}[(R_s^{(r)})^k] \le \alpha_e^{(r)}\mathbb{E}[(T_s^{(r)})^{k+1}] + (k+1)(\min_{i=1,2} c_i^{(r)})\mathbb{E}[T_s^k], \tag{4.42}$$

and, for $i = 1, 2$, as $r \downarrow 0$,

$$|\mathbb{E}[(\widehat{T}_u^{(r)})^k] - \mathbb{E}[(T_u^{(r)})^k]| = \mathbb{E}[(T_u^{(r)})^k 1(T_u^{(r)} > 1/r)] = o(r^{3-k}), \quad u = (e, i), s, \tag{4.43}$$

$$|\mathbb{E}[(\widehat{R}_v^{(r)})^k] - \mathbb{E}[(R_v^{(r)})^k]| = \mathbb{E}[(R_v^{(r)})^k 1(R_v^{(r)} > 1/r)] = o(r^{2-k}), \quad v = e, s. \tag{4.44}$$

In particular, $\{(R_e^{(r)})^k; r \in (0, 1]\}$ and $\{(R_s^{(r)})^k 1(L^{(r)}(0) \ge 1); r \in (0, 1]\}$ are uniformly bounded for $k = 1, 2$ by (A1).

This lemma is easily proved using the BAR (4.17), but its proof is lengthy, so we defer it to Appendix A.4. We prepare one more lemma and its corollary, which are helpful for the expansion of the BAR.

Lemma 6. *For $k = 1, 2, 3$ and $r \in (0, 1]$,*

$$\begin{aligned}&c_1^{(r)}k\mathbb{E}[(R_s^{(r)})^{k-1}(1 - 1(R_s^{(r)} \ge 1/r))1(1 \le L^{(r)} \le \ell_1^{(r)})]\\ &\quad = \alpha_e^{(r)}(-\mathbb{E}_e^{(r)}[(\widehat{R}_{s-}^{(r)})^k 1(L^{(r)}(0-) = \ell_1^{(r)})] + \mathbb{E}[(\widehat{T}_s^{(r)})^k]\mathbb{P}_s^{(r)}[L^{(r)}(0) \le \ell_1^{(r)}]),\end{aligned} \tag{4.45}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$c_2^{(r)} k\mathbb{E}[(R_s^{(r)})^{k-1}(1-1(R_s^{(r)} \ge 1/r))1(L^{(r)} > \ell_1^{(r)})]$$
$$= \alpha_e^{(r)}(\mathbb{E}_e^{(r)}[(\widehat{R}_{s-}^{(r)})^k 1(L^{(r)}(0-) = \ell_1^{(r)})] + \mathbb{E}[(\widehat{T}_s^{(r)})^k]\mathbb{P}_s^{(r)}[L^{(r)}(0) > \ell_1^{(r)}]), \tag{4.46}$$

$$k\mathbb{E}[(R_e^{(r)})^{k-1}(1-1(R_e^{(r)} \ge 1/r))1(L^{(r)} \le \ell_1^{(r)})]$$
$$= \alpha_e^{(r)}(\mathbb{E}_s[(\widehat{R}_e^{(r)})^k 1(L^{(r)}(0) = \ell_1^{(r)})] + \mathbb{E}[(\widehat{T}_{e,1}^{(r)})^k]\mathbb{P}_e^{(r)}[L^{(r)}(0-) < \ell_1^{(r)}]), \tag{4.47}$$

$$k\mathbb{E}[(R_e^{(r)})^{k-1}(1-1(R_e^{(r)} \ge 1/r))1(L^{(r)} > \ell_1^{(r)})]$$
$$= \alpha_e^{(r)}(-\mathbb{E}_s[(\widehat{R}_e^{(r)})^k 1(L^{(r)}(0) = \ell_1^{(r)})] + \mathbb{E}[(\widehat{T}_{e,2}^{(r)})^k]\mathbb{P}_e^{(r)}[L^{(r)}(0-) \ge \ell_1^{(r)}]). \tag{4.48}$$

Proof. For $k \ge 1$, we first prove (4.45). We apply $f(z, x) = (x_2 \wedge 1/r)^k 1(z \le \ell_1^{(r)})$ to (4.17), then we have (4.45) because

$$\frac{d}{dt} f(X^{(r)}(t)) = k(R_s^{(r)}(t))^{k-1} 1(R_s^{(r)} < 1/r,\ 1 \le L^{(r)}(t) \le \ell_1^{(r)}).$$

Similarly, (4.46), (4.47), and (4.48) are obtained from (4.17) by applying $f(z, x) = (x_2 \wedge 1/r)^k 1(z > \ell_1^{(r)})$, $f(z, x) = (x_1 \wedge 1/r)^k 1(z \le \ell_1^{(r)})$, and $f(z, x) = (x_1 \wedge 1/r)^k 1(z > \ell_1^{(r)})$, respectively. □

The following corollary is immediate from (4.45), (4.48), (A1), and Lemma 5.

Corollary 5. *For each* $k = 1, 2, 3$, $\mathbb{E}_e[(\widehat{R}_{s-}^{(r)})^k 1(L^{(r)}(0-) = \ell_1^{(r)})]$ *and* $\mathbb{E}_s[(\widehat{R}_e^{(r)})^k 1(L^{(r)}(0) = \ell_1^{(r)})]$ *are uniformly bounded for* $r \in (0, 1]$.

Note that the limits of $\psi_1^{(r)}(\theta_1)$ and $\psi_2^{(r)}(\theta_2)$ should be obtained from (4.40). However, we see that it requires the relation between the limits of $\varphi_1^{(r)}(0)$ and $\varphi_2^{(r)}(0)$, which is not easy to get from (4.40). To consider this problem, we introduce the following quantities. Define, for $i = 1, 2$,

$$\Delta_i^{(r)} = \mu^{(r)}\mathbb{E}_e^{(r)}[\widehat{R}_{s-}^{(r)} 1(L^{(r)}(0-) = \ell_1^{(r)})] + \lambda_i^{(r)}\mathbb{E}_s^{(r)}[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})] - \mathbb{P}_e^{(r)}[L^{(r)}(0-) = \ell_1^{(r)}]. \tag{4.49}$$

Then, $\Delta_i^{(r)}$ is well defined and finite by Corollary 5, and we have the following lemma.

Lemma 7. *We have*

$$\alpha_e^{(r)}\Delta_1^{(r)} = -rb_1\mu\mathbb{P}[L^{(r)} \le \ell_1^{(r)}] + c_1^{(r)}\mu^{(r)}\mathbb{P}[L^{(r)} = 0] + o(r), \tag{4.50}$$

$$\alpha_e^{(r)}\Delta_2^{(r)} = r\mu b_2\mathbb{P}[L^{(r)} > \ell_1^{(r)}] + o(r). \tag{4.51}$$

Proof. For $k = 1$, multiply both sides of (4.45) and (4.47) by $\mu^{(r)}$ and $\lambda_1^{(r)}$, respectively, and take their difference, then, noting that $c_1^{(r)}\mu^{(r)} - \lambda_1^{(r)} = r\mu^{(r)}b_1 + o(r)$, we have

$$(r\mu^{(r)}b_1 + o(r))\mathbb{P}[L^{(r)} \le \ell_1^{(r)}] - c_1^{(r)}\mu^{(r)}\mathbb{P}[L^{(r)} = 0]$$
$$= \alpha_e^{(r)}(\mathbb{P}_e^{(r)}[L^{(r)}(0-) = \ell_1^{(r)}] - \mu^{(r)}\mathbb{E}_e^{(r)}[\widehat{R}_{s-}^{(r)} 1(L^{(r)}(0-) = \ell_1^{(r)})]$$
$$- \lambda_1^{(r)}\mathbb{E}_s^{(r)}[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})]) + o(r),$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

because $\mathbb{P}_s^{(r)}[L^{(r)}(0) \le \ell_1^{(r)}] - \mathbb{P}_e^{(r)}[L^{(r)}(0-) < \ell_1^{(r)}] = \mathbb{P}_e^{(r)}[L^{(r)}(0-) = \ell_1^{(r)}]$ by Lemma 2. This proves (4.50). Similarly, (4.51) is obtained from (4.46) and (4.48), □

We use Lemma 7 to determine the limits of $\mathbb{P}[L^{(r)} \le \ell_1^{(r)}]$ and $\mathbb{P}[L^{(r)} > \ell_1^{(r)}]$, equivalently, $\varphi_1^{(r)}(0)$ and $\varphi_2^{(r)}(0)$, as $r \downarrow 0$ under some conditions.

Step 2: In this step we expand $D^{(r)}(\theta, 0)$ and $D^{(r)}(0, \theta)$ concerning r as $r \downarrow 0$. We will see that the $\Delta_i^{(r)}$s describe the expansion terms of $D^{(r)}(\theta, 0)$ of order r. We introduce here the $\mathcal{E}_i^{(r)}$s for describing their expansion terms of order r^2.

We start with the following elementary expansions. Recall Remark 5 for our convention of the order terms.

Lemma 8. *For each fixed θ_1 and θ_2, as $r \downarrow 0$,*

$$e^{r\theta_i \ell_1^{(r)}} = e^{\theta_i \ell_1}(1 + o(\theta_i r)), \quad e^{r\theta_i(\ell_1^{(r)}+1)} = e^{\theta_i \ell_1}(1 + \theta_i r + o(\theta_i r)), \quad i = 1, 2, \tag{4.52}$$

$$e^{r\theta_1 \ell_1^{(r)}} - e^{r\theta_2 \ell_1^{(r)}} = e^{\theta_1 \ell_1} - e^{\theta_2 \ell_1} + e^{(\theta_1 \vee \theta_2)\ell_1}(o(\theta_1 r) + o(\theta_2 r)), \tag{4.53}$$

$$e^{r\theta_1(\ell_1^{(r)}+1)} - e^{r\theta_2(\ell_1^{(r)}+1)} = e^{\theta_1 \ell_1} - e^{\theta_2 \ell_1} + r(e^{\theta_1 \ell_1}\theta_1 - e^{\theta_2 \ell_1}\theta_2) + e^{(\theta_1 \vee \theta_2)\ell_1}(o(\theta_1 r) + o(\theta_2 r)). \tag{4.54}$$

Lemma 9. *Let $S_n(x) = \sum_{k=1}^n x^k$ for $x \in \mathbb{R}$ and integer $n \ge 1$, then, for each fixed θ_1 and θ_2, as $r \downarrow 0$, for $i = 1, 2$,*

$$\begin{aligned} e^{-\eta_i^{(r)}(r\theta_i)\widehat{R}_e^{(r)}} &= 1 - r\theta_i(\lambda_i^{(r)}\widehat{R}_e^{(r)}) - 2^{-1}(r\theta_i)^2[(\lambda_i^{(r)}\sigma_{e,i}^{(r)})^2(\lambda_i^{(r)}\widehat{R}_e^{(r)}) - (\lambda_i^{(r)}\widehat{R}_e^{(r)})^2] \\ &\quad + S_3(\widehat{R}_e^{(r)})o((r\theta)^2), \end{aligned} \tag{4.55}$$

$$\begin{aligned} e^{-\zeta^{(r)}(r\theta_i)\widehat{R}_{s-}^{(r)}} &= 1 + r\theta_i(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - 2^{-1}(r\theta_i)^2[(\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - (\mu^{(r)}\widehat{R}_{s-}^{(r)})^2] \\ &\quad + S_3(\widehat{R}_{s-}^{(r)})o((r\theta)^2), \end{aligned} \tag{4.56}$$

$$\begin{aligned} &e^{-\eta_1^{(r)}(r\theta_1)\widehat{R}_e^{(r)}} - e^{-\eta_2^{(r)}(r\theta_2)\widehat{R}_e^{(r)}} \\ &\quad = r(\theta_2\lambda_2^{(r)} - \theta_1\lambda_1^{(r)})\widehat{R}_e^{(r)} + 2^{-1}r^2[\theta_2^2(\lambda_2^{(r)}\sigma_{e,2}^{(r)})^2\lambda_2^{(r)} - \theta_1^2(\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2\lambda_1^{(r)}]\widehat{R}_e^{(r)} \\ &\quad\quad - 2^{-1}r^2[\theta_2^2(\lambda_2^{(r)})^2 - \theta_1^2(\lambda_1^{(r)})^2](\widehat{R}_e^{(r)})^2 + S_3(\widehat{R}_e^{(r)})o((r\theta)^2), \end{aligned} \tag{4.57}$$

$$\begin{aligned} &e^{-\zeta^{(r)}(r\theta_2)\widehat{R}_{s-}^{(r)}} - e^{-\zeta^{(r)}(r\theta_1)\widehat{R}_{s-}^{(r)}} \\ &\quad = r(\theta_2 - \theta_1)(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - 2^{-1}r^2(\theta_2^2 - \theta_1^2)[(\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - (\mu^{(r)}\widehat{R}_{s-}^{(r)})^2] \\ &\quad\quad + S_3(\widehat{R}_{s-}^{(r)})o((r\theta)^2). \end{aligned} \tag{4.58}$$

These lemmas are proved in Appendix A.5. We first compute $D^{(r)}(\theta, 0)$:

$$\begin{aligned} D^{(r)}(\theta, 0) &= \alpha_e^{(r)}(1 - e^{r\theta\ell_1^{(r)}})\mathbb{P}_e^{(r)}[L^{(r)}(0-) = \ell_1^{(r)}] \\ &\quad + \alpha_e^{(r)}e^{r\theta\ell_1^{(r)}}\mathbb{E}_e^{(r)}[(1 - e^{-\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}})1(L^{(r)}(0-) = \ell_1^{(r)})] \\ &\quad + \alpha_e^{(r)}(e^{r\theta(\ell_1^{(r)}+1)} - 1)\mathbb{E}_s^{(r)}[e^{-\eta_1^{(r)}(r\theta)\widehat{R}_e^{(r)}}1(L^{(r)}(0) = \ell_1^{(r)})] \end{aligned}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$
\begin{aligned}
&+\alpha_e^{(r)}\mathbb{E}_s^{(r)}[(e^{-\eta_1^{(r)}(r\theta)\widehat{R}_e^{(r)}}-1)1(L^{(r)}(0)=\ell_1^{(r)})]\\
&=-\alpha_e^{(r)}e^{r\theta\ell_1^{(r)}}\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad+\alpha_e^{(r)}e^{r\theta\ell_1^{(r)}}e^{r\theta}\mathbb{E}_s^{(r)}[e^{-\eta_1^{(r)}(r\theta)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]. \qquad (4.59)
\end{aligned}
$$

Define $\mathcal{E}_1^{(r)}$ as

$$
\begin{aligned}
\mathcal{E}_1^{(r)}&=2^{-1}\mathbb{E}_e^{(r)}[\{(\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)})-(\mu^{(r)}\widehat{R}_{s-}^{(r)})^2\}1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad-2^{-1}\mathbb{E}_s^{(r)}[\{((\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2+2)(\lambda_1^{(r)}\widehat{R}_e^{(r)})-(\lambda_1^{(r)}\widehat{R}_e^{(r)})^2\}1(L^{(r)}(0)=\ell_1^{(r)})]. \qquad (4.60)
\end{aligned}
$$

Applying (4.56) and

$$
\begin{aligned}
e^{r\theta}e^{-\eta_1^{(r)}(r\theta)\widehat{R}_e^{(r)}}&=(1+r\theta+o(r\theta))\{1-r\theta(\lambda_1^{(r)}\widehat{R}_e^{(r)})\\
&\quad-2^{-1}(r\theta)^2[(\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2(\lambda_1^{(r)}\widehat{R}_e^{(r)})-(\lambda_1^{(r)}\widehat{R}_e^{(r)})^2]+o((r\theta)^2)\}\\
&=1-r\theta(-1+\lambda_1^{(r)}\widehat{R}_e^{(r)})-2^{-1}(r\theta)^2[((\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2+2)(\lambda_1^{(r)}\widehat{R}_e^{(r)})-(\lambda_1^{(r)}\widehat{R}_e^{(r)})^2]\\
&\quad+S_3(\widehat{R}_e^{(r)})o((r\theta)^2)
\end{aligned}
$$

to (4.59) and recalling the definition of $\Delta_1^{(r)}$, we have

$$
D^{(r)}(\theta,0)=\alpha_e^{(r)}e^{\theta\ell_1}(-r\theta\Delta_1^{(r)}+(r\theta)^2\mathcal{E}_1^{(r)})+o((r\theta)^2), \qquad (4.61)
$$

because $\mathbb{E}_s^{(r)}[S_3(\widehat{R}_e^{(r)})1(L^{(r)}(0)=\ell_1^{(r)})]$ is uniformly bounded for $r\in(0,1]$ by Corollary 5.

Similarly, define $\mathcal{E}_2^{(r)}$ as

$$
\begin{aligned}
\mathcal{E}_2^{(r)}&=-2^{-1}\mathbb{E}_e^{(r)}[(\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)})-(\mu^{(r)}\widehat{R}_{s-}^{(r)})^2]1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad+2^{-1}\mathbb{E}_s^{(r)}[\{(\lambda_2^{(r)}(\sigma_{e,2}^{(r)})^2+2)(\lambda_2^{(r)}\widehat{R}_e^{(r)})-(\lambda_2^{(r)}\widehat{R}_e^{(r)})^2\}1(L^{(r)}(0)=\ell_1^{(r)})]. \qquad (4.62)
\end{aligned}
$$

Then, $D^{(r)}(0,\theta)$ is computed from (4.39) as

$$
\begin{aligned}
&D^{(r)}(0,\theta)\\
&=\alpha_e^{(r)}(e^{r\theta\ell_1^{(r)}}-1)\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta)\widehat{R}_s^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad+\alpha_e^{(r)}\mathbb{E}_e^{(r)}[(e^{-\zeta^{(r)}(r\theta)\widehat{R}_s^{(r)}}-1)1(L^{(r)}(0-)=\ell_1^{(r)})]+\alpha_e^{(r)}(1-e^{r\theta(\ell_1^{(r)}+1)})\mathbb{P}_s^{(r)}[L^{(r)}(0)=\ell_1^{(r)}]\\
&\quad+\alpha_e^{(r)}e^{r\theta(\ell_1^{(r)}+1)}\mathbb{E}_s^{(r)}[(1-e^{-\eta_2^{(r)}(r\theta)\widehat{R}_e^{(r)}})1(L^{(r)}(0)=\ell_1^{(r)})]\\
&=\alpha_e^{(r)}e^{r\theta\ell_1^{(r)}}\{\mathbb{E}_e^{(r)}[e^{-\zeta^{(r)}(r\theta)\widehat{R}_s^{(r)}}1(L^{(r)}(0-)=\ell_1^{(r)})]-e^{r\theta}\mathbb{E}_s^{(r)}[e^{-\eta_2^{(r)}(r\theta)\widehat{R}_e^{(r)}}1(L^{(r)}(0)=\ell_1^{(r)})]\}\\
&=\alpha_e^{(r)}e^{\theta\ell_1}\{\mathbb{E}_e^{(r)}[1+r\theta\mu^{(r)}\widehat{R}_{s-}^{(r)}\\
&\quad-2^{-1}(r\theta)^2[(\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)})-(\mu^{(r)}\widehat{R}_{s-}^{(r)})^2]1(L^{(r)}(0-)=\ell_1^{(r)})]\\
&\quad-(1+r\theta+o(r\theta))\mathbb{E}_s^{(r)}[1-r\theta(\lambda_2^{(r)}\widehat{R}_e^{(r)})\\
&\quad-2^{-1}(r\theta)^2[(\lambda_2^{(r)}\sigma_{e,2}^{(r)})^2(\lambda_2^{(r)}\widehat{R}_e^{(r)})-(\lambda_2^{(r)}\widehat{R}_e^{(r)})^2]1(L^{(r)}(0)=\ell_1^{(r)})]\}+o((r\theta)^2)\\
&=\alpha_e^{(r)}e^{\theta\ell_1}(r\theta\Delta_2^{(r)}+(r\theta)^2\mathcal{E}_2^{(r)})+o((r\theta)^2). \qquad (4.63)
\end{aligned}
$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

Step 3: In this step we derive asymptotic expansions for $\psi_i^{(r)}(\theta_i)$ for $i=1,2$. First, note that (4.50) and (4.51) can be written as

$$\alpha_e^{(r)}\Delta_1^{(r)} = -rb_1\mu\varphi_1^{(r)}(0) + c_1^{(r)}\mu^{(r)}\mathbb{P}[L^{(r)}=0] + o(r), \tag{4.64}$$

$$\alpha_e^{(r)}\Delta_2^{(r)} = r\mu b_2\varphi_2^{(r)}(0) + o(r). \tag{4.65}$$

Then, substituting (4.61) into the asymptotic BAR (4.40) with $\theta_2=0$ and applying (4.64), we have

$$\begin{aligned}
&\tfrac{1}{2}rc_1\mu\sigma_1^2\big(\beta_1-\theta_1\big)\psi_1^{(r)}(\theta_1)\\
&\quad= c_1\mu\mathbb{P}[L^{(r)}=0] - e^{\theta_1\ell_1}\alpha_e^{(r)}(\Delta_1^{(r)} - r\theta_1\mathcal{E}_1^{(r)}) + \theta_1 o(r),\\
&\quad= c_1\mu(1-e^{\theta_1\ell_1})\mathbb{P}[L^{(r)}=0] + re^{\theta_1\ell_1}\big(b_1\mu\varphi_1^{(r)}(0) + \theta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}\big) + \theta_1 o(r),
\end{aligned} \tag{4.66}$$

Similarly, substituting (4.63) into (4.40) with $\theta_1=0$ and applying (4.65), we have

$$\begin{aligned}
\tfrac{1}{2}rc_2\mu\sigma_2^2\big(\beta_2-\theta_2\big)\psi_2^{(r)}(\theta_2) &= \alpha_e^{(r)}e^{\theta_2\ell_1}\big(\Delta_2^{(r)} + r\theta_2\mathcal{E}_2^{(r)}\big) + \theta_2 o(r)\\
&= re^{\theta_2\ell_1}\big(\mu b_2\varphi_2^{(r)}(0) + \theta_2\alpha_e^{(r)}\mathcal{E}_2^{(r)}\big) + \theta_2 o(r).
\end{aligned} \tag{4.67}$$

We next compute $\mathbb{P}[L^{(r)}=0]$ for (4.66).

Lemma 10. *We have*

$$\mathbb{P}[L^{(r)}=0] = \begin{cases} \dfrac{r}{2c_1\mu\ell_1}(c_1\mu\sigma_1^2\varphi_1^{(r)}(0) + 2\alpha_e^{(r)}\mathcal{E}_1^{(r)}) + o(r), & b=0,\\ \dfrac{re^{\beta_1\ell_1}}{c_1\mu(e^{\beta_1\ell_1}-1)}\big(b_1\mu\varphi_1^{(r)}(0) + \beta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}\big) + o(r), & b\neq 0. \end{cases} \tag{4.68}$$

Proof. From (4.66), we have

$$\mathbb{P}[L^{(r)}=0] = r\frac{c_1\mu\sigma_1^2\big(\theta_1-\beta_1\big)\psi_1^{(r)}(\theta_1) + 2e^{\theta_1\ell_1}(b_1\mu\varphi_1^{(r)}(0) + \theta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)})}{2c_1\mu(e^{\theta_1\ell_1}-1)} + o(r). \tag{4.69}$$

If $b_1=0$ then $\beta_1=b_1=0$, and therefore, (4.69) yields

$$\begin{aligned}
\mathbb{P}[L^{(r)}=0] &= r\lim_{\theta_1\to 0}\frac{\theta_1(c_1\mu\sigma_1^2\psi_1^{(r)}(\theta_1) + 2e^{\theta_1\ell_1}\alpha_e^{(r)}\mathcal{E}_1^{(r)})}{2c_1\mu(e^{\theta_1\ell_1}-1)} + o(r)\\
&= \frac{r}{2c_1\mu\ell_1}(c_1\mu\sigma_1^2\varphi_1^{(r)}(0) + 2\alpha_e^{(r)}\mathcal{E}_1^{(r)}) + o(r),
\end{aligned}$$

which proves (4.68) for $b_1=0$. If $b_1\neq 0$ then substituting $\theta_1=\beta_1$ into (4.69) yields

$$\mathbb{P}[L^{(r)}=0] = \frac{re^{\beta_1\ell_1}}{c_1\mu(e^{\beta_1\ell_1}-1)}(b_1\mu\varphi_1^{(r)}(0) + \beta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}) + o(r),$$

which is (4.68) for $b_1\neq 0$. □

We also compute $\alpha_e^{(r)}\Delta_1^{(r)}$ to be used later.

Lemma 11. *We have*

$$\alpha_e^{(r)}\Delta_1^{(r)} = \begin{cases} \dfrac{r}{2\ell_1}(c_1\mu\sigma_1^2\varphi_1^{(r)}(0) + 2\alpha_e^{(r)}\mathcal{E}_1^{(r)}) + o(r), & b_1=0,\\ \dfrac{r}{e^{\beta_1\ell_1}-1}\Big(b_1\mu\varphi_1^{(r)}(0) + \dfrac{1}{c_1\mu}\beta_1 e^{\beta_1\ell_1}\alpha_e^{(r)}\mathcal{E}_1^{(r)}\Big) + o(r), & b\neq 0. \end{cases} \tag{4.70}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This lemma is immediate from substituting $\mathbb{P}_e^{(r)}[L^{(r)}=0]$ of Lemma 10 into (4.64). We are now ready to compute $\psi_i^{(r)}(\theta_i)$ for $i=1,2$, which will be a key to prove Theorem 1.

Lemma 12. *We have*

$$\psi_1^{(r)}(\theta_1)=\begin{cases}\dfrac{e^{\theta_1\ell_1}-1}{\ell_1\theta_1}\Big(\varphi_1^{(r)}(0)+\dfrac{2\alpha_e^{(r)}\mathcal{E}_1^{(r)}}{c_1\mu\sigma_1^2}\Big)-\dfrac{2\alpha_e^{(r)}\mathcal{E}_1^{(r)}}{c_1\mu\sigma_1^2}e^{\theta_1\ell_1}+o(1), & b_1=0,\\ \dfrac{\beta_1}{e^{\beta_1\ell_1}-1}\dfrac{e^{\beta_1\ell_1}-e^{\theta_1\ell_1}}{\beta_1-\theta_1}\Big(\varphi_1^{(r)}(0)+\dfrac{2\alpha_e^{(r)}\mathcal{E}_1^{(r)}}{c_1\mu\sigma_1^2}\Big)-\dfrac{2\alpha_e^{(r)}\mathcal{E}_1^{(r)}}{c_1\mu\sigma_1^2}e^{\theta_1\ell_1}+o(1), & b_1\neq 0,\end{cases} \tag{4.71}$$

$$\psi_2^{(r)}(\theta_2)=\frac{\beta_2 e^{\theta_2\ell_1}}{\beta_2-\theta_2}\Big(\varphi_2^{(r)}(0)+\frac{2\alpha_e^{(r)}\mathcal{E}_2^{(r)}}{c_2\mu\sigma_2^2}\Big)-\frac{2\alpha_e^{(r)}\mathcal{E}_2^{(r)}}{c_2\mu\sigma_2^2}e^{\theta_2\ell_1}+o(1). \tag{4.72}$$

Proof. We first compute $\psi_1^{(r)}(\theta_1)$. Assume that $b_1=0$. Substituting $\mathbb{P}_e^{(r)}[L^{(r)}=0]$ of Lemma 10 into (4.66) for $\beta_1=b_1=0$, we have

$$-rc_1\mu\sigma_1^2\theta_1\psi^{(r)}(\theta_1)=(1-e^{\theta_1\ell_1})\frac{r}{\ell_1}(c_1\mu\sigma_1^2\varphi_1^{(r)}(0)+2\alpha_e^{(r)}\mathcal{E}_1^{(r)})+r2e^{\theta_1\ell_1}\theta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}+o(r),$$

which proves (4.71) for $b_1=0$. Next assume that $b_1\neq 0$. Similarly to the case of $b_1=0$, it follows from Lemma 10 and (4.66) for $b_1\neq 0$ that

$$\begin{aligned}&\frac{r}{2}c_1\mu\sigma_1^2(\beta_1-\theta_1)\psi^{(r)}(\theta_1)\\&\quad=\frac{re^{\beta_1\ell_1}(1-e^{\theta_1\ell_1})}{(e^{\beta_1\ell_1}-1)}\big(b_1\mu\varphi_1^{(r)}(0)+\beta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}\big)+re^{\theta_1\ell_1}\big(b_1\mu\varphi^{(r)}(0)+\theta_1\alpha_e^{(r)}\mathcal{E}_1^{(r)}\big)+o(r)\\&\quad=\frac{rb_1\mu\varphi^{(r)}(0)}{e^{\beta_1\ell_1}-1}\big(e^{\beta_1\ell_1}-e^{\theta_1\ell_1}\big)+r\frac{\beta_1e^{\beta_1\ell_1}(1-e^{\theta_1\ell_1})+\theta_1e^{\theta_1\ell_1}(e^{\beta_1\ell_1}-1)}{(e^{\beta_1\ell_1}-1)}\alpha_e^{(r)}\mathcal{E}_1^{(r)}+o(r).\end{aligned}$$

Hence, we have

$$\begin{aligned}\psi_1^{(r)}(\theta_1)&=\frac{\beta_1\varphi^{(r)}(0)}{e^{\beta_1\ell_1}-1}\frac{e^{\beta_1\ell_1}-e^{\theta_1\ell_1}}{\beta_1-\theta_1}\\&\quad+\frac{\beta_1e^{\beta_1\ell_1}-\theta_1e^{\theta_1\ell_1}+(\theta_1-\beta_1)e^{\theta_1\ell_1}e^{\beta_1\ell_1}}{c_1\mu\sigma_1^2(e^{\beta_1\ell_1}-1)(\beta_1-\theta_1)}2\alpha_e^{(r)}\mathcal{E}_1^{(r)}+o(r)\\&=\frac{\beta_1\varphi^{(r)}(0)}{e^{\beta_1\ell_1}-1}\frac{e^{\beta_1\ell_1}-e^{\theta_1\ell_1}}{\beta_1-\theta_1}\\&\quad+\frac{\beta_1(e^{\beta_1\ell_1}-e^{\theta_1\ell_1})+(\beta_1-\theta_1)e^{\theta_1\ell_1}-(\beta_1-\theta_1)e^{\theta_1\ell_1}e^{\beta_1\ell_1}}{c_1\mu\sigma_1^2(e^{\beta_1\ell_1}-1)(\beta_1-\theta_1)}2\alpha_e^{(r)}\mathcal{E}_1^{(r)}+o(r)\\&=\frac{\beta_1}{e^{\beta_1\ell_1}-1}\frac{e^{\beta_1\ell_1}-e^{\theta_1\ell_1}}{\beta_1-\theta_1}\Big(\varphi^{(r)}(0)+\frac{2\alpha_e^{(r)}\mathcal{E}_1^{(r)}}{c_1\mu\sigma_1^2}\Big)-\frac{2e^{\theta_1\ell_1}}{c_1\mu\sigma_1^2}\alpha_e^{(r)}\mathcal{E}_1^{(r)}+o(r).\end{aligned}$$

Thus, we have (4.71) for $b_1\neq 0$. We next compute $\psi_2^{(r)}(\theta_2)$. From (4.67), we have

$$\psi_2^{(r)}(\theta_2)=\frac{e^{\theta_2\ell_1}}{\beta_2-\theta_2}\Big(\beta_2\varphi_2^{(r)}(0)+(\beta_2-(\beta_2-\theta_2))\frac{2\alpha_e^{(r)}\mathcal{E}_2^{(r)}}{c_2\mu\sigma_2^2}\Big)+\theta_2o(r),$$

which proves (4.72). □

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

5. Proof of Theorem 1

From Lemma 12, we can see that, if

$$\lim_{r\downarrow 0} \mathcal{E}_1^{(r)} = \lim_{r\downarrow 0} \mathcal{E}_2^{(r)} = 0, \tag{5.73}$$

then the limits of $\psi_1^{(r)}(\theta_1)$ and $\psi_1^{(r)}(\theta_2)$ can be obtained as long as the limits of $\varphi_1^{(r)}(0)$ and $\varphi_1^{(r)}(0)$ exist. This suggests that (5.73) must be a key condition for Theorem 1. Taking this into account, we separate the proof of Theorem 1 into two steps.

Proposition 2. *If (5.73) holds then Theorem 1 holds without assumption (A6).*

Proposition 3. *Under the assumptions (A1)–(A5), if*

$$\lim_{r\downarrow 0} (\mathcal{E}_1^{(r)} + \mathcal{E}_2^{(r)}) = 0, \tag{5.74}$$

then (5.73) holds. In particular, (A6) implies (5.74) and, therefore, (5.73).

These propositions are proved in the next two subsections. Obviously, they prove Theorem 1. To prove Proposition 3, we use the following lemma, which is proved in Appendix 7.6.

Lemma 13. *Under the assumptions (A1)–(A5), we have*

$$\lim_{r\downarrow 0} \mathbb{E}_s^{(r)}[(\widehat{R}_e^{(r)})^i 1(L^{(r)}(0) = \ell_1^{(r)})] = 0, \qquad i = 1, 2, \tag{5.75}$$

if either $\lambda_2 \neq \lambda_2$ *or*

$$\lim_{r\downarrow 0} \mathbb{P}^{(r)}[L^{(r)} = \ell_1^{(r)}] = 0. \tag{5.76}$$

Lemma 14. *Under the assumptions (A1)–(A5), (5.75) implies (5.74).*

Remark 6. By Lemmas 13 and 14 and Proposition 3, (5.76) is sufficient for (5.73). We conjecture that (5.76) is true because $L^{(r)}(\cdot)$ would weakly converge to a reflected diffusion whose stationary distribution has a density. However, the verification of (5.76) would need a process limit for $L^{(r)}(\cdot)$, which is beyond the present framework.

5.1. Proof of Proposition 2

Throughout this proof, parts (i), (ii), and (iii) of Theorem 1 under the condition (5.73) instead of (A6) are denoted by part (i)', (ii)', and (iii)', respectively.

We first prove part (i)'. Since (3.4) holds, Corollary 4 implies that

$$\lim_{r\downarrow 0} \left| \frac{\psi_i^{(r)}(\theta_i)}{\varphi_i^{(r)}(0)} - \varphi_{i|cd}^{(r)}(\theta_i) \right| = 0. \tag{5.77}$$

By this formula, we can immediately see that Lemma 12 and (5.73) yield

$$\lim_{r\downarrow 0} \varphi_{1|cd}^{(r)}(\theta_1) = \widetilde{\varphi}_{1|cd}(\theta_1) \equiv \begin{cases} \dfrac{e^{\theta_1\ell_1} - 1}{\ell_1\theta_1}, & b_1 = 0, \\ \dfrac{\beta_1}{e^{\beta_1\ell_1} - 1} \dfrac{e^{\beta_1\ell_1} - e^{\theta_1\ell_1}}{\beta_1 - \theta_1}, & b_1 \neq 0, \end{cases} \tag{5.78}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$\lim_{r\downarrow 0}\varphi^{(r)}_{2|cd}(\theta_2)=\widetilde{\varphi}_{2|cd}(\theta_2)\equiv e^{\theta_2\ell_1}\frac{\beta_2}{\beta_2-\theta_2}. \tag{5.79}$$

Since $\widetilde{\varphi}_{i|cd}$ is the moment generating function of $\widetilde{\nu}_{i|cd}$, (5.78) and (5.79) are equivalent to the fact that $\nu^{(r)}_{i|cd}$ weakly converges to $\widetilde{\nu}_{i|cd}$ for $i=1,2$, respectively. Thus, part (i)' is proved.

We next prove part (ii)'. For this, we compute the limits of $\varphi^{(r)}_i(0)$ for $i=1,2$ under the assumptions in part (ii)'. First consider the case that $T^{(r)}_s$ is exponentially distributed for all $r\in(0,t]$. In this case, $R^{(r)}_s(0-)$ has the same distribution as $T^{(r)}_s$ under $\mathbb{P}^{(r)}_e$ by Lemma 2.1 and Theorem 3.1 of [12], so, by Lemma 5 for $k=1$,

$$\mu^{(r)}\mathbb{E}^{(r)}_e[\widehat{R}^{(r)}_{s-}1(L^{(r)}(0-)=\ell^{(r)}_1)]=\mathbb{P}^{(r)}_e[L^{(r)}(0-)=\ell^{(r)}_1]+o(r).$$

Applying this fact to the definitions of $\Delta^{(r)}_i$ yields

$$\begin{aligned}\lambda^{(r)}_2\Delta^{(r)}_1-\lambda^{(r)}_1\Delta^{(r)}_2&=(\lambda^{(r)}_2-\lambda^{(r)}_1)(\mu^{(r)}\mathbb{E}^{(r)}_e[\widehat{R}^{(r)}_{s-}1(L^{(r)}(0-)=\ell^{(r)}_1)]-\mathbb{P}^{(r)}_e[L^{(r)}(0-)=\ell^{(r)}_1])\\&=o(r).\end{aligned} \tag{5.80}$$

Compute $\alpha^{(r)}_e\Delta^{(r)}_1$ by Lemma 11 and (5.73), and recall (4.65), then

$$\alpha^{(r)}_e\Delta^{(r)}_1=\begin{cases}\dfrac{r}{2\ell_1}c_1\mu\sigma_1^2\varphi^{(r)}_1(0)+o(r), & b_1=0,\\ \dfrac{r}{e^{\beta_1\ell_1}-1}b_1\mu\varphi^{(r)}_1(0)+o(r), & b\neq 0,\end{cases} \tag{5.81}$$

$$\alpha^{(r)}_e\Delta^{(r)}_2=r\mu b_2\varphi^{(r)}_2(0)+o(r). \tag{5.82}$$

Then, for $b_1=0$, (5.80) implies that

$$\lambda_2\frac{c_1\mu\sigma_1^2}{2\ell_1}\varphi^{(r)}_1(0)-\lambda_1\mu b_2\varphi^{(r)}_2(0)=o(1).$$

Since $\varphi^{(r)}_1(0)+\varphi^{(r)}_2(0)=1$ and $\mu c_i=\lambda_i$ for $i=1,2$, this can be written as

$$\frac{\sigma_1^2}{2\ell_1}\varphi^{(r)}_1(0)=\frac{b_2}{c_2}(1-\varphi^{(r)}_1(0))+o(1). \tag{5.83}$$

Hence,

$$\lim_{r\downarrow 0}\varphi^{(r)}_i(0)=\widetilde{\varphi}_i(0)\equiv\begin{cases}\dfrac{2b_2\ell_1}{c_2\sigma_1^2+2b_2\ell_1}=\widetilde{A}_1>0, & i=1,\\ \dfrac{c_2\sigma_1^2}{c_2\sigma_1^2+2b_2\ell_1}=\widetilde{A}_2>0, & i=2.\end{cases} \tag{5.84}$$

Similarly, for $b_1\neq 0$, from (5.81) and (5.82), we have

$$\frac{\lambda_2 b_1\mu}{e^{\beta_1\ell_1}-1}\varphi^{(r)}_1(0)-\lambda_1\mu b_2\varphi^{(r)}_2(0)=o(1). \tag{5.85}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

Hence, $\mu c_i = \lambda_i$ for $i = 1, 2$ yields

$$\lim_{r\downarrow 0} \varphi_i^{(r)}(0) = \widetilde{\varphi}_i(0) \equiv \begin{cases} \dfrac{c_1 b_2 (e^{\beta_1 \ell_1} - 1)}{c_2 b_1 + c_1 b_2 (e^{\beta_1 \ell_1} - 1)} = \widetilde{A}_1 > 0, & i = 1, \\ \dfrac{c_2 b_1}{c_2 b_1 + c_1 b_2 (e^{\beta_1 \ell_1} - 1)} = \widetilde{A}_2 > 0, & i = 2. \end{cases} \tag{5.86}$$

We next consider the case that $\lambda_1^{(r)} - \lambda_2^{(r)} = O(r)$. Define $\overline{\varphi}_1(0)$ and $\underline{\varphi}_1(0)$ as

$$\overline{\varphi}_1(0) = \limsup_{r\downarrow 0} \varphi_1^{(r)}(0), \qquad \underline{\varphi}_1(0) = \liminf_{r\downarrow 0} \varphi_1^{(r)}(0),$$

and let $\{\overline{r}_n \in (0, 1]; n \geq 1\}$ and $\{\underline{r}_n \in (0, 1]; n \geq 1\}$ be the sequences such that $\varphi_1^{(\overline{r}_n)}(0)$ and $\varphi_1^{(\underline{r}_n)}(0)$ attain $\overline{\varphi}_1(0)$ and $\underline{\varphi}_1(0)$ as $n \to \infty$, respectively. We aim to show that $\overline{\varphi}_1(0) = \underline{\varphi}_1(0)$ using $\lambda_1^{(r)} - \lambda_2^{(r)} = O(r)$.

To see this, we first consider $\overline{\varphi}_1(0)$. Since $\varphi_1^{(\overline{r}_n)}(0)$ converges to $\overline{\varphi}_1(0)$, applying (5.73) and (4.32) to Lemma 12, we have

$$\lim_{n\to\infty} \varphi_1^{(\overline{r}_n)}(\theta_1) = \lim_{n\to\infty} \psi_1^{(\overline{r}_n)}(\theta_1) = \widetilde{\varphi}_{1|cd}(\theta_1) \lim_{n\to\infty} \varphi_1^{(\overline{r}_n)}(0) = \widetilde{\varphi}_{1|cd}(\theta_1)\overline{\varphi}_1(0). \tag{5.87}$$

Since the distribution $\widetilde{\nu}_1$ has the bounded density $\overline{\varphi}_1(0)\, h_{1|cd}$ on $[0, \ell_1]$, this implies that

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}[L^{(\overline{r}_n)} = \ell_0^{(\overline{r}_n)}] &\leq \lim_{n\to\infty} \mathbb{P}\left[\overline{r}_n L^{(\overline{r}_n)} \in [\ell_1 - \varepsilon, \overline{r}_n \ell_0^{(\overline{r}_n)}]\right] \\ &\leq \lim_{n\to\infty} \overline{\varphi}_1(0) \int_{\ell_1-\varepsilon}^{\ell_1} h_{1|cd}(x)\, dx \\ &\leq \varepsilon \sup_{x\in[\ell_1-\varepsilon,\ell_1]} h_{1|cd}(x) \quad \text{for all } \varepsilon > 0, \end{aligned} \tag{5.88}$$

because $0 \leq \overline{\varphi}_1(0) \leq 1$. Hence, $\mathbb{P}[L^{(\overline{r}_n)} = \ell_0^{(\overline{r}_n)}] = o(1)$ as $n \to \infty$. By Lemma 13, this implies that

$$\lim_{n\to\infty} \mathbb{E}_s^{(\overline{r}_n)}[\widehat{R}_e^{(\overline{r}_n)} 1(L^{(\overline{r}_n)}(0) = \ell_0^{(\overline{r}_n)})] = 0. \tag{5.89}$$

Then, from the definitions of $\Delta_i^{(r)}$ and $\lambda_1^{(\overline{r}_n)} - \lambda_2^{(\overline{r}_n)} = O(\overline{r}_n)$, we have

$$\Delta_1^{(\overline{r}_n)} - \Delta_2^{(\overline{r}_n)} = (\lambda_1^{(\overline{r}_n)} - \lambda_2^{(\overline{r}_n)}) \mathbb{E}_s^{(\overline{r}_n)}[\widehat{R}_e^{(\overline{r}_n)} 1(L^{(\overline{r}_n)}(0) = \ell_0^{(\overline{r}_n)})] = o(\overline{r}_n). \tag{5.90}$$

If $b_1 = 0$ then (4.65) and Lemma 11 yield

$$\alpha_e^{(\overline{r}_n)}(\Delta_1^{(\overline{r}_n)} - \Delta_2^{(\overline{r}_n)}) = \frac{r c_1 \mu \sigma_1^2}{2\ell_1} \varphi_1^{(\overline{r}_n)}(0) - r\mu b_2 \varphi_2^{(\overline{r}_n)}(0) + o(\overline{r}_n).$$

Hence, by (5.90) and $\varphi_1^{(\overline{r}_n)}(0) + \varphi_2^{(\overline{r}_n)}(0) = 1$,

$$\frac{c_1 \mu \sigma_1^2}{2\ell_1} \varphi_1^{(\overline{r}_n)}(0) = \mu b_2 (1 - \varphi_1^{(\overline{r}_n)}(0)) + o(1).$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This is identical to (5.83) for $r=\overline{r}_n$ because $c_1=c_2$ by $\lambda_1=\lambda_2$. Hence, we have (5.84) for $r=\overline{r}_n$ and $b_1=0$. Similarly, if $b_1\neq 0$ then, again by (4.65), (5.82) and (5.90) for $r=\overline{r}_n$, we have

$$\frac{b_1\mu}{e^{\beta_1\ell_1}-1}\varphi_1^{(\overline{r}_n)}(0)-\mu b_2\varphi_2^{(\overline{r}_n)}(0)=o(1).$$

This is identical to (5.85) for $r=\overline{r}_n$ by $\lambda_1=\lambda_2$. Hence, we have (5.86) for $r=\overline{r}_n$ and $b_1\neq 0$. Thus, we have

$$\overline{\varphi}_1(0)=\lim_{n\to\infty}\varphi_1^{(\overline{r}_n)}(0)=\widetilde{\varphi}_1(0)>0.$$

We can repeat the exact same arguments for $\{\underline{r}_n;n\geq 0\}$, and therefore, we also have

$$\underline{\varphi}_1(0)=\lim_{n\to\infty}\varphi_1^{(\underline{r}_n)}(0)=\widetilde{\varphi}_1(0)>0.$$

Hence, $\overline{\varphi}_1(0)=\underline{\varphi}_1(0)$ as claimed, and therefore, $\lim_{r\downarrow 0}\varphi_1^{(r)}(0)=\widetilde{\varphi}_1(0)>0$. Hence, the condition (3.4) in part (i)' is satisfied, so we can use the results in that part. Namely, we have (5.78) and (5.79), equivalently, (3.5). From this together with (5.84) and (5.86), we have (3.9).

It remains to prove (3.10) in part (ii). To see this, apply assumption (A3), (5.84), (5.86), and Lemma 10 to (4.20), then we have

$$\lim_{r\downarrow 0}\alpha_e^{(r)}=\lim_{r\downarrow 0}(c_1^{(r)}\mu^{(r)}(\varphi_1^{(r)}(0)-\mathbb{P}[L^{(r)}=0])+c_2^{(r)}\mu^{(r)}\varphi_2^{(r)}(0))=\lambda_1\widetilde{\varphi}_1(0)+\lambda_2\widetilde{\varphi}_2(0),$$

which proves (3.10).

Finally, we prove part (iii). Assume that $\lambda_1>\lambda_2$, then $\Delta_1^{(r)}>\Delta_2^{(r)}$ by (5.90). For $b=0$, this yields

$$\frac{c_1\mu\sigma_1^2}{2\ell_1}\varphi_1^{(r)}(0)>\mu c_2 b_2(1-\varphi_1^{(r)}(0))+o(1),$$

and therefore, $\varphi_1^{(r)}(0)>\widetilde{A}_1>0$. Similarly, for $b\neq 0$, we also have $\varphi_1^{(r)}(0)>\widetilde{A}_1>0$. Hence, (3.4) holds for $i=1$, so $\nu_{1|cd}^{(r)}$ weakly converges to $\widetilde{\nu}_{1|cd}$ as $r\downarrow 0$ by part (i). Similarly, if $\lambda_1<\lambda_2$ then (3.4) holds for $i=2$, so $\nu_{2|cd}^{(r)}$ weakly converges to $\widetilde{\nu}_{2|cd}$. Thus, the proof of Proposition 2 is completed.

5.2. Proof of Proposition 3

We prove that (5.74) implies (5.73) by *reductio ad absurdum*. First, assume that $\mathcal{E}_1^{(r)}$ does not vanish as $r\downarrow 0$. Then, there are a sequence $\{r_n\in(0,1];n\geq 0\}$ and a constant $\delta_1\neq 0$ such that $r_n\downarrow 0$ and $\mathcal{E}_1^{(r_n)}\to\delta_1$ as $n\to\infty$, because $\mathcal{E}_1^{(r)}$ is uniformly bounded for $r\in(0,1]$ by Corollary 5.

We next choose a subsequence $\{r'_n\in(0,1];n\geq 0\}\subset\{r_n\in(0,1];n\geq 0\}$ such that $\alpha_e^{(r)}$ and $\varphi_1^{(r'_n)}(0)$ converge to constants $a_e>0$ and $d_1\in[0,1]$, respectively, as $n\to\infty$, which is possible because $\alpha_e^{(r)}$ is uniformly bounded away from 0 and above by Corollary 2 and $\varphi_1^{(r'_n)}(0)\in[0,1]$. Then, by Lemma 12, we have

$$\lim_{n\to\infty}\varphi_1^{(r)}(\theta_1)=\begin{cases}\dfrac{e^{\theta_1\ell_1}-1}{\ell_1\theta_1}\left(d_1+\dfrac{2a_e\delta_1}{c_1\mu\sigma_1^2}\right)-\dfrac{2a_e\delta_1}{c_1\mu\sigma_1^2}e^{\theta_1\ell_1}, & b_1=0,\\ \dfrac{\beta_1}{e^{\beta_1\ell_1}-1}\dfrac{e^{\beta_1\ell_1}-e^{\theta_1\ell_1}}{\beta_1-\theta_1}\left(d_1+\dfrac{2a_e\delta_1}{c_1\mu\sigma_1^2}\right)-\dfrac{2a_e\delta_1}{c_1\mu\sigma_1^2}e^{\theta_1\ell_1}, & b_1\neq 0.\end{cases}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This shows that $\nu_1^{(r)}$ weakly converges to a finite measure on $[0, \ell_1]$, but this measure cannot have a negative mass at ℓ_1, so we must have $\delta_1 < 0$ because $\delta_1 \neq 0$.

We next consider the sequence $\{\mathcal{E}_2^{(r'_n)}; n \geq 1\}$. If this sequence does not vanish as $n \to \infty$ then there is a subsequence $\{r''_n \in (0, 1]; n \geq 0\} \subset \{r'_n \in (0, 1]; n \geq 0\}$ such that $\mathcal{E}_2^{(r''_n)}$ converges some $\delta_2 \neq 0$ as $n \to \infty$. Then, by the exact same argument as used for $\{\mathcal{E}_1^{(r_n)}; n \geq 1\}$ using (4.72) instead of (4.71), we must have $\delta_2 < 0$. Hence, we have

$$\lim_{r''_n \downarrow 0} (\mathcal{E}_1^{(r''_n)} + \mathcal{E}_2^{(r''_n)}) = \delta_1 + \delta_2 < 0.$$

This contradicts (5.74), and therefore, $\mathcal{E}_1^{(r)}$ must vanish as $r \downarrow 0$. By (5.74), this further implies that $\mathcal{E}_2^{(r)}$ vanishes as $r \downarrow 0$. Hence, we have (5.73).

We now prove that (A6) implies (5.74). By the definitions of $\mathcal{E}_1^{(r)}$ and $\mathcal{E}_2^{(r)}$, we have

$$\begin{aligned}\mathcal{E}_1^{(r)} + \mathcal{E}_2^{(r)} = {} & -2^{-1}\mathbb{E}_s^{(r)}[\{((\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2 + 2)(\lambda_1^{(r)}\widehat{R}_e^{(r)}) - (\lambda_1^{(r)}\widehat{R}_e^{(r)})^2\}1(L^{(r)}(0) = \ell_1^{(r)})] \\ & + 2^{-1}\mathbb{E}_s^{(r)}[\{(\lambda_2^{(r)}(\sigma_{e,2}^{(r)})^2 + 2)(\lambda_2^{(r)}\widehat{R}_e^{(r)}) - (\lambda_2^{(r)}\widehat{R}_e^{(r)})^2\}1(L^{(r)}(0) = \ell_1^{(r)})]. \quad (5.91)\end{aligned}$$

We separately consider the two cases that $\lambda_1 = \lambda_2$ and $\lambda_1 \neq \lambda_2$. First, consider the case that $\lambda_1 = \lambda_2$. In this case, $\sigma_{e,1}^2 = \sigma_{e,2}^2$ by (A6). Hence, (5.91) implies (5.74). Next consider the case that $\lambda_1 \neq \lambda_2$. In this case, we have (5.75) by Lemma 13. Applying (5.75) to (5.91), we have (5.74) because $\lambda_i^{(r)}$ and $\sigma_{e,i}^{(r)}$ converge to constants λ_i and $\sigma_{e,i}$, respectively, as $r \downarrow 0$ for $i = 1, 2$. This completes the proof of Proposition 3.

6. Concluding remarks

In this section we intuitively discuss two questions: what is a process limit for the two-level $GI/G/1$ queue in heavy traffic and how to verify (3.11) in Remark 2?

6.1. Process limit

Up to now, we have considered the limit of the sequence of the scaled stationary distributions in heavy traffic. From the assumptions (A1)–(A6), one can guess this limiting distribution must be the stationary distribution of the reflected diffusion on $[0, \infty)$ whose Brownian component changes its drift and variance at the level ℓ_1.

Recall that, if one of the conditions in part (ii) of Theorem 1 is satisfied then, for $b_1 \neq 0$, the stationary distribution $\widetilde{\nu}$ is characterized by the stationary equation, that is, BAR, i.e.

$$\begin{aligned}&\sum_{i=1}^{2}\left(-b_i\theta_i + \frac{1}{2}c_i\sigma_i^2\theta_i^2\right)\widetilde{\varphi}_i(\theta_i) + \frac{\beta_1 e^{\beta_1\ell_1}}{e^{\beta_1\ell_1} - 1}\theta_1\frac{c_1\sigma_1^2}{2}\widetilde{\varphi}_1(0) - \frac{\beta_1 e^{\theta_1\ell_1}}{e^{\beta_1\ell_1} - 1}\theta_1\frac{c_1\sigma_1^2}{2}\widetilde{\varphi}_1(0) \\ &+ \beta_2 e^{\theta_2\ell_1}\theta_2\frac{c_2\sigma_2^2}{2}\widetilde{\varphi}_2(0) = 0, \qquad \theta \equiv (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}_-, \quad (6.92)\end{aligned}$$

which can be obtained from (4.40) and (4.39) similarly to $D^{(r)}(\theta, \theta)$ from (A.135) and (A.136), recall that $\beta_i = 2b_i/(c_i\sigma_i^2)$ for $i = 1, 2$. The BAR for $b_1 = 0$ is obtained from (6.92) by letting $b_1 \downarrow 0$. So, we only consider the case that $b_1 \neq 0$.

We like to identify this stationary distribution $\widetilde{\nu}$ by that of a reflected diffusion process. Assume that the diffusion scaled process $L_{\mathrm{dif}}^{(r)}(\cdot) \equiv \{rL^{(r)}(r^{-2}t); t \geq 0\}$ weakly converges to a

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

nonnegative valued stochastic process $Z(\cdot) \equiv \{Z(t); t \geq 0\}$ as $r \downarrow 0$. We aim to characterize this $Z(\cdot)$ by a stochastic integral equation. To this end, define real-valued functions $\sigma(x)$ and $b(x)$ for $x \in \mathbb{R}_+$ as

$$b(x) = b_1 \mu 1(x \leq \ell_1) + b_2 \mu 1(x > \ell_1),$$
$$\sigma(x) = \sqrt{c_1}\mu\sigma_1 1(x \leq \ell_1) + \sqrt{c_2}\mu\sigma_2 1(x > \ell_1),$$

then one may expect that $Z(\cdot)$ satisfies the following stochastic integral equation:

$$Z(t) = Z(0) + \int_0^t \sigma(Z(u))\, dW(u) - \int_0^t b(Z(u))\, du + Y(t), \quad t \geq 0. \tag{6.93}$$

Here $W(\cdot)$ is the standard Brownian motion, $Y(\cdot) \equiv \{Y(t); t \geq 0\}$ is a nondecreasing process satisfying $\int_0^t 1(Z(u) > 0) Y(du) = 0$ for $t \geq 0$. Recently, Miyazawa [11] (see also [1, 2, 13]) showed that (6.93) has a unique weak solution and $Z(t)$ has a stationary distribution if only if $b_2 > 0$, which corresponds to $\rho_2^{(r)} < 1$.

Assume that $Z(\cdot)$ is stationary, then, applying the Ito formula to (6.93) for test function $f(x) = e^{\theta x}$, we have

$$\sum_{i=1}^{2} \Big(- b_i\theta + \frac{1}{2} c_i \sigma_i^2 \theta^2 \Big) \widehat{\varphi}_i(\theta) + \frac{\beta_1 e^{\beta_1 \ell_1}}{e^{\beta_1 \ell_1} - 1} \theta \frac{c_1 \sigma_1^2}{2} \widehat{\varphi}_1(0) = 0, \quad \theta \leq 0, \tag{6.94}$$

where $\widehat{\varphi}_1(\theta) = \mathbb{E}[e^{\theta Z} 1(Z \leq \ell_1)]$ and $\widehat{\varphi}_2(\theta) = \mathbb{E}[e^{\theta Z} 1(Z > \ell_1)]$ for random variable Z subject to the stationary distribution of the $Z(\cdot)$. Equation (6.94) is the stationary equation for $Z(\cdot)$. Then, we can see that (6.92) for $\theta_1 = \theta_2 = \theta$ agrees with (6.94) if and only if $c_1 = c_2$, equivalently, $\lambda_1 = \lambda_2$. Hence, if $L_{\rm dif}^{(r)}(\cdot)$ weakly converges to $Z(\cdot)$ as $r \downarrow 0$ then we must have $\lambda_1 = \lambda_2$. Of course, this does not exclude the possibility that $L_{\rm dif}^{(r)}(\cdot)$ weakly converges to some process other than $Z(\cdot)$. We conjecture that this is indeed possible because the stationary distribution of $L_{\rm dif}^{(r)}(\cdot)$ weakly converges to the $\widetilde{\nu}$ of (3.7) for $\lambda_1 \neq \lambda_2$ at least when T_s is exponentially distributed by part (ii) of Theorem 1. Thus, we have two challenging problems.

(a) Prove that $L_{\rm dif}^{(r)}(\cdot)$ weakly converges to $Z(\cdot)$ when $\lambda_1 = \lambda_2$.

(b) Find a process limit of $L_{\rm dif}^{(r)}(\cdot)$ when $\lambda_1 \neq \lambda_2$.

Both problems require theoretical investigations based on process limits, which is beyond the present framework. So, we leave them for future study.

6.2. Intuitive verification of (3.11)

We intuitively show (3.11). First assume that $b_1 \neq 0$. Under the stationary setting, let $\tau_1^{(r)}$ be the first time for $L^{(r)}(t)$ to get into $\ell_1^{(r)}$ from below or above. Suppose that $L^{(r)}(t)$ first moves to $\ell_1^{(r)} + 1$ after time $\tau_1^{(r)}$, and returns to $\ell_1^{(r)}$ at time $\tau_2^{(r)}$ after time $B_2^{(r)}$. Since the relative frequency of moving up at level $\ell_1^{(r)}$ is guessed to be $f_{\rm up}^{(r)} \equiv \lambda_2^{(r)}/(\lambda_2^{(r)} + c_1^{(r)}\mu^{(r)})$ and $B_2^{(r)}$ is stochastically almost identical with the busy period of the $GI/G/1$ queue with interarrival time $T_{e,2}^{(r)}$ and service time $T_s^{(r)}/c_2^{(r)}$, the mean sojourn time for $L^{(r)}(t)$ to be above $\ell_1^{(r)}$ in the cycle $(\tau_1^{(r)}, \tau_2^{(r)}]$ is

$$f_{\rm up}^{(r)} \mathbb{E}[B_2^{(r)}] \sim \frac{\lambda_2^{(r)}}{\lambda_2^{(r)} + c_1^{(r)}\mu^{(r)}} \times \frac{1}{c_2^{(r)}\mu^{(r)}(1 - \rho_2^{(r)})} = \frac{c_2}{c_1 + c_2} \times \frac{1}{r b_2 \mu} + o(1), \tag{6.95}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

where '$\sim$' stands for the same order as $r \downarrow 0$, that is, for nonzero valued functions f_1 and f_2, $f_1(r) \sim f_2(r)$ means that $\lim_{r\downarrow 0} f_1(r)/f_2(r) = 1$.

On the other hand, when $L^{(r)}(t)$ first moves to $\ell_1^{(r)} - 1$ after time $\tau_1^{(r)}$, let $B_1^{(r)}$ be the time to return to $\ell_1^{(r)}$ from below. Let $f_{\rm down}^{(r)} = 1 - f_{\rm up}^{(r)}$, then, similarly to $B_2^{(r)}$ but, taking $b_1 \neq 0$ and the reflecting boundary at level 0 into account, it follows from (4.25) of [12] that the mean sojourn time for $L^{(r)}(t)$ to be below $\ell_1^{(r)}$ in the cycle $(\tau_1^{(r)}, \tau_2^{(r)}]$ is

$$\begin{aligned} f_{\rm down}^{(r)}\mathbb{E}[B_1^{(r)}] &\sim \frac{c_1^{(r)}\mu^{(r)}}{\lambda_2^{(r)} + c_1^{(r)}\mu^{(r)}} \times \frac{1}{\lambda_1^{(r)}\mathbb{P}_s^{(r)}[L^{(r)}(0) = \ell_1^{(r)}]} \\ &= \frac{c_1}{c_1 + c_2} \times \frac{e^{\beta_1 \ell_1} - 1}{r b_1 \mu} + o(1). \end{aligned} \tag{6.96}$$

Since $f_{\rm up}^{(r)} B_2^{(r)} + f_{\rm down}^{(r)} B_1^{(r)}$ is the mean cycle time to return to $\ell_1^{(r)}$, it follows from the cycle formula of a regenerative process that

$$\lim_{r\downarrow 0} \nu_1^{(r)}(\mathbb{R}_+) = \lim_{r\downarrow 0} \frac{f_{\rm down}^{(r)}\mathbb{E}[B_1^{(r)}]}{f_{\rm down}^{(r)}\mathbb{E}[B_1^{(r)}] + f_{\rm up}^{(r)}\mathbb{E}[B_2^{(r)}]} = \frac{c_1 b_2 (e^{\beta_1 \ell_1} - 1)}{c_2 b_1 + c_1 b_2 (e^{\beta_1 \ell_1} - 1)} = \widetilde{A}_1.$$

Thus, for $b_1 \neq 0$, (3.11) is intuitively obtained for $i = 1$, and it is similarly obtained for $i = 2$. The case for $b_1 = 0$ is similarly derived. Thus, we have support of (3.11).

Appendix A.

A.1. Proof of Lemma 1

Let $N_{e,\wedge}^{(r)}(\cdot)$ and $N_{e,\vee}^{(r)}(\cdot)$ be the renewal processes whose inter-counting times are $T_{e,1}(n) \wedge T_{e,2}(n)$ and $T_{e,1}(n) \vee T_{e,2}(n)$, respectively, for $n \geq 1$. Obviously, $0 < 1/\mathbb{E}[T_{e,1} \vee T_{e,2}] = \mathbb{E}[N_{e,\vee}^{(r)}(t)] \leq \mathbb{E}[N_e^{(r)}(t)] \leq \mathbb{E}[N_{e,\wedge}^{(r)}(t)] = 1/\mathbb{E}[T_{e,1} \wedge T_{e,2}] < \infty$ under the stationary framework. Hence, $\alpha_e^{(r)} = \mathbb{E}[N_e^{(r)}(1)]$ is finite and positive. To prove $\alpha_s^{(r)} = \alpha_e^{(r)}$, consider the stationary process $X_a^{(r)}(t) \equiv L^{(r)}(t) \wedge a$ for each constant $a > 0$. Then, $X_a^{(r)}(t)$ is piecewise constant, has the increment $X_a^{(r)}(u) - X_a^{(r)}(u-) = 1(L^{(r)}(u-) \leq a - 1)$ at the counting instant u of $N_e^{(r)}$, and has the increment $X_a^{(r)}(u) - X_a^{(r)}(u-) = -1(L^{(r)}(u) \leq a)$ at the counting instant u of $N_s^{(r)}$. Hence, from an elementary calculus similar to (4.16), we have

$$X_a^{(r)}(1) - X_a^{(r)}(0) = \int_0^1 1(L^{(r)}(u-) \leq a - 1) N_e^{(r)}(du) - \int_0^1 1(L^{(r)}(u) \leq a) N_s^{(r)}(du).$$

Since $X_a^{(r)}(t)$ is stationary by (A5) and bounded by a, this equation yields

$$\mathbb{E}\Big[\int_0^1 1(L^{(r)}(u-) \leq a - 1) N_e^{(r)}(du)\Big] = \mathbb{E}\Big[\int_0^1 1(L^{(r)}(u) \leq a) N_s^{(r)}(du)\Big].$$

Letting $a \to \infty$, the left-hand side of this equation converges to $\mathbb{E}[N_e^{(r)}(1)] = \alpha_e^{(r)}$, while it right-hand side converges to $\mathbb{E}[N_s^{(r)}(1)] = \alpha_s^{(r)}$. Hence, we have $\alpha_s^{(r)} = \alpha_e^{(r)}$.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

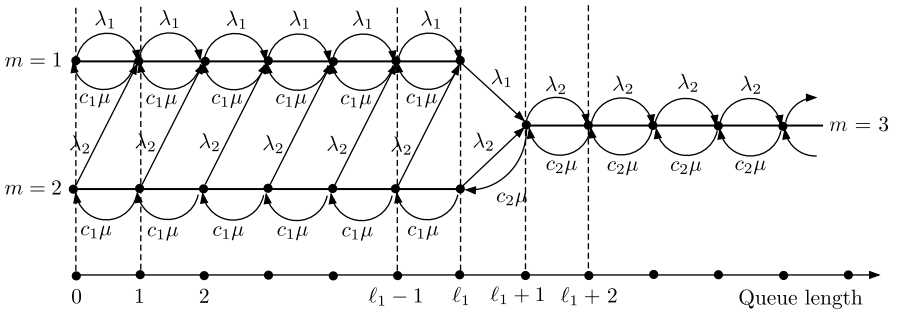


FIGURE 2. Transition diagram.

A.2. Proofs of Corollary 1

We only verify (3.13) for $b_1 \neq 0$ because it is similarly proved for $b_1 = 0$. Since $\mathcal{E}_1^{(r)} = o(1)$ by Proposition 3, it follows from (4.68) and (3.8) that

$$\begin{aligned}\frac{1}{r}\mathbb{P}[L^{(r)}=0] &= \frac{b_1 e^{\beta_1 \ell_1}}{c_1(e^{\beta_1\ell_1}-1)}\frac{c_1 b_2(e^{\beta_1\ell_1}-1)}{c_2 b_1 + c_1 b_2(e^{\beta_1\ell_1}-1)} + o(1)\\ &= \frac{b_1 b_2 e^{\beta_1\ell_1}}{c_2 b_1 + c_1 b_2(e^{\beta_1\ell_1}-1)} + o(1).\end{aligned}$$

On the other hand, from (3.6),

$$\begin{aligned}&\lim_{r\downarrow 0}\mathbb{E}[rL^{(r)}]\\ &= \frac{c_1\sigma_1^2(e^{\beta_1\ell_1}-1-\beta_1\ell_1)}{2b_1(e^{\beta_1\ell_1}-1)}\frac{c_1 b_2(e^{\beta_1\ell_1}-1)}{c_2 b_1 + c_1 b_2(e^{\beta_1\ell_1}-1)} + \left(\ell_1 + \frac{c_2\sigma_2^2}{2b_2}\right)\frac{c_2 b_1}{c_2 b_1 + c_1 b_2(e^{\beta_1\ell_1}-1)}\\ &= \frac{c_1^2 b_2^2\sigma_1^2(e^{\beta_1\ell_1}-1) + 2b_1 b_2\ell_1(c_2 b_1 - c_1 b_2) + c_2^2 b_1^2\sigma_2^2}{2b_1 b_2(c_2 b_1 + c_1 b_2(e^{\beta_1\ell_1}-1))}.\end{aligned}$$

Hence, we have (3.13).

A.3. Proof of Proposition 1

We prove Proposition 1 only for $b_1 \neq 0$ because the case for $b_1 = 0$ can be similarly proved. We first consider this two-level $M/M/1$ queue without index r for notational convenience. Because of the exponential assumption, this model can be described by a continuous-time Markov chain. However, $L(\cdot)$ itself cannot be a Markov chain because the arrival rate may depend on the queue length at the last arrival instant as discussed at the end of Section 3.2. Taking this into account, we describe this model by the continuous-time Markov chain with state space

$$S = \{(j,\ell); j = 1, 2, \ell = 0, 1, \ldots, \ell_1\} \cup \{(3,\ell); \ell \geq \ell_1 + 1\},$$

and its transition diagram is depicted in Figure 2. Denote this Markov chain by $X_{\exp}(\cdot) \equiv \{(J(t), L(t)); t \geq 0\}$.

We partition the state space S of $X_{\exp}(\cdot)$ into the following three disjoint subsets:

$$S_j = \{(j,\ell) \in S; \ell \in [0, \ell_1]\}, \quad j = 1, 2, \qquad S_3 = \{(3,\ell): \ell \geq \ell_1 + 1\}.$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

By the assumptions of Proposition 1, $\mathrm{n}X_M(\cdot)$ has a stationary distribution. Let (J, L) be a random vector subject to this stationary distribution, and define $p_{j,\ell}$ as

$$p_{j,\ell} = \mathbb{P}[J = j, L = \ell], \quad j = 1, 2, 3, (j, \ell) \in S_j.$$

Obviously, the stationary distribution, $\{p_{j,\ell}; (j, \ell) \in S\}$, is uniquely determined by the following stationary equations. For states in S_1,

$$\lambda_1 p_{1,0} = c_1 \mu p_{1,1}, \tag{A.97}$$

$$(\lambda_1 + c_1\mu)p_{1,\ell} = \lambda_1 p_{1,\ell-1} + c_1\mu p_{1,\ell+1} + \lambda_2 p_{2,\ell-1}, \quad \ell \in [1, \ell_1 - 1], \tag{A.98}$$

$$(\lambda_1 + c_1\mu)p_{1,\ell_1} = \lambda_1 p_{1,\ell_1-1} + \lambda_2 p_{2,\ell_1-1}, \tag{A.99}$$

and, for states in S_2,

$$\lambda_2 p_{2,0} = c_1 \mu p_{2,1}, \tag{A.100}$$

$$(\lambda_2 + c_1\mu)p_{2,i} = c_1\mu p_{2,\ell+1}, \qquad \ell \in [1, \ell_1 - 1], \tag{A.101}$$

$$(\lambda_2 + c_1\mu)p_{2,\ell_1} = c_2\mu p_{3,\ell_1+1}, \tag{A.102}$$

For states in S_3, it follows from the equation for $(3, \ell_1 + 1)$ and the local balance that

$$(\lambda_2 + c_2\mu)p_{3,\ell_1+1} = \lambda_1 p_{1,\ell_1} + \lambda_2 p_{2,\ell_1} + c_2\mu p_{\ell_1+2}, \tag{A.103}$$

$$\lambda_2 p_{3,\ell} = c_2\mu p_{3,\ell+1}, \quad \ell \geq \ell_1 + 1. \tag{A.104}$$

In what follows, we use the following notations:

$$\rho_{j,k} = \frac{\lambda_j}{c_k\mu}, \qquad \gamma_{j,k} = \frac{c_j\mu}{c_1\mu + \lambda_k}, \quad j, k = 1, 2.$$

We first note that, from (A.104),

$$p_{3,\ell} = \rho_{22}^{\ell-(\ell_1+1)} p_{3,\ell_1+1}, \quad \ell \geq \ell_1 + 1. \tag{A.105}$$

Thus, $p_{3,\ell}$ for $\ell \geq \ell_1 + 1$ is simply determined by p_{3,ℓ_1+1}.

We next consider $p_{2,\ell}$ for $\ell \in [0, \ell_1 - 1]$. From (A.100) and (A.101), we have

$$p_{2,0} = \frac{c_1\mu}{\lambda_2} p_{2,1} = \rho_{21}^{-1} p_{2,1}, \tag{A.106}$$

$$p_{2,\ell} = \gamma_{12} p_{2,\ell+1} = \gamma_{12}^{\ell_1-\ell} p_{2,\ell_1}, \quad \ell \in [1, \ell_1 - 1]. \tag{A.107}$$

Thus, $p_{2,\ell}$ for $\ell \in [0, \ell_1 - 1]$ is determined by p_{2,ℓ_1}. Furthermore, (A.107) can be written by (A.100) as

$$p_{2,\ell} = \gamma_{12}^{-1} p_{2,\ell-1} = \gamma_{12}^{-\ell+1} p_{2,1} = \gamma_{12}^{-\ell+1} \rho_{21} p_{2,0}, \quad \ell \in [1, \ell_1 - 1]. \tag{A.108}$$

Let us consider relations among the stationary probabilities, $p_{1,0}$, p_{1,ℓ_1}, p_{2,ℓ_1}, and p_{3,ℓ_1+1} at the boundary states of S_j for $j = 1, 2, 3$. We call them boundary probabilities. There is one more boundary probability, $p_{2,0}$, but this probability is ignorable compared with p_{2,ℓ_1} for large ℓ_1 as $\rho_{21} p_{2,0} = \gamma_{12}^{\ell_1-1} p_{2,\ell_1-1} = \gamma_{12}^{\ell_1} p_{2,\ell_1}$ by (A.108) and (A.107), so we do not consider it. We first note that, from (A.102),

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$p_{2,\ell_1} = \gamma_{22} p_{3,\ell_1+1}. \tag{A.109}$$

Since $c_2\mu p_{\ell_1+2} = \lambda_2 p_{3,\ell_1+1}$ by (A.109) for $\ell = \ell_1 + 1$, (A.103) can be written as

$$p_{3,\ell_1+1} = \rho_{12} p_{1,\ell_1} + \rho_{22} p_{2,\ell_1}, \tag{A.110}$$

and, substituting $p_{3,\ell_1+1} = \gamma_{22}^{-1} p_{2,\ell_1}$ of (A.109) to this equation, we have

$$(\gamma_{22}^{-1} - \rho_{22}) p_{2,\ell_1} = \rho_{12} p_{1,\ell_1}.$$

Since $\gamma_{22}^{-1} - \rho_{22} = ((c_1\mu + \lambda_2)/c_2\mu) - \lambda_2/c_2\mu = c_1/c_2 = \rho_{12}\rho_{11}^{-1}$, this equation yields

$$p_{2,\ell_1} = \rho_{11} p_{1,\ell_1}. \tag{A.111}$$

Thus, the boundary probabilities p_{1,ℓ_1}, p_{2,ℓ_1}, and p_{3,ℓ_1+1} are explicitly related by (A.109) and (A.111). We would also like to see such a relation between $p_{1,0}$ and p_{1,ℓ_1}, but this is not easy.

To resolve this problem, we consider $p_{1,\ell}$ for $\ell \in [0, \ell_1 - 1]$. From (A.98),

$$\lambda_1 p_{1,\ell} - c_1\mu p_{1,\ell+1} = \lambda_1 p_{1,\ell-1} - c_1\mu p_{1,\ell} + \lambda_2 p_{2,\ell-1}, \quad \ell \in [1, \ell_1 - 1],$$

then this equation can be written as

$$\rho_{11} p_{1,\ell} - p_{1,\ell+1} = \rho_{11} p_{1,\ell-1} - p_{1,\ell} + \rho_{21} p_{2,\ell-1}, \quad \ell \in [1, \ell_1 - 1].$$

Summing up both sides of this equation for $\ell \in [1, m]$ for $m \in [1, \ell_1 - 1]$, we have

$$\rho_{11} p_{1,m} - p_{1,m+1} = \rho_{21} \sum_{\ell=1}^{m} p_{2,\ell-1}, \quad m \in [0, \ell_1 - 1], \tag{A.112}$$

because $\rho_{11} p_{1,0} - p_{1,1} = 0$ by (A.97). This equation is valid also for $m = 0$ by the convention that the empty sum vanishes, which we always take.

Hence, applying (A.106) and (A.107) to (A.112), we have

$$\begin{aligned}
\rho_{11} p_{1,m} - p_{1,m+1} &= \rho_{21} \left(\rho_{21}^{-1} p_{2,1} + \sum_{\ell=1}^{m-1} \gamma_{12}^{\ell_1-\ell} p_{2,\ell_1} \right) \\
&= \left(1 + \rho_{21} \sum_{\ell=1}^{m-1} \gamma_{12}^{-(\ell-1)} \right) \gamma_{12}^{\ell_1-1} p_{2,\ell_1} \\
&= \left(1 + \rho_{21} \frac{(\gamma_{12}^{-(m-1)} - 1)\gamma_{12}}{1 - \gamma_{12}} \right) \gamma_{12}^{\ell_1-1} p_{2,\ell_1}, \qquad m \in [1, \ell_1 - 1].
\end{aligned}$$

Multiplying both sides by $\rho_{11}^{-(m+1)}$, we obtain

$$\begin{aligned}
&\rho_{11}^{-m} p_{1,m} - \rho_{11}^{-(m+1)} p_{1,m+1} \\
&\quad = \rho_{11}^{-(m+1)} \left(1 + \rho_{21} \frac{(\gamma_{12}^{-(m-1)} - 1)\gamma_{12}}{1 - \gamma_{12}} 1(m \geq 1) \right) \gamma_{12}^{\ell_1-1} p_{2,\ell_1}, \quad m \in [0, \ell_1 - 1].
\end{aligned}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

Summing up both sides of this equation from $m=0$ to $m=k-1\le \ell_1-1$, we have

$$\begin{aligned} p_{1,0}-\rho_{11}^{-k}p_{1,k} &= \sum_{m=0}^{k-1}\rho_{11}^{-(m+1)}\gamma_{12}^{\ell_1-1}p_{2,\ell_1}+\sum_{m=1}^{k-1}\frac{(\rho_{11}\gamma_{12})^{-(m-1)}-\rho_{11}^{-(m-1)}}{1-\gamma_{12}}\rho_{11}^{-2}\rho_{21}\gamma_{12}^{\ell_1}p_{2,\ell_1}\\ &= \frac{\rho_{11}^{-k}-1}{1-\rho_{11}}\gamma_{12}^{\ell_1-1}p_{2,\ell_1}+\left(\frac{(\rho_{11}\gamma_{12})^{-(k-1)}-1}{(\rho_{11}\gamma_{12})^{-1}-1}-\frac{\rho_{11}^{-(k-1)}-1}{\rho_{11}^{-1}-1}\right)\frac{\rho_{11}^{-2}\rho_{21}\gamma_{12}^{\ell_1}p_{2,\ell_1}}{1-\gamma_{12}}\\ &= \frac{\rho_{11}^{-k}-1}{1-\rho_{11}}\gamma_{12}^{\ell_1-1}p_{2,\ell_1}+\left(\frac{(\rho_{11}\gamma_{12})^{-(k-1)}-1}{(\rho_{11}\gamma_{12})^{-1}-1}\rho_{11}^{-2}-\frac{\rho_{11}^{-k}-\rho_{11}^{-1}}{1-\rho_{11}}\right)\frac{\rho_{21}\gamma_{12}^{\ell_1}p_{2,\ell_1}}{1-\gamma_{12}} \end{aligned}$$

for $k\in[1,\ell_1]$. Hence, multiplying both sides of this equation by ρ_{11}^{k}, we have

$$\begin{aligned} p_{1,k} &= \rho_{11}^{k}p_{1,0}-\frac{1-\rho_{11}^{k}}{1-\rho_{11}}\gamma_{12}^{\ell_1-1}p_{2,\ell_1}+\frac{1-\rho_{11}^{k-1}}{1-\rho_{11}}\frac{\rho_{21}}{1-\gamma_{12}}\gamma_{12}^{\ell_1}p_{2,\ell_1}\\ &\quad-\frac{\gamma_{12}^{-(k-1)}-\rho_{11}^{k-1}}{(\rho_{11}\gamma_{12})^{-1}-1}\frac{\rho_{11}^{-1}\rho_{21}}{1-\gamma_{12}}\gamma_{12}^{\ell_1}p_{2,\ell_1}, \qquad k\in[1,\ell_1]. \end{aligned} \tag{A.113}$$

In particular, for $k=\ell_1$, substituting $p_{2,\ell_1}=\rho_{11}p_{1,\ell_1}$ (A.111), we have

$$\rho_{11}^{\ell_1}\frac{p_{1,0}}{p_{1,\ell_1}}=1+\frac{1-\rho_{11}^{\ell_1}}{1-\rho_{11}}\gamma_{12}^{\ell_1-1}\rho_{11}-\frac{1-\rho_{11}^{\ell_1-1}}{1-\rho_{11}}\frac{\rho_{11}\rho_{21}}{1-\gamma_{12}}\gamma_{12}^{\ell_1}+\frac{\rho_{11}\gamma_{12}-\rho_{11}^{\ell_1}\gamma_{12}^{\ell_1}}{(\rho_{11}\gamma_{12})^{-1}-1}\frac{\rho_{11}^{-1}\rho_{21}}{1-\gamma_{12}}. \tag{A.114}$$

Thus, $p_{1,0}$ is explicitly related to p_{1,ℓ_1}.

We now apply these computations to the rth two-level $M/M/1$ queue, and compute the limit of its scaled stationary distribution through moment generating functions. For the rth system, we use the superscript $^{(r)}$ as we have used for the rth two-level $GI/G/1$ queue. For example, $p_{j,\ell}^{(r)}$ is the stationary probability of $\{(J,L)=(j,\ell)\}$ for $(j,\ell)\in S$.

Define the moment generating functions $g_j^{(r)}(\theta)$ for $j=1,2,3$ as

$$g_j^{(r)}(\theta)=\sum_{k=0}^{\ell_1^{(r)}}e^{r\theta k}p_{j,k}^{(r)},\quad \theta\in\mathbb{R}, j=1,2, \qquad g_3^{(r)}(\theta)=\sum_{k=\ell_1^{(r)}+1}^{\infty}e^{r\theta k}p_{3,k}^{(r)},\quad \theta\le 0,$$

and define $g^{(r)}(\theta)=\sum_{j=1}^{3}g_j^{(r)}(\theta)$. Obviously, $g^{(r)}(\theta)$ is the moment generating function of $rL^{(r)}$. Thus, our task is to compute the limit of $g^{(r)}(\theta)$ as $r\downarrow 0$. To this end, we compute $g_j^{(r)}(\theta)$ for $j=1,2,3$, using the normalization condition that $g^{(r)}(0)=1$. However, it should be noted that this condition is not sufficient to prove the weak convergence of the distribution of $rL^{(r)}$, for which it is required that $\lim_{\theta\uparrow 0}\lim_{r\downarrow 0}g^{(r)}(\theta)=1$. This convergence will be verified through explicitly computing $g_j^{(r)}(\theta)$ for $j=1,2,3$.

We first compute $g_j^{(r)}(\theta)$ for $j=2,3$ because they are easier. It is immediate from (A.107), (A.106), and (A.105) that

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

$$g_2^{(r)}(\theta) = p_{2,0}^{(r)} + e^{\ell_1^{(r)} r\theta} \sum_{\ell=1}^{\ell_1^{(r)}} e^{-(\ell_1^{(r)}-\ell)r\theta} (\gamma_{12}^{(r)})^{\ell_1^{(r)}-\ell} p_{2,\ell_1^{(r)}}^{(r)}$$
$$= \left((\rho_{21}^{(r)})^{-1} (\gamma_{12}^{(r)})^{\ell_1^{(r)}-1} + e^{r\theta} \frac{e^{\ell_1^{(r)} r\theta} - (\gamma_{12}^{(r)})^{\ell_1^{(r)}}}{e^{r\theta} - \gamma_{12}^{(r)}} \right) p_{2,\ell_1^{(r)}}^{(r)}, \tag{A.115}$$

$$g_3^{(r)}(\theta) = \frac{1}{1 - e^{r\theta}\rho_2^{(r)}} e^{(\ell_1^{(r)}+1)r\theta} p_{3,\ell_1^{(r)}+1}^{(r)}. \tag{A.116}$$

On the other hand, we compute $g_1^{(r)}(\theta)$ from (A.113) as

$$g_1^{(r)}(\theta) = \sum_{k=0}^{\ell_1^{(r)}} (e^{r\theta}\rho_1^{(r)})^k p_{1,0}^{(r)} - \sum_{k=1}^{\ell_1^{(r)}} \frac{e^{r\theta k} - (e^{r\theta}\rho_1^{(r)})^k}{1 - \rho_1^{(r)}} (\gamma_{12}^{(r)})^{\ell_1^{(r)}-1} p_{2,\ell_1^{(r)}}^{(r)}$$
$$+ \sum_{k=1}^{\ell_1^{(r)}} \frac{e^{r\theta(k-1)} - (e^{r\theta}\rho_1^{(r)})^{k-1}}{1 - \rho_1^{(r)}} \frac{e^{r\theta}\rho_{21}^{(r)}}{1 - \gamma_{12}^{(r)}} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)}$$
$$- \sum_{k=1}^{\ell_1^{(r)}} \frac{(e^{r\theta}(\gamma_{12}^{(r)})^{-1})^{k-1} - (e^{r\theta}\rho_1^{(r)})^{k-1}}{(\rho_1^{(r)}\gamma_{12}^{(r)})^{-1} - 1} \frac{e^{r\theta}(\rho_1^{(r)})^{-1}\rho_{21}^{(r)}}{1 - \gamma_{12}^{(r)}} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1}^{(r)}. \tag{A.117}$$

To ease further compuations, we introduce the following functions:

$$f_1^{(r)}(\theta) = \sum_{k=0}^{\ell_1^{(r)}-1} e^{r\theta k}, \qquad f_{\rho_1}^{(r)}(\theta) = \sum_{k=0}^{\ell_1^{(r)}-1} (e^{r\theta}\rho_1^{(r)})^k, \qquad f_{\gamma_{12}^{-1}}^{(r)}(\theta) = \sum_{k=0}^{\ell_1^{(r)}-1} (e^{r\theta}(\gamma_{12}^{(r)})^{-1})^k,$$

which are computed as

$$f_1^{(r)}(\theta) = \frac{1 - (e^{r\theta})^{\ell_1^{(r)}}}{1 - e^{r\theta}} 1(\theta \neq 0) + \ell_1^{(r)} 1(\theta = 0), \qquad f_{\rho_1}^{(r)}(\theta) = \frac{1 - (e^{r\theta}\rho_1^{(r)})^{\ell_1^{(r)}}}{1 - e^{r\theta}\rho_1^{(r)}},$$
$$f_{\gamma_{12}^{-1}}^{(r)}(\theta) = \frac{e^{r\theta\ell_1^{(r)}} - (\gamma_{12}^{(r)})^{\ell_1^{(r)}}}{e^{r\theta} - \gamma_{12}^{(r)}} (\gamma_{12}^{(r)})^{-\ell_1^{(r)}+1}.$$

Then, using these function, (A.117) becomes

$$g_1^{(r)}(\theta) = \left(f_{\rho_1}^{(r)}(\theta) + (e^{r\theta}\rho_1^{(r)})^{\ell_1^{(r)}} \right) p_{1,0}^{(r)} - \left(f_1^{(r)}(\theta) - \rho_1^{(r)} f_{\rho_1}^{(r)}(\theta) \right) \frac{e^{r\theta}(\gamma_{12}^{(r)})^{\ell_1^{(r)}-1}}{1 - \rho_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)}$$
$$+ \left(f_1^{(r)}(\theta) - f_{\rho_1}^{(r)}(\theta) \right) \frac{e^{r\theta}\rho_{21}^{(r)}}{(1 - \rho_1^{(r)})(1 - \gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)}$$
$$- \left(f_{\gamma_{12}^{-1}}^{(r)}(\theta) - f_{\rho_1}^{(r)}(\theta) \right) \frac{e^{r\theta}(\rho_1^{(r)})^{-1}\rho_{21}^{(r)}}{((\rho_1^{(r)}\gamma_{12}^{(r)})^{-1} - 1)(1 - \gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)}. \tag{A.118}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

We now compute the limits of $g_j^{(r)}(0)$ for $j=1,2,3$. For this, we note that the following asymptotic properties are obtained from the assumptions (A2)–(A5):

$$(\gamma_{12}^{(r)})^{\ell_1^{(r)}} = O(\gamma_{12}^{\ell_1/r}) \quad \text{as } r \downarrow 0. \tag{A.119}$$

Hence, $(\gamma_{12}^{(r)})^{\ell_1^{(r)}}$ decays exponentially fast as the function of $1/r$. On the other hand,

$$\lim_{r\downarrow 0} \frac{1-\rho_\ell^{(r)}}{r} = \beta_\ell, \quad \ell = 1, 2, \tag{A.120}$$

$$\lim_{r\downarrow 0} (\rho_1^{(r)})^{\ell_1^{(r)}} = \lim_{r\downarrow 0} (1-\beta_1 r + o(r))^{\ell_1/r + o(r)} = e^{-\beta_1 \ell_1}, \tag{A.121}$$

where we recall that $\beta_\ell = b_\ell / c_\ell$. Other important asymptotic properties are the ratios of the boundary probabilities, which are presented in the next lemma.

Lemma 15. *For the sequence of the two-level $M/M/1$ queue satisfying (A2)–(A5),*

$$\lim_{r\downarrow 0} \frac{p_{2,\ell_1}^{(r)}}{p_{1,\ell_1}^{(r)}} = 1, \qquad \lim_{r\downarrow 0} \frac{p_{3,\ell_1+1}^{(r)}}{p_{1,\ell_1}^{(r)}} = \frac{\lambda_1+\lambda_2}{\lambda_2}, \qquad \lim_{r\downarrow 0} \frac{p_{1,0}^{(r)}}{p_{1,\ell_1}^{(r)}} = \frac{\lambda_1+\lambda_2}{\lambda_2} e^{\beta_1 \ell_1}. \tag{A.122}$$

Proof. The first two limits in (A.122) are immediate from (A.111) and (A.109) because

$$\lim_{r\downarrow 0} \gamma_{22}^{(r)} = \lim_{r\downarrow 0} \frac{c_2^{(r)} \mu^{(r)}}{c_1^{(r)} \mu + \lambda_2^{(r)}} = \frac{c_2 \mu}{c_1 \mu + \lambda_2} = \frac{\lambda_2}{\lambda_1+\lambda_2}.$$

For the third limit, we have, from (A.114) for the rth system,

$$\lim_{r\downarrow 0} \frac{p_{1,0}^{(r)}}{p_{1,\ell_1}^{(r)}} = \lim_{r\downarrow 0} (\rho_1^{(r)})^{-\ell_1^{(r)}} \left(1 + \frac{\rho_1^{(r)} \gamma_{12}^{(r)}}{(\rho_1^{(r)} \gamma_{12}^{(r)})^{-1} - 1} \frac{(\rho_1^{(r)})^{-1} \rho_{21}^{(r)}}{1-\gamma_{12}^{(r)}}\right).$$

Hence, applying (A.121), we have the third limit of (A.122). □

We now tentatively assume that there exists finite $\widetilde{p}_{1,\ell_1}$ such that

$$\widetilde{p}_{1,\ell_1} = \lim_{r\downarrow 0} \frac{p_{1,\ell_1^{(r)}}^{(r)}}{r}. \tag{A.123}$$

The existence of this $\widetilde{p}_{1,\ell_1}$ will be verified at the final step (see (A.128)). Applying these asymptotic properties, it follows from (A.115) and (A.116) with help of (A.123) that

$$\begin{aligned} \lim_{r\downarrow 0} g_2^{(r)}(0) &= \lim_{r\downarrow 0} \left((\rho_{21}^{(r)})^{-1} (\gamma_{12}^{(r)})^{\ell_1^{(r)}-1} + \frac{1-(\gamma_{12}^{(r)})^{\ell_1^{(r)}}}{1-\gamma_{12}^{(r)}} \right) p_{2,\ell_1^{(r)}}^{(r)} \\ &= \frac{c_1\mu+\lambda_2}{\lambda_2} \lim_{r\downarrow 0} p_{2,\ell_1^{(r)}}^{(r)} = \frac{\lambda_1+\lambda_2}{\lambda_2} \lim_{r\downarrow 0} p_{1,\ell_1^{(r)}}^{(r)} = 0, \end{aligned} \tag{A.124}$$

$$\begin{aligned} \lim_{r\downarrow 0} g_3^{(r)}(0) &= \lim_{r\downarrow 0} \frac{1}{1-\rho_2^{(r)}} p_{3,\ell_1^{(r)}+1}^{(r)} = \frac{c_2}{b_2} \lim_{r\downarrow 0} \frac{1}{r} p_{3,\ell_1^{(r)}+1}^{(r)} \\ &= \beta_2^{-1} \gamma_{22}^{-1} \lim_{r\downarrow 0} \frac{1}{r} p_{2,\ell_1^{(r)}}^{(r)} = \frac{\lambda_1+\lambda_2}{\lambda_2} \beta_2^{-1} \widetilde{p}_{1,\ell_1}. \end{aligned} \tag{A.125}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

To compute the limit of $g_1^{(r)}(0)$, we first compute it from (A.118) as

$$\begin{aligned} g_1^{(r)}(0) &= \left(f_{\rho_1}^{(r)}(0) + (\rho_1^{(r)})^{\ell_1^{(r)}}\right) p_{1,0}^{(r)} - \left(f_1^{(r)}(0) - \rho_1^{(r)} f_{\rho_1}^{(r)}(0)\right) \frac{(\gamma_{12}^{(r)})^{\ell_1^{(r)}-1}}{1-\rho_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)} \\ &\quad + \left(f_1^{(r)}(0) - f_{\rho_1}^{(r)}(0)\right) \frac{\rho_{21}^{(r)}}{(1-\rho_1^{(r)})(1-\gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)} \\ &\quad - \left(f_{\gamma_{12}^{-1}}^{(r)}(0) - f_{\rho_1}^{(r)}(0)\right) \frac{(\rho_1^{(r)})^{-1}\rho_{21}^{(r)}}{((\rho_1^{(r)}\gamma_{12}^{(r)})^{-1}-1)(1-\gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)}. \end{aligned} \tag{A.126}$$

To proceed to further compute, we note that

$$\lim_{r\downarrow 0} r f_1^{(r)}(0) = \lim_{r\downarrow 0} r\ell_1^{(r)} = \ell_1,$$

$$\lim_{r\downarrow 0} r f_{\rho_1}^{(r)}(0) = \lim_{r\downarrow 0} r \frac{1-(\rho_1^{(r)})^{\ell_1^{(r)}}}{1-\rho_1^{(r)}} = \beta_1^{-1}(1-e^{-\beta_1\ell_1}),$$

$$\lim_{r\downarrow 0} f_{\gamma_{12}^{-1}}^{(r)}(0)(\gamma_{12}^{(r)})^{\ell_1^{(r)}} = \lim_{r\downarrow 0} \frac{1-(\gamma_{12}^{(r)})^{\ell_1^{(r)}}}{1-\gamma_{12}^{(r)}} \gamma_{12}^{(r)} = \frac{c_1\mu}{\lambda_2} = \rho_{21}^{-1}.$$

Then, we have, using Lemma 15,

$$\begin{aligned} \lim_{r\downarrow 0} \left(f_{\rho_1}^{(r)}(0) + (\rho_1^{(r)})^{\ell_1^{(r)}}\right) p_{1,0}^{(r)} &= \beta_1^{-1}(1-e^{-\beta_1\ell_1}) \lim_{r\downarrow 0} \frac{1}{r} p_{1,0}^{(r)} \\ &= \beta_1^{-1}\left(e^{\beta_1\ell_1}-1\right) \frac{\lambda_1+\lambda_2}{\lambda_2} \widetilde{p}_{1,\ell_1}, \end{aligned}$$

$$\lim_{r\downarrow 0} \left(f_1^{(r)}(0) - \rho_1^{(r)} f_{\rho_1}^{(r)}(0)\right) \frac{(\gamma_{12}^{(r)})^{\ell_1^{(r)}-1}}{1-\rho_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)} = 0,$$

$$\lim_{r\downarrow 0} \left(f_1^{(r)}(0) - f_{\rho_1}^{(r)}(0)\right) \frac{\rho_{21}^{(r)}}{(1-\rho_1^{(r)})(1-\gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)} = 0,$$

and

$$\begin{aligned} &\lim_{r\downarrow 0} \left(f_{\gamma_{12}^{-1}}^{(r)}(0) - f_{\rho_1}^{(r)}(0)\right) \frac{(\rho_1^{(r)})^{-1}\rho_{21}^{(r)}}{((\rho_1^{(r)}\gamma_{12}^{(r)})^{-1}-1)(1-\gamma_{12}^{(r)})} (\gamma_{12}^{(r)})^{\ell_1^{(r)}} p_{2,\ell_1^{(r)}}^{(r)} \\ &\quad = \lim_{r\downarrow 0} f_{\gamma_{12}^{-1}}^{(r)}(0)(\gamma_{12}^{(r)})^{\ell_1^{(r)}} \frac{(\rho_1^{(r)})^{-1}\rho_{21}^{(r)}}{((\rho_1^{(r)}\gamma_{12}^{(r)})^{-1}-1)(1-\gamma_{12}^{(r)})} p_{2,\ell_1^{(r)}}^{(r)} = 0. \end{aligned}$$

Hence, from (A.126)

$$\lim_{r\downarrow 0} g_1^{(r)}(0) = \frac{\lambda_1+\lambda_2}{\lambda_2} \beta_1^{-1}(e^{\beta_1\ell_1}-1)\widetilde{p}_{1,\ell_1}. \tag{A.127}$$

Hence, by (A.124) and (A.125), we have

$$1 = \lim_{r\downarrow 0} g^{(r)}(0) = \frac{\lambda_1+\lambda_2}{\lambda_2} \left(\beta_2^{-1} + \beta_1^{-1}(e^{\beta_1\ell_1}-1)\right) \widetilde{p}_{1,\ell_1}. \tag{A.128}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

This proves that $\widetilde{p}_{1,\ell_1}$ indeed exists and is finite, and given by

$$\widetilde{p}_{1,\ell_1} = \frac{\lambda_2\beta_1\beta_2}{(\lambda_1+\lambda_2)(\beta_1+\beta_2(e^{\beta_1\ell_1}-1))}. \tag{A.129}$$

We are now in a final step to find $\widetilde{g}(\theta) \equiv \lim_{r\downarrow 0} g^{(r)}(\theta)$ for $\theta < 0$. This can be obtained in the following way. From (A.116) and Lemma 15,

$$\begin{aligned}
\lim_{r\downarrow 0} g_3^{(r)}(\theta) &= \lim_{r\downarrow 0} \frac{1}{1-e^{r\theta}\rho_2^{(r)}} e^{(\ell_1^{(r)}+1)r\theta} p_{3,\ell_1^{(r)}+1}^{(r)} \\
&= \lim_{r\downarrow 0} \frac{r}{1-(1+r\theta+o(r\theta))(1-r\beta_2+o(r))} e^{\ell_1^{(r)}\theta} \frac{1}{r} p_{3,\ell_1^{(r)}+1}^{(r)} \\
&= \frac{e^{\ell_1\theta}}{\beta_2-\theta} \frac{\lambda_1+\lambda_2}{\lambda_2} \lim_{r\downarrow 0} \frac{1}{r} p_{1,\ell_1^{(r)}}^{(r)} \\
&= \frac{e^{\ell_1\theta}\beta_2}{\beta_2-\theta} \frac{\beta_1}{\beta_1+\beta_2(e^{\beta_1\ell_1}-1)},
\end{aligned}$$

while, from (A.118) and Lemma 15,

$$\begin{aligned}
\lim_{r\downarrow 0} g_1^{(r)}(\theta) &= \lim_{r\downarrow 0} \left(f_{\rho_1}^{(r)}(\theta) + (e^{r\theta}\rho_1^{(r)})^{\ell_1^{(r)}} \right) p_{1,0}^{(r)} \\
&= \lim_{r\downarrow 0} \frac{1-(e^{r\theta}\rho_1^{(r)})^{\ell_1^{(r)}+1}}{1-e^{r\theta}\rho_1^{(r)}} p_{1,0}^{(r)} \\
&= \frac{1-e^{\ell_1(\theta-\beta_1)}}{\beta_1-\theta} \frac{\lambda_1+\lambda_2}{\lambda_2} e^{\beta_1\ell_1} \widetilde{p}_{1,\ell_1} \\
&= \frac{\beta_1}{e^{\beta_1\ell_1}-1} \frac{e^{\beta_1\ell_1}-e^{\ell_1\theta}}{\beta_1-\theta} \frac{\beta_2(e^{\beta_1\ell_1}-1)}{\beta_1+\beta_2(e^{\beta_1\ell_1}-1)}.
\end{aligned}$$

Furthermore, $\lim_{r\downarrow 0} g_2^{(r)}(\theta) = 0$ by (A.124). Hence, we have, for $\theta \le 0$ satisfying $\theta \ne \beta_1$,

$$\begin{aligned}
\lim_{r\downarrow 0} g^{(r)}(\theta) &= \lim_{r\downarrow 0} g_1^{(r)}(\theta) + \lim_{r\downarrow 0} g_3^{(r)}(\theta) \\
&= \frac{\beta_1}{e^{\beta_1\ell_1}-1} \frac{e^{\beta_1\ell_1}-e^{\ell_1\theta}}{\beta_1-\theta} \frac{\beta_2(e^{\beta_1\ell_1}-1)}{\beta_1+\beta_2(e^{\beta_1\ell_1}-1)} + \frac{e^{\ell_1^{(r)}\theta}\beta_2}{\beta_2-\theta} \frac{\beta_1}{\beta_1+\beta_2(e^{\beta_1\ell_1}-1)}.
\end{aligned}$$

This completes the proof of Proposition 1.

A.4. Proofs of Lemma 5

We first prove (4.41) and (4.42). Apply $f(z,x) = x_1^{k+1}$ to (4.17), then

$$\begin{aligned}
(k+1)\mathbb{E}[(R_e^{(r)})^k] = \alpha_e^{(r)}\mathbb{E}[(T_{e,1}^{(r)})^{k+1})]\mathbb{P}_e^{(r)}[L^{(r)}(0) \le \ell_1^{(r)}] \\
+ \alpha_e^{(r)}\mathbb{E}[(T_{e,2}^{(r)})^{k+1})]\mathbb{P}_e^{(r)}[L^{(r)}(0) > \ell_1^{(r)}],
\end{aligned}$$

where (4.17) is applicable since the right-hand side of this equation is finite. Hence, (4.41) is obtained. Similarly, we have (4.42) applying $f(z,x) = x_2^{k+1}$ to (4.17) because

$$f'(X^{(r)}(t)) = -(k+1)(R_s^{(r)}(t))^k(1-1(L^{(r)}(t)=0)),$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

and $R^{(r)}(t)$ has the same distribution as $T_s^{(r)}$ when $L^{(r)}(t)=0$. We next prove (4.43). For this, we use the following inequality:

$$|(x\wedge 1/r)^k - x^k| = x^k 1(x>1/r) < rx^{k+1}1(x>1/r) < r^2x^{k+2}1(x>1/r), \quad x\ge 0,\ k\ge 1. \tag{A.130}$$

Hence, the last inequality in (A.130) for $k=1$ and the uniform boundedness of $\mathbb{E}[(T_u^{(r)})^3$ yield

$$|\mathbb{E}[\widehat{T}_u^{(r)}] - \mathbb{E}[T_u^{(r)}]| \le r^2\mathbb{E}[(T_u^{(r)})^3 1(T_u^{(r)}>1/r)] = o(r^2). \tag{A.131}$$

On the other hand, the second inequality in (A.130) yields

$$|\mathbb{E}[(\widehat{T}_u^{(r)})^2] - \mathbb{E}[(T_u^{(r)})^2]| \le r\mathbb{E}[(T_u^{(r)})^2 1(T_u^{(r)}>1/r)] = o(r). \tag{A.132}$$

Thus, (4.43) is proved for $k=1,2$. We next prove (4.44) for $v=e$. Since $\mathbb{E}[(R_e^{(r)})^k]$ is uniformly bounded for $k=1,2$ by (4.42), we can prove (4.44) by the same argument as the proof of (4.42), but the order of r is one less because of (4.42). We omit the proof for $v=s$ because it is similar to the case of $v=e$ except for the fact that (4.41) is used instead of (4.42).

A.5. Proofs of Lemmas 8 and 9

Proof of Lemma 8. For a $(n+1)$-times continuously differentiable function f, we have, from Taylor expansion (see, e.g., [14, Theorem 19.9]),

$$\begin{aligned}\left|f(x) - \sum_{k=0}^{n}\frac{1}{k!}f^{(k)}(0)x^k\right| &= \frac{1}{n!}\left|\int_0^x (x-y)^n f^{(n+1)}(y)\,dy\right| \\ &\le \frac{1}{(n+1)!}|x|^{n+1}\max_{|u|\le|x|}|f^{(n+1)}(u)|, \quad x\in\mathbb{R},\ n\ge 0.\end{aligned} \tag{A.133}$$

Applying $f(x)=e^{-x}$ to this inequality, we have $|e^{-x}-1+x|\le\frac{1}{2}|x|^2e^{|x|}$. Hence, for each fixed $\theta\in\mathbb{R}$, as $r\downarrow 0$,

$$\begin{aligned}e^{r\theta\ell_1^{(r)}} &= e^{\theta(\ell_1+o(r))} = e^{\theta\ell_1}(1+\theta o(r)) + o(\theta o(r)) = e^{\theta\ell_1}(1+o(\theta r)),\\ e^{r\theta(\ell_1^{(r)}+1)} &= e^{\theta r\ell_1^{(r)}}e^{r\theta} = e^{\theta\ell_1}(1+o(\theta r))(1+\theta r+o((\theta r)^2)) = e^{\theta\ell_1}(1+\theta r+o(\theta r)).\end{aligned}$$

Thus, (4.52) is obtained. Equations (4.53) and (4.54) are immediate from these asymptotic expansions.

Proof of Lemma 9. Since $\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}$ and $\eta_i^{(r)}(r\theta)\widehat{R}_e^{(r)}$ are uniformly bounded in $r<a$ for a sufficiently small $a>0$ and vanish as $r\downarrow 0$, it follows from Lemma 3 and (A.133) that

$$\begin{aligned}&e^{-\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}}\\ &\quad= 1-\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)} + \tfrac{1}{2}(\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)})^2 + O\Big(|\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}|^3\max_{r|\theta|<a}e^{|\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}|}\Big)\\ &\quad= 1-(-\mu^{(r)}r\theta+\tfrac{1}{2}(\mu^{(r)})^3(\sigma_s^{(r)})^2(r\theta)^2+o((r\theta)^2))\widehat{R}_{s-}^{(r)}\\ &\qquad+\tfrac{1}{2}(-\mu^{(r)}r\theta+\tfrac{1}{2}(\mu^{(r)})^3(\sigma_s^{(r)})^2(r\theta)^2+o((r\theta)^2))^2(\widehat{R}_{s-}^{(r)})^2\\ &\qquad+(\widehat{R}_{s-}^{(r)})^3O\Big((r\theta)^3\max_{r|\theta|<a}e^{|\zeta^{(r)}(r\theta)\widehat{R}_{s-}^{(r)}|}\Big)\\ &\quad= 1+r\theta(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - \tfrac{1}{2}(r\theta)^2((\mu^{(r)}\sigma_s^{(r)})^2(\mu^{(r)}\widehat{R}_{s-}^{(r)}) - (\mu^{(r)}\widehat{R}_{s-}^{(r)})^2) + S_3(\widehat{R}_{s-}^{(r)})o((r\theta)^2),\end{aligned}$$

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

where we recall that $S_3(\widehat{R}_{s-}^{(r)}) = \widehat{R}_{s-}^{(r)} + (\widehat{R}_{s-}^{(r)})^2 + (\widehat{R}_{s-}^{(r)})^3$. This proves (4.56), and (4.58) and (4.57) are similarly proved.

A.6. Proof of Lemma 13

We first assume that $\lambda_1 \neq \lambda_2$, and prove (5.75) for $i = 1$. Consider the BAR (4.40) for $\theta_2 = \theta_1 = \theta \leq 0$. From (4.39), we have

$$D^{(r)}(\theta, \theta) = \alpha_e^{(r)} e^{r\theta(\ell_1^{(r)}+1)} \mathbb{E}_s^{(r)}[(e^{-\eta_1^{(r)}(r\theta)\widehat{R}_e^{(r)}} - e^{-\eta_2^{(r)}(r\theta)\widehat{R}_e^{(r)}}) 1(L^{(r)}(0) = \ell_1^{(r)})]. \tag{A.134}$$

Applying (4.52), (4.57), and Corollary 2 to (A.134), we have

$$\begin{aligned} D^{(r)}(\theta, \theta) &= \alpha_e^{(r)} e^{\theta\ell_1}(1 + r\theta + o(r\theta)) \mathbb{E}_s^{(r)}\big[(r\theta(\lambda_2^{(r)} - \lambda_1^{(r)})\widehat{R}_e^{(r)} \\ &\quad + 2^{-1} r^2\theta^2\{(\lambda_2^{(r)}\sigma_{e,2}^{(r)})^2\lambda_2^{(r)} - (\lambda_1^{(r)}\sigma_{e,1}^{(r)})^2\lambda_1^{(r)}\}\widehat{R}_e^{(r)} \\ &\quad - 2^{-1} r^2\theta^2\{(\lambda_2^{(r)})^2 - (\lambda_1^{(r)})^2\}(\widehat{R}_e^{(r)})^2) 1(L^{(r)}(0) = \ell_1^{(r)})\big] + o((r\theta)^2), \\ &= \alpha_e^{(r)} e^{\theta\ell_1}\big\{r\theta(\lambda_2^{(r)} - \lambda_1^{(r)}) \mathbb{E}_s^{(r)}\big[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})\big] + (r\theta)^2(\mathcal{E}_1^{(r)} + \mathcal{E}_2^{(r)})\big\} \\ &\quad + o((r\theta)^2). \end{aligned}$$

Since $\mathcal{E}_1^{(r)} + \mathcal{E}_2^{(r)}$ is uniformly bounded in $r \in (0, 1]$, this implies that

$$D^{(r)}(\theta, \theta) = r\theta\alpha_e^{(r)} e^{\theta\ell_1}(\lambda_2^{(r)} - \lambda_1^{(r)}) \mathbb{E}_s^{(r)}\big[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})\big] + O((r\theta)^2). \tag{A.135}$$

Substituting (A.135) into the BAR (4.40) for $\theta_2 = \theta_1 = \theta \leq 0$, which is given by

$$\begin{aligned} &\frac{1}{2}\sum_{i=1,2} c_i\sigma_i^2\big((\beta_i - \theta)\mu\theta r^2 + o((\theta r)^2)\big)\psi_i^{(r)}(\theta) \\ &\quad = c_1\left(\mu r\theta - \frac{\mu^3\sigma_s^2 r^2\theta^2}{2} + o((\theta r)^2)\right) \mathbb{E}[1(L^{(r)} = 0) g_{1,r\theta}^{(r)}(R^{(r)})] + D^{(r)}(\theta, \theta), \end{aligned} \tag{A.136}$$

then picking up the terms of order $r\theta$, we have

$$c_1\mu\mathbb{E}[1(L^{(r)} = 0) g_{1,r\theta}^{(r)}(R^{(r)})] + \alpha_e^{(r)} e^{\theta\ell_1}(\lambda_2 - \lambda_1)\mathbb{E}_s[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})] = O(r\theta).$$

Since $\mathbb{E}[1(L^{(r)} = 0) g_{1,r\theta}^{(r)}(R^{(r)})] = O(r)$ by Corollary 3 and Lemma 10, this yields

$$(\lambda_2 - \lambda_1)\mathbb{E}_s[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})] = O(r). \tag{A.137}$$

Since $\lambda_1 \neq \lambda_2$, (A.137) implies (5.75) for $i = 1$. To prove (5.75) for $i = 2$, we derive the following formula by Schwartz's inequality:

$$\begin{aligned} \big\{\mathbb{E}_s^{(r)}[(\widehat{R}_e^{(r)})^2 1(L^{(r)}(0) = \ell_1^{(r)}]\big\}^2 &= \big\{\mathbb{E}_s^{(r)}[(\widehat{R}_e^{(r)})^{3/2}(\widehat{R}_e^{(r)})^{1/2} 1^2(L^{(r)}(0) = \ell_1^{(r)})]\big\}^2 \\ &\leq \mathbb{E}_s^{(r)}[(\widehat{R}_e^{(r)})^3 1(L^{(r)}(0) = \ell_1^{(r)})] \mathbb{E}_s^{(r)}[\widehat{R}_e^{(r)} 1(L^{(r)}(0) = \ell_1^{(r)})]. \end{aligned} \tag{A.138}$$

Since $\mathbb{E}_s^{(r)}[(\widehat{R}_e^{(r)})^3 1(L^{(r)}(0) = \ell_1^{(r)})]$ is uniformly bounded in r by Corollary 5, (5.75) for $i = 1$ implies that for $i = 2$. Thus, (5.75) is proved.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

We next prove that (5.76) implies (5.75). For this proof, we use the BAR (4.17) for $f(z, x_1, x_2) = (x_1 \wedge 1/r)(x_2 \wedge 1/r)1(z = \ell_0^{(r)})$, then we have

$$\begin{aligned}&\mathbb{E}[R_e^{(r)}1(R_e^{(r)} \le 1/r, L^{(r)} = \ell_0^{(r)})] + \mathbb{E}[R_s^{(r)}1(R_s^{(r)} \le 1/r, L^{(r)} = \ell_0^{(r)})] \\ &\quad = \alpha_e^{(r)}\mathbb{E}[\widehat{T}_{e,2}^{(r)}]\mathbb{E}_e^{(r)}[\widehat{R}_{s-}^{(r)}1(L^{(r)}(0) = \ell_0^{(r)})] + \alpha_e^{(r)}\mathbb{E}[\widehat{T}_s^{(r)}]\mathbb{E}_s^{(r)}[\widehat{R}_e^{(r)}1(L^{(r)}(0) = \ell_0^{(r)})]. \quad \text{(A.139)}\end{aligned}$$

Then, we can see that the left-hand side of (A.139) vanishes as $r \downarrow 0$ by (5.76) because, by Schwartz's inequality,

$$\left(\mathbb{E}[R_v^{(r)}1(R_v^{(r)} \le 1/r, L^{(r)} = \ell_0^{(r)})]\right)^2 \le \mathbb{E}[(R_v^{(r)})^2]\mathbb{P}[L^{(r)} = \ell_0^{(r)}], \quad v = e, s.$$

Hence, we have (5.75) for $i = 1$. It for $i = 2$ follows from (A.138) as we have argued.

A.7. Proof of Lemma 14

In the proof of Proposition 3, under the assumption that $\lambda_1 \neq \lambda_2$, (5.74) is obtained by applying (5.75) to (5.91). Obviously, if (5.75) holds then this proof is also valid without any extra assumption including $\lambda_1 \neq \lambda_2$. Hence, the lemma is proved.

Acknowledgements

This study was motivated by the BAR approach, coined by Jim Dai (see, e.g., [4]). I am grateful to him for his grand design about it, and to Rami Atar for providing his recent papers [1, 2]. I also benefited from Evsey Morozov for his comments on this paper.

Funding information

There are no funding bodies to thank relating to the creation of this paper.

Competing interests

There were no competing interests to declare that arose during the preparation or publication process of this paper.

References

[1] Atar, R., Castiel, E. and Reiman, M. (2024). Asymptotic optimality of switched control policies in a simple parallel server system under an extended heavy traffic condition. *Stochastic Systems* (to appear).
[2] Atar, R. and Wolansky, G. (2024). *Invariance principle and McKean-Vlasov limit for randomized load balancing in heavy traffic*. Preprint. Technion - Israel Institute of Technology.
[3] Braverman, A., Dai, J. and Miyazawa, M. (2017). Heavy traffic approximation for the stationary distribution of a generalized Jackson network: the BAR approach. *Stochastic Systems* **7**, 143–196.
[4] Braverman, A., Dai, J. and Miyazawa, M. (2024). The BAR-approach for multiclass queueing networks with SBP service policies. *Stochastic Systems*. Published online in Articles in Advance.
[5] Davis, M. H. A. (1993). *Markov Models and Optimization*. Monographs on Statistics and Applied Probability 49. Chapman & Hall.
[6] Harrison, J. M. (2013). *Brownian Models of Performance and Control*. Cambridge University Press, New York.
[7] Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society* **57**, 902–904.
[8] Mandelbaum, A. and Pats, G. (1998). State-dependent stochastic networks. Part I: approximations and applications with continuous diffusion limits. *Annals of Applied Probability* **8**, 569–646.
[9] Miyazawa, M. (1994). Rate conservation laws: a survey. *Queueing Systems* **15**, 1–58.
[10] Miyazawa, M. (2017). A unified approach for large queue asymptotics in a heterogeneous multiserver queue. *Adv. in Appl. Probab.* **49**, 182–220.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press

[11] MIYAZAWA, M. (2024). Multi-level reflecting Brownian motion on the half line and its stationary distribution. *Journal of the Indian Society for Probability and Statistics* **25**, 543–574.
[12] MIYAZAWA, M. (2024). Palm problems arising in bar approach and its applications. *Queueing Systems* **108**, 253–273.
[13] MIYAZAWA, M. (2024). The stationary distributions of state-dependent diffusions reflected at one and two sides. Technical report.
[14] RUDIN, W. (1987). *Real complex analysis* 3rd ed. International Series in Pure and Applied Mathematics. McGraw-Hill.
[15] YAMADA, K. (1995). Diffusion approximation for open state-dependent queueing networks in the heavy traffic situation. *Annals of Applied Probability* **5**, 958–982.

https://doi.org/10.1017/apr.2025.5 Published online by Cambridge University Press