MATHEMATICS OF OPERATIONS RESEARCH
Vol. 34, No. 1, February 2009, pp. 142–179
ISSN 0364-765X | EISSN 1526-5471 | 09 | 3401 | 0142
DOI 10.1287/moor.1080.0350

Multiserver Loss Systems with Subscribers

Ramandeep S. Randhawa
McCombs School of Business, The University of Texas at Austin, Austin, Texas 78712,
ramandeep.randhawa@mccombs.utexas.edu, http://www.mccombs.utexas.edu/faculty/ramandeep.randhawa

Sunil Kumar
Graduate School of Business, Stanford University, Stanford, California 94305
kumar_sunil@gsb.stanford.edu, http://www.stanford.edu/~skumar

We study a multiserver loss system with two kinds of customers: subscribers and infrequent users. We model the infrequent users' requests for service by a Poisson process. However, noting that the Poisson process is unable to capture repeated interactions as well as retrials, we propose a Markovian on-off-hold model for the subscribers' requests for service that takes into account retrials by subscribers denied service. We analyze this system in an asymptotic regime where the number of subscribers and the arrival rate of the Poisson stream, and consequently the number of servers, grow without bound. In this regime, we identify and prove convergence to diffusion limits for the system. We also prove the convergence of the invariant distributions to the invariant distribution of the limiting diffusions.

Key words: loss systems; subscribers with retrials; on-off-hold request model
MSC2000 subject classification: Primary: 60K25, 68M20; secondary: 60J60, 93E03, 60B10
OR/MS subject classification: Primary: queues: applications, limit theorems, Markovian; secondary: probability: stochastic model application, diffusion
History: Received: August 8, 2006; revised April 22, 2007, August 28, 2007, April 19, 2008, and July 2, 2008.

1. Introduction. The study of multiserver loss models, where customers who are not served on request are lost, has a long history going back to Erlang (Erlang [13], Gross and Harris [16]). Such models are appropriate for many applications like telephone circuits, rental systems, etc. Most of these models assume Poisson or, at best, renewal arrival streams, independent of the number of servers being utilized. This assumption is justifiable in some applications, but it is not appropriate in settings where the servers are used by a fixed pool of subscribers and the number of servers is not negligible when compared to the size of the subscriber pool. Such a scenario arises in rental systems, for example, where the number of units available to rent might be comparable to the number of subscribers, and a nonnegligible fraction of the subscribers could be renters at any given time.

The setting of a rental system serves as a canonical application of the models studied in this paper. In particular, our model corresponds to rental systems that have a pool of subscribers as well as an exogenous stream of pay-per-use customers. The subscribers repeatedly request for service whereas the pay-per-use customers constitute a Poisson arrival stream and can be thought of as infrequent users of the system. We model each subscriber as a Markov chain having three states: on, off, and hold. A subscriber spends an exponentially distributed amount of time with mean $1/\lambda$ in the off state, after which she requests for a server. If no servers are available, she transits to the hold state. Otherwise, a server is assigned to her and she transits to the on state. We do not allow a server to be assigned to more than one subscriber at any time. In the hold state, she retries to obtain a server after an exponentially distributed amount of time with mean $1/\nu$ until a server is available. Once a server is assigned to her, she transits to the on state. In the on state, she uses the server for an exponentially distributed amount of time with mean $1/\mu$, after which she releases the server and transits to the off state. All the times are independent and identically distributed and independent of each other. Figure 1 illustrates the transitions and a precise formulation using Poisson processes is provided in Equation (1). This subscriber model is related to the classical Engset model (see Kleinrock [21]) used in telecommunications. In fact, for the special case in which the hold and off states are indistinguishable, i.e., when the rates at which a subscriber tries to obtain a server from the hold and off states are equal, the subscribers can be characterized by a two-state Markov chain; this is identical to the Engset model.

The natural performance metric in this system is the steady state probability that the servers are fully occupied. The objective of this paper is to provide a means of computing this probability via asymptotic analysis. To this end, we characterize the convergence of the number-in-system process under suitable scaling and parameter regimes to diffusion limits and characterize the invariant distributions of these limits. These process limits are interesting and important in their own right. Furthermore, we show that the invariant distributions of the scaled systems converge to the invariant distributions of the limiting diffusions. In addition, for the special case in which the hold and off states are indistinguishable, we show that the steady state distribution of the number-in-system depends on the means of the on and off times and not on any other moments of their distributions. This justifies the use of a Markovian model to obtain asymptotic results.

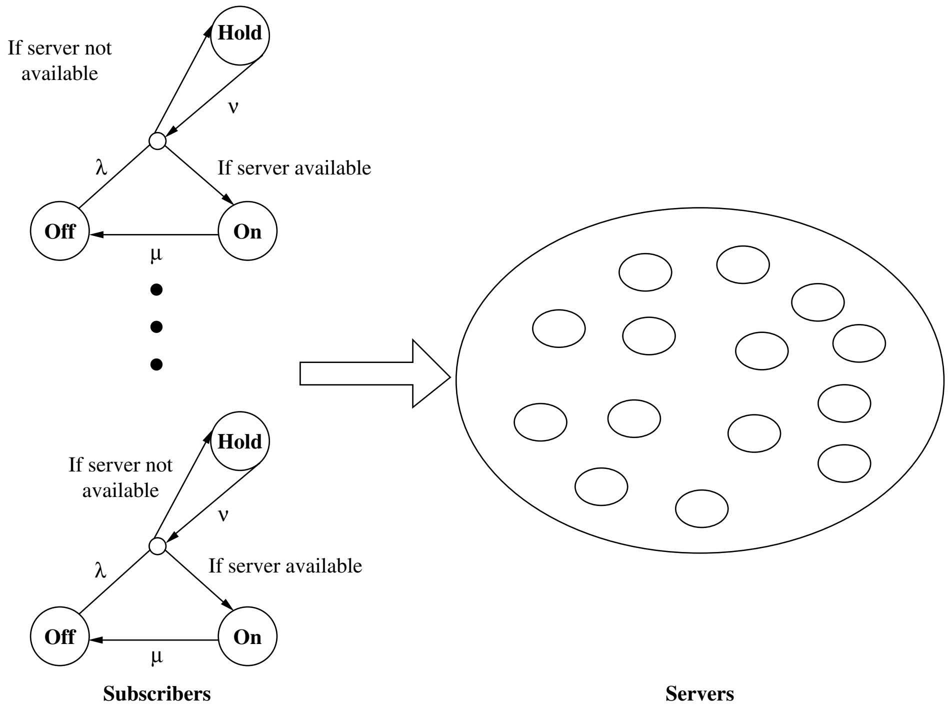

FIGURE 1. The system with n subscribers and k servers.

We study the asymptotic behavior of this system as the number of subscribers n grows without bound. In order to get an appreciable but not excessive impact from the Poisson stream, we set its rate $\lambda_p = \lambda_1 n + \lambda_2\sqrt{n}$. Further, we choose the number of servers to be of the form $k = k_1 n + k_2\sqrt{n}$ for some k_1 and k_2, as in Halfin and Whitt [17]. When k_1 is set at a level that corresponds to the nominal utilization of the system, we derive diffusion limits for the underlying system. In particular, we prove that a scaled and centered version of the number-in-system process converges to a reflected affine-drift diffusion process. When k_1 is set at a higher level, the limits obtained are similar to those for the critically loaded case, except that asymptotically this system does not behave as a loss system, i.e., the resulting diffusion is an affine-drift diffusion process without reflection. The asymptotic results for this system without reflection are quite similar to those in Mandelbaum and Pats [27] and we do not treat it here. The limits obtained when k_1 is lower than the critical level are uninformative as well as trivial. For the critically loaded case, we extend the process level convergence to steady state convergence by proving the convergence of the corresponding invariant distribution. When the hold and off states are indistinguishable for the subscribers, we are able to completely characterize the limiting invariant distribution as a truncated two-dimensional Gaussian distribution.

The convergence of the invariant distribution provides a basis for performing a static economic analysis, for example, setting prices and capacity levels as in Maglaras and Zeevi [26]. In the meantime, the diffusion limits may provide a useful basis for dynamic control, for example, dynamic allocation of servers between subscribers and exogenous arrivals along the lines of Savin et al. [38] may be feasible. Such an analysis also provides economic justification for the choice of parameters under which we obtain the asymptotic results. This is the topic of Randhawa and Kumar [34].

The model studied in this paper is related to many-server closed queuing systems. The literature for limit theory of closed queuing systems is relatively small, with single server stations studied in Harrison and Williams [18] and Kumar [24]. Krichagina and Puhalskii [23] and Kogan et al. [22] study queuing systems that allow stations to have state-dependent service rates but infinite buffers. A very recent related paper is de Véricourt and Jennings [11], which develops limit theory for a multiserver queuing system serving a pool of subscribers. This work models the system as a queue, where requests that are not accepted immediately are queued. Using the probability of queuing as the relevant performance metric, de Véricourt and Jennings [11] use diffusion limits to size capacity. Critically loaded loss systems with customers arriving as a Poisson process are a subject of study in papers such as Hunt and Kelly [19], Reiman [35], and Puhalskii and Reiman [32].

Though there are many reports in the area of limit theorems for Markovian systems, to the best of our knowledge the combination of state-dependent rates, retrials, and loss systems has not been studied before in the literature. The tools existing in the literature for proving limit theorems are not entirely adequate to handle these systems. In particular, handling boundary behavior in states where all servers are occupied and losses occur requires a new argument via an intermediate system. This argument occupies most of §5.

The paper is organized as follows. In §3, we analyze the subscriber model and obtain the asymptotic limits. In §4, we derive the asymptotic results for the general system with an exogenous Poisson stream along with the subscribers. Section 5 contains a proof of the main diffusion level convergence and §6 contains a proof of the convergence of the invariant distribution. Finally, in §7, we offer concluding comments and a comparison of the subscriber model and the exogenous arrival model.

2. Mathematical preliminaries. All the random quantities in this paper are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For some appropriate dimension m, $D_{\mathbb{R}^m}[0,\infty)$ is the space of right continuous $\mathbb{R}^m$-valued functions defined on $[0,\infty)$ with left limits endowed with the Skorohod-J_1 topology, and M_m is the Borel σ-algebra on $D_{\mathbb{R}^m}[0,\infty)$ (cf. §3.5 of Ethier and Kurtz [14]). The stochastic processes that we shall consider in this paper are measurable functions from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(D_{\mathbb{R}^m}[0,\infty), M_m)$. Let d denote the metric on $D_{\mathbb{R}^m}[0,\infty)$ as defined on p. 117 of Ethier and Kurtz [14] using the Euclidean norm on $\mathbb{R}^m$. We can write $x \in D_{\mathbb{R}^m}[0,\infty)$ as $x = (x_1, x_2, \ldots, x_m)$, where each $x_i \in D_{\mathbb{R}}[0,\infty)$. Then, for $T \geq 0$, we denote $\|x\|_T \equiv \sup_{t\leq T} \max_{i=1,\ldots,m} |x_i(t)|$[1] and use the convention that for any function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$, $\|x(f(\cdot))\|_T = \sup_{t\leq T} \max_{i=1,\ldots,m} |x_i(f(t))|$. When a stochastic process X takes values in $D_{\mathbb{R}^m}[0,\infty)$, we shall abuse notation to say that $X \in D_{\mathbb{R}^m}[0,\infty)$. We note that by the definition of the metric d, when a limit point of a sequence of random elements in $D_{\mathbb{R}^m}[0,\infty)$ has continuous paths a.s., convergence in the Skorohod topology is equivalent to uniform convergence on compact time intervals. Thus, we shall say that for a sequence of random elements $\{X^n\}$ in $D_{\mathbb{R}^m}[0,\infty)$ and a random element $X \in D_{\mathbb{R}^m}[0,\infty)$ that has continuous paths a.s., $X^n \to X$ if $\|X^n - X\|_T \to 0$, a.s. for each $T > 0$.

For a collection of probability measures P^n and P defined on $(S, \mathcal{S})$, where S is a general metric space and $\mathcal{S}$ its Borel σ-field, we say that $P^n \Rightarrow P$, i.e., P^n weakly converges to P if and only if $\int_S f\,dP^n \to \int_S f\,dP$ for all bounded, continuous real-valued functions f on S. Further, if X^n and X are random elements of this space such that P^n and P are the probability measures associated with X^n and X, respectively, then $X^n \Rightarrow X$ if and only if $P^n \Rightarrow P$. For any two random elements X, Y of $(S, \mathcal{S})$, we use $X \stackrel{d}{=} Y$ to denote equality in distribution. If $S = \mathbb{R}$ and X^n and X are random elements of this space, we define convergence in probability as $X^n \stackrel{p}{\to} X$ if $\lim_{n\to\infty} \mathbb{P}(|X^n - X| \geq \epsilon) = 0$ for all $\epsilon > 0$. Unless stated otherwise, all convergence results in this paper take place as the index $n \to \infty$.

For any stochastic process $X \in D_{\mathbb{R}^m}[0,\infty)$ and any measurable function $f\colon \mathbb{R}^m \to \mathbb{R}$, $\mathbb{E}_x f(X(t)) \equiv \mathbb{E}[f(X(t)) \mid X(0) = x]$, and if μ denotes a probability distribution on $(\mathbb{R}^m, \mathcal{R}_m)$, where $\mathcal{R}_m$ is the Borel σ-algebra on $\mathbb{R}^m$, then $\mathbb{E}_\mu f(X(t)) \equiv \int_{\mathbb{R}^m} \mathbb{E}_x f(X(t))\,d\mu(x)$. $C^{i,j,k}$ is the space of functions defined on $\mathbb{R}^3$ that are i, j, and k times continuously differentiable with respect to the first, second, and third argument, respectively.

We use the convention that for a vector $a \in \mathbb{R}^m$, a' denotes its transpose and $\|a\| = \max_{i=1,\ldots,m} |a_i|$. For two vectors $a, b \in \mathbb{R}^m$, we say $a \leq b$ if $a_i \leq b_i$ for each $i = 1, 2, \ldots, m$. We shall denote $e = (1,1,1)'$ as the unit vector in $\mathbb{R}^3$. Finally, for $f, g\colon \mathbb{R} \to \mathbb{R}$, we say that $f(n) = O(g(n))$ if there are positive constants c and k such that $0 \leq f(n) \leq cg(n)$ for all $n \geq k$, and $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.

3. Subscriber model with retrials. We begin by introducing the subscriber model and carrying out the asymptotic analysis in the regime, where the number of subscribers and the number of servers increase without bound. We initially provide a verbal description of the model, followed by a mathematical definition in Equation (1).

3.1. Model. We consider a system with n subscribers and k servers, where a server cannot be assigned to more than one subscriber at any time. Each subscriber has an underlying continuous time Markov chain $J(\cdot)$, where $J(t) \in \{\text{on}, \text{off}, \text{hold}\}$ for $t \geq 0$. A subscriber spends an exponentially distributed amount of time with mean $1/\lambda$ in the off state, after which she requests for a server. If no servers are available, she transits to the hold state; otherwise, a server is assigned to her and she transits to the on state. In the hold state, she retries to obtain a server after every exponentially distributed amount of time with mean $1/\nu$ until a server is available. Once a server is assigned to her, she transits to the on state. In the on state, she uses the server for an exponentially distributed amount of time with mean $1/\mu$, after which she returns the server and transits to the off state. We assume that all the times are independent and identically distributed and independent of each other.

[1] We would like to point out to the reader that our usage of $\|\cdot\|$ is merely notational; we do not imply that this is a norm on $D_{\mathbb{R}^m}[0,\infty)$.

This system is a special case of that discussed in §4, where in addition to the subscribers an external arrival stream of customers is also present. However, the novelty of this subscriber model and the interesting nature of its diffusion limits motivate us to present the asymptotic results for this special case before doing so for the general system.

We shall begin by introducing some notation. Let $Q^n \in D_{\mathbb{R}^2}[0,\infty)$ such that $Q_1^n(t)$ and $Q_2^n(t)$ represent the number of servers in use at time t and the number of subscribers in the hold state, respectively, when there are a total of n subscribers, i.e., $Q_1^n(t)=\sum_{i=1}^n 1_{\{\text{Subscriber } i \text{ is in the on state at time } t\}}$ and $Q_2^n(t)=\sum_{i=1}^n 1_{\{\text{Subscriber } i \text{ is in the hold state at time } t\}}$. Q_1^n, Q_2^n can be obtained from the following fundamental equation[2]:

$$\begin{aligned} Q_1^n(t) &= Q_1^n(0)+N^a\left(\int_0^t 1_{\{Q_1^n(u)<k\}}(n-Q_1^n(u)-Q_2^n(u))\lambda\,du\right)+N^r\left(\int_0^t 1_{\{Q_1^n(u)<k\}}\nu Q_2^n(u)\,du\right)\\ &\quad -N^d\left(\int_0^t \mu Q_1^n(u)\,du\right),\\ Q_2^n(t) &= Q_2^n(0)+N^a\left(\int_0^t (n-Q_1^n(u)-Q_2^n(u))\lambda\,du\right)-N^a\left(\int_0^t 1_{\{Q_1^n(u)<k\}}(n-Q_1^n(u)-Q_2^n(u))\lambda\,du\right)\\ &\quad -N^r\left(\int_0^t 1_{\{Q_1^n(u)<k\}}\nu Q_2^n(u)\,du\right) \end{aligned} \tag{1}$$

for $t\ge 0$, where $N^a(\cdot)$, $N^d(\cdot)$, and $N^r(\cdot)$ are three independent, one-dimensional unit rate Poisson processes.

As we are interested in asymptotic results, we consider capacity levels of the form $k^n \equiv k_1 n + k_2\sqrt{n}$ for some $k_1\in\mathbb{R}_+$ and $k_2\in\mathbb{R}$. Define

$$m \equiv \frac{\lambda\mu}{\lambda+\mu} \tag{2}$$

and the centered and scaled process

$$\widehat{Q}^n(\cdot)=\frac{Q^n(\cdot)-(k^n,0)'}{\sqrt{n}}\le 0.$$

Further defining $q^n(\cdot)=Q^n(\cdot)/n$ and $\bar{q}=(\bar{q}_1,\bar{q}_2)'$ with $\bar{q}_1\equiv\min(k_1,\lambda/(\lambda+\mu))$ and $\bar{q}_2\equiv\lambda/(\lambda+\mu)-\bar{q}_1$, we have the following asymptotic results for this system.

Proposition 3.1. *If $q^n(0)\Rightarrow\bar{q}$, then*
(a) *$\|q^n-\bar{q}\|_T\Rightarrow 0$ for any $T\ge 0$.*
(b) *If $k_1=\lambda/(\lambda+\mu)$ and $\widehat{Q}^n(0)\Rightarrow\widehat{Q}(0)$, then $\widehat{Q}^n\Rightarrow\widehat{Q}$, where*

$$\widehat{Q}_1(t)=\widehat{Q}_1(0)-\int_0^t\big((\lambda+\mu)(\widehat{Q}_1(u)+k_2)+(\lambda-\nu)\widehat{Q}_2(u)\big)\,du+\sqrt{2m}B(t)-Y(t), \tag{3}$$

$$\widehat{Q}_2(t)=\widehat{Q}_2(0)-\int_0^t \nu\widehat{Q}_2(u)\,du+Y(t), \tag{4}$$

where B is a standard Brownian motion independent of $\widehat{Q}(0)$ and Y is the nonnegative, nondecreasing continuous process such that $\widehat{Q}_1(t)\le 0$ and $\int_0^t\widehat{Q}_1(u)\,dY(u)=0$, $\forall\, t\ge 0$ and $Y(0)=0$, and $\widehat{Q}_2(t)\ge 0$, $\forall\, t\ge 0$.

(c) *The invariant distribution of $\widehat{Q}^n(\cdot)$, $\hat{\pi}^n\Rightarrow\hat{\pi}$, where $\hat{\pi}$ is the unique invariant distribution of the diffusion process given by Equations* (3)–(4).

All the asymptotic results in the paper are structured in a manner similar to Proposition 3.1. We first state the fluid level convergence as in (a), followed by weak convergence to diffusion limits as in (b). In doing so, we first establish that the equations describing the diffusion limit (Equations (3)–(4)) have a unique strong solution. We also establish that the limiting diffusions have unique stationary distributions and establish the convergence of the invariant distributions of $\widehat{Q}^n$ to the invariant distribution of the limiting diffusions as in (c). Finally, where possible, we characterize this limiting invariant distribution.

This is a good place to point out the organization of proofs of results in this paper. The main result of this paper, Theorem 4.1, has two parts. The first part is proved in §5 and the second is proved in §6. All propositions are proved in Appendices A and B contains auxiliary results that are useful in proving the propositions.

[2] Proposition 4.1 proves the existence of a unique Q that satisfies a relation similar to that in Equation (1) for the general case in §4.1. The proof can be adapted to prove the same for this case.

The diffusion process given by Equations (3)–(4) is quite interesting. Note that there is a single one-dimensional Brownian motion driving this diffusion and the only stochasticity in the process $\widehat{Q}_2(\cdot)$ arises through the "pushing process" at the boundary. This process is fairly complicated and it renders a further investigation into its invariant distribution futile. (The reader is directed to §4 for an illustration of the difficulties that arise in the computation of the invariant distribution of this process.) However, we shall see that for the case $\nu = \lambda$, where the off and hold states are indistinguishable, we obtain a far more tractable process for which we will be able to compute the invariant distribution; this case is tackled in §4.1 in a more general setting.

Having investigated the subscriber model, we now incorporate an exogenous customer stream in addition to the subscribers in the system, allowing for the eventual analysis of whether or not subscription is a desirable option for the system designer. These customers do not subscribe but rather are considered to be one-time users who attempt to get service on arrival and are lost if no servers are available. We will again develop asymptotic results for this model and use them to obtain approximations for the invariant distributions.

4. Subscribers with a Poisson customer stream. Consider the system with a total of n subscribers as before and an exogenous stream of customers that arrive according to a Poisson process with rate λ_p^n. Let $Q^n \in D_{\mathbb{R}^3}[0,\infty)$, where $Q_1^n(\cdot)$, $Q_2^n(\cdot)$, and $Q_3^n(\cdot)$ are the processes that denote the number of servers in use by the subscribers, the number of servers in use by the exogenous customers, and the number of subscribers in the hold state, respectively. To be precise, our model results in Q^n being as follows:

$$
\begin{aligned}
Q_1^n(t) &= Q_1^n(0) + N_1^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}(n - Q_1^n(u) - Q_3^n(u))\lambda\,du\right) - N_1^d\left(\int_0^t \mu Q_1^n(u)\,du\right)\\
&\quad + N^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right),\\
Q_2^n(t) &= Q_2^n(0) + N_2^a\left(\lambda_p^n\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\,du\right) - N_2^d\left(\int_0^t \mu Q_2^n(u)\,du\right),\\
Q_3^n(t) &= Q_3^n(0) - N^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right) + N_1^a\left(\int_0^t (n - Q_1^n(u) - Q_3^n(u))\lambda\,du\right)\\
&\quad - N_1^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}(n - Q_1^n(u) - Q_3^n(u))\lambda\,du\right),
\end{aligned}
$$

for $t \geq 0$, where $N_i^a(\cdot)$ for $i = 1, 2$, $N_j^d(\cdot)$ for $j = 1, 2$, and $N^r(\cdot)$ are five independent unit rate Poisson processes. For $(x, y, z) \in \mathbb{R}_+^3$, let $\lambda^n(x, y, z) = ((n - x - z)\lambda, \lambda_p^n, 0)'$ and $\mu^n(x, y, z) = (\mu x, \mu y, 0)'$. Then, using the notation $N^u(x, y, z) = (N_1^u(x), N_2^u(y), N_3^u(z))'$ for $u = a, d$ with $N_3^a(t) = N_3^d(t) = 0$ for $t \geq 0$, the above relation can be reexpressed as

$$
\begin{aligned}
Q^n(t) &= Q^n(0) + N^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\lambda^n(Q^n(u))\,du\right) - N^d\left(\int_0^t \mu^n(Q^n(u))\,du\right)\\
&\quad + N^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right)(1, 0, -1)' + \big(0, 0, \Delta N_1^a(Q_1^n, Q_2^n, Q_3^n)(t)\big)',
\end{aligned} \tag{5}
$$

where for $A \in D_{\mathbb{R}}[0,\infty)$,

$$
\begin{aligned}
\Delta A(Q_1^n, Q_2^n, Q_3^n)(\cdot) \equiv\; & A\left(\int_0^{\cdot} (n - Q_1^n(u) - Q_3^n(u))\lambda\,du\right)\\
& - A\left(\int_0^{\cdot} 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}(n - Q_1^n(u) - Q_3^n(u))\lambda\,du\right).
\end{aligned}
$$

The following result proves the existence of a unique Q^n that satisfies Equation (5).

Proposition 4.1. *There exists a unique strong solution Q^n defined on $(\Omega, \mathcal{F}, \mathbb{P})$ to Equation* (5).

Note that Equation (5) can be rewritten as

$$
\begin{aligned}
Q^n(t) &= Q^n(0) + \int_0^t \big[\lambda^n(Q^n(u)) + \nu^n(Q^n(u)) - \mu^n(Q^n(u))\big]\,du\\
&\quad + \bar{N}^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\lambda^n(Q^n(u))\,du\right)' - \bar{N}^d\left(\int_0^t \mu^n(Q^n(u))\,du\right)'
\end{aligned}
$$

$$+\bar{N}^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right)(1,0,-1)'+(0,0,\Delta\bar{N}_1^a(Q_1^n,Q_2^n,Q_3^n))'$$
$$-\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)=k^n\}}\zeta^n(Q^n(u))\,du, \tag{6}$$

where for $(x,y,z)\in\mathbb{R}_+^3$, $\zeta^n(x,y,z)=((n-x-z)\lambda+\nu z,\lambda_p^n,-(n-x-z)\lambda+\nu z)'$ and

$$\lambda^n(x,y,z)=\big((n-x-z)\lambda,\lambda_1 n+\lambda_2\sqrt{n},0\big)', \tag{7}$$
$$\mu^n(x,y,z)=(\mu x,\mu y,0)',\qquad\text{and} \tag{8}$$
$$\nu^n(x,y,z)=(\nu z,0,-\nu z)'. \tag{9}$$

We use the convention that for any given Poisson process N, $\bar{N}$ denotes its centered version, namely $\bar{N}(t)=N(t)-t$, $t\ge 0$.

Before continuing, we introduce the following notation:

$$\kappa^n=\frac{k^n}{n}, \tag{10}$$
$$\theta^n(x,y,z)=\lambda^n(x,y,z)+\nu^n(x,y,z)-\mu^n(x,y,z), \tag{11}$$
$$M^{a,n}(t)=\bar{N}^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2(u)<k^n\}}\lambda^n(Q^n(u))\,du\right)$$
$$+\bar{N}^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right)(1,0,0)', \tag{12}$$
$$M^{d,n}(t)=\bar{N}^d\left(\int_0^t \mu^n(Q^n(u))\,du\right)+\bar{N}^r\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\nu Q_3^n(u)\,du\right)(0,0,1)', \tag{13}$$
$$\alpha^n=\frac{1}{n}(M^{a,n}-M^{d,n}), \tag{14}$$
$$\delta^n=\frac{1}{n}(0,0,\Delta\bar{N}_1^a(Q_1^n,Q_2^n,Q_3^n))', \tag{15}$$
$$S^n=\big\{(x,y,z)\in\mathbb{R}^3\mid x,y,z\ge 0,\ x+y\le k^n,\ x+z\le n\big\}, \tag{16}$$

where S^n denotes the state space of the process Q^n. Rearranging terms in Equation (6), we obtain the following characterization of $q^n(\cdot)\equiv Q^n(\cdot)/n$.

Proposition 4.2. *For $t\ge 0$, $q^n(t)$ can be written as*

$$q^n(t)=\Phi_{\kappa^n}^n(X^n)(t)\equiv X^n(t)+\int_0^t R^n(u)\,dY^n(u), \tag{17}$$

where $X^n, R^n\in D_{\mathbb{R}^3}[0,\infty)$ and $Y^n\in D_{\mathbb{R}}[0,\infty)$ are defined as

$$X^n(t)\equiv q^n(0)+\frac{1}{n}\int_0^t\theta^n(nq^n(u))\,du+\alpha^n(t)+\delta^n(t), \tag{18}$$
$$R^n(t)\equiv-\big((1-q_1^n(t)-q_3^n(t))\lambda+\nu q_3^n(t),\ \lambda_1+\lambda_2/\sqrt{n},-\big(1-q_1^n(t)-q_3^n(t)\big)\lambda+\nu q_3^n(t)\big)', \tag{19}$$
$$Y^n(t)\equiv\int_0^t 1_{\{q_1^n(u)+q_2^n(u)=\kappa^n\}}\,du. \tag{20}$$

In addition, we have

$$\hat{n}'\int_0^t R^n(u)\,dY^n(u)=-\sup_{0\le s\le t}\big(X_1^n(s)+X_2^n(s)-\kappa^n\big)^+, \tag{21}$$

where $\hat{n}=(1,1,0)'$.

Note that the process $-\int_0^t R^n(u)\,dY^n(u)$ counts the number of requests by the subscribers and the Poisson stream denied by time t. As the total number of denied requests must equal $\sup_{0\le s\le t}(X_1^n(s)+X_2^n(s)-\kappa^n)^+$, the last relation follows.

We now proceed with obtaining the asymptotic limits. To obtain meaningful limits, we shall choose $\lambda_p^n=\lambda_1 n+\lambda_2\sqrt{n}$, where $\lambda_1,\lambda_2\ge 0$ and the capacity level $k^n=k_1 n+k_2\sqrt{n}$ as before. For convenience, we shall recycle our earlier notation and define $\bar{q}=(\lambda/(\lambda+\mu),\lambda_1/\mu,0)'$. We now state the asymptotic results for this system corresponding to the fluid limit:

PROPOSITION 4.3.
(a) *If* $k_1 \geq \lambda/(\lambda+\mu)+\lambda_1/\mu$ *and* $q^n(0) \Rightarrow \bar{q}$, *then* $\|q^n - \bar{q}\|_T \Rightarrow 0$ *for any* $T \geq 0$.
(b) *If* $k_1 < \lambda/(\lambda+\mu)+\lambda_1/\mu$ *and* $q^n(0) \Rightarrow \bar{\bar{q}}$, *where* $\bar{\bar{q}}_1 = (-b-\sqrt{b^2-4ac})/2a$ *with* $a = \nu + (\mu/\lambda)(\nu-\lambda)$,

$$b = -\left(\nu(k_1+1)+\lambda_1+\frac{\mu}{\lambda}(\nu-\lambda)k_1\right),$$

and $c = k_1\nu$, $\bar{\bar{q}}_2 = k_1 - \bar{\bar{q}}_1$, *and* $\bar{\bar{q}}_3 = 1-((\lambda+\mu)/\lambda)\bar{\bar{q}}_1$, *then* $\|q^n - \bar{\bar{q}}\|_T \Rightarrow 0$ *for any* $T \geq 0$.

A point worth noting about this result is that we require $q^n(0)$ to converge to the stable point (or equilibrium point) of the fluid limit as $n \to \infty$. If this condition does not hold, i.e., if there exists a sequence along which $q^n(0)$ converges to a point different from the equilibrium point of the fluid limit, then along this sequence $q^n(t)$ shall converge to the equilibrium point only as both $n \to \infty$ and $t \to \infty$.

We now wish to study the diffusion limits for this system. To this effect, define $\widehat{Q}^n(\cdot) = \sqrt{n}(q^n(\cdot)-\bar{q}) - (k_2, 0, 0)'$. Note that if $k_1 < \lambda/(\lambda+\mu)+\lambda_1/\mu$, the correction process $\widehat{Q}^n$ is asymptotically not well-defined. If $k_1 > \lambda/(\lambda+\mu)+\lambda_1/\mu$, this case is identical to an infinite server setting and is treated in Mandelbaum and Pats [27]. Thus, we focus on the case $k_1 = \lambda/(\lambda+\mu)+\lambda_1/\mu$ and state the corresponding diffusion result:

THEOREM 4.1.
(a) *If* $k_1 = \lambda/(\lambda+\mu)+\lambda_1/\mu$, $q^n(0) \Rightarrow \bar{q}$ *and* $\widehat{Q}^n(0) \Rightarrow \widehat{Q}(0)$, *then* $\widehat{Q}^n \Rightarrow \widehat{Q}$, *the unique strong solution to*

$$\widehat{Q}_1(t) = \widehat{Q}_1(0) + \int_0^t [(\nu-\lambda)\widehat{Q}_3(u) - (\lambda+\mu)(\widehat{Q}_1(u)+k_2)]\,du + \sqrt{2m}B_1(t) - mY(t), \tag{22}$$

$$\widehat{Q}_2(t) = \widehat{Q}_2(0) + \int_0^t (\lambda_2 - \mu\widehat{Q}_2(u))\,du + \sqrt{2\lambda_1}B_2(t) - \lambda_1 Y(t), \tag{23}$$

$$\widehat{Q}_3(t) = \widehat{Q}_3(0) - \int_0^t \nu\widehat{Q}_3(u)\,du + mY(t), \tag{24}$$

where m is given by Equation (2), B_1 *and* B_2 *are two independent one-dimensional Brownian motions that are independent of* $\widehat{Q}(0)$, *and Y is the nonnegative, nondecreasing continuous process such that* $\widehat{Q}_1(t)+\widehat{Q}_2(t) \leq 0$ *and* $\int_0^t (\widehat{Q}_1(u)+\widehat{Q}_2(u))\,dY(u) = 0$, $\forall\, t \geq 0$ *and* $Y(0) = 0$, *and* $\widehat{Q}_3(t) \geq 0$, $\forall\, t \geq 0$.
(b) *The invariant distribution of* $\widehat{Q}^n(\cdot)$, $\hat{\pi}^n \Rightarrow \hat{\pi}$, *where* $\hat{\pi}$ *is the unique invariant distribution of the diffusion process* $\widehat{Q}(\cdot)$.

Noting the convergence in Theorem 4.1(b), it would be quite useful if we can characterize the invariant distribution $\hat{\pi}$. However, we shall now demonstrate using a nonrigorous argument the difficulty in estimating $\hat{\pi}$. Let us assume this invariant distribution has a density $\hat{p} \in C^{2,2,1}$. Denote the state space of the diffusion process (22)–(24) by S, i.e., $S = \{(x, y, z)\colon x+y \leq 0, z \geq 0\}$ and its boundary by $\partial S = \{(x, y, z)\colon x+y = 0\}$. Let $L = ((\nu-\lambda)z - (\lambda+\mu)(x+k_2))(\partial/\partial x) + (\lambda_2 - \mu y)(\partial/\partial y) - \nu z(\partial/\partial z) + m(\partial^2/\partial x^2) + \lambda_1(\partial^2/\partial y^2)$ denote the generator of the diffusion process; its domain is $C^{2,2,1}$. Pick any $f \in C^{2,2,1}$ with compact support such that $m(\partial f/\partial x) + \lambda_1(\partial f/\partial y) = 0$ on ∂S. Then, applying Ito's lemma to f for this diffusion process and taking expectations with respect to the invariant distribution, we obtain the condition

$$\int_S \hat{p}(v) L f(v)\,dv = 0. \tag{25}$$

Assuming sufficient regularity conditions on $\hat{p}$, repeated use of integration by parts leads us to contend that $\hat{p}$ should solve the following partial differential equation (p.d.e.):

$$m\frac{\partial^2\hat{p}}{\partial x^2} + \lambda_1\frac{\partial^2\hat{p}}{\partial y^2} + (\lambda+\mu)(x+k_2)\frac{\partial\hat{p}}{\partial x} - (\lambda_2-\mu y)\frac{\partial\hat{p}}{\partial y} + \nu z\frac{\partial\hat{p}}{\partial z} + (\lambda+2\mu+\nu)\hat{p} = 0$$

for $(x, y, z) \in S \setminus \partial S$, with the boundary condition

$$-(\lambda x + (\lambda+\mu)k_2 - \lambda_2)\hat{p} - m\frac{\partial\hat{p}}{\partial x} - \lambda_1\frac{\partial\hat{p}}{\partial y} = 0, \quad \text{for } (x, y, z) \in \partial S.$$

We are unable to solve this p.d.e. and thus cannot provide a better characterization of the invariant distribution. We shall see in the following section that for the case $\nu = \lambda$, we can actually compute the invariant distribution of the limiting diffusion process.

4.1. Special case, $\nu = \lambda$. When $\nu = \lambda$, the off and hold state are indistinguishable, which implies that we do not need to keep track of the number of subscribers in the hold state. This allows us to focus on a two-dimensional process $Q^n \in D_{\mathbb{R}^2}[0, \infty)$, where $Q_1^n(\cdot)$ and $Q_2^n(\cdot)$ denote the number of servers in use by the subscribers and the exogenous customer stream, respectively. For $(x, y) \in \mathbb{R}_+^2$, let $\lambda^n(x, y) = ((n-x)\lambda, \lambda_p^n)'$ and $\mu(x, y) = (\mu x, \mu y)'$. Then, we obtain the following characterization of Q^n:

$$Q^n(t) = Q^n(0) + N^a\left(\int_0^t 1_{\{Q_1^n(u)+Q_2^n(u)<k^n\}}\lambda(Q^n(u))\,du\right) - N^d\left(\int_0^t \mu(Q^n(u))\,du\right) \tag{26}$$

for $t \geq 0$, where we use the notation $N^z(x, y) = (N_1^z(x), N_2^z(y))$ for $z = a, d$. Defining $q^n(\cdot) = Q^n(\cdot)/n$, $\bar{q} = (\lambda/(\lambda+\mu), \lambda_1/\mu)'$ and $\widehat{Q}^n(\cdot) = \sqrt{n}(q^n(\cdot) - \bar{q}) - (k_2, 0)'$, the corresponding asymptotic results for this system are as follows.

Proposition 4.4. *If $k_1 = \lambda/(\lambda+\mu) + \lambda_1/\mu$ and $q^n(0) \Rightarrow \bar{q}$, then*
(a) *$\|q^n - \bar{q}\|_T \Rightarrow 0$ for any $T \geq 0$.*
(b) *If $\widehat{Q}^n(0) \Rightarrow \widehat{Q}(0)$, then $\widehat{Q}^n \Rightarrow \widehat{Q}$, where*

$$\widehat{Q}_1(t) = \widehat{Q}_1(0) - \int_0^t (\lambda+\mu)(\widehat{Q}_1(u) + k_2)\,du + \sqrt{2m}B_1(t) - mY(t), \tag{27}$$

$$\widehat{Q}_2(t) = \widehat{Q}_2(0) + \int_0^t (\lambda_2 - \mu\widehat{Q}_2(u))\,du + \sqrt{2\lambda_1}B_2(t) - \lambda_1 Y(t), \tag{28}$$

where B_1 and B_2 are two independent one-dimensional Brownian motions that are independent of $\widehat{Q}(0)$, and Y is the nonnegative, nondecreasing continuous process such that $\widehat{Q}_1(t) + \widehat{Q}_2(t) \leq 0$ and

$$\int_0^t (\widehat{Q}_1(u) + \widehat{Q}_2(u))\,dY(u) = 0, \quad \forall\, t \geq 0, \qquad \text{and} \qquad Y(0) = 0.$$

(c) *The invariant distribution of $\widehat{Q}^n(\cdot)$, $\widehat{\pi}^n \Rightarrow \widehat{\pi}$, where $\widehat{\pi}$ is the unique invariant distribution of the diffusion process $\widehat{Q}(\cdot)$.*

We shall now characterize the invariant distribution of this diffusion process.

Proposition 4.5.
(a) *If $\lambda_1 > 0$, the density corresponding to the invariant distribution of the diffusion process* (27)–(28) *is*

$$\hat{p}(x, y) = \begin{cases} \dfrac{\exp(-(1/2)(((x+k_2)^2(\lambda+\mu))/m + (y-\lambda_2/\mu)^2\mu/\lambda_1))}{\int_{-\infty}^{\infty}\int_{-\infty}^{-x}\exp(-(1/2)(((x+k_2)^2(\lambda+\mu))/m + (y-\lambda_2/\mu)^2\mu/\lambda_1))\,dy\,dx}, & \text{if } x+y \leq 0, \\ 0, & \text{else.} \end{cases} \tag{29}$$

(b) *If $\lambda_1 = 0$, the density corresponding to the invariant distribution of the diffusion process $\widehat{Q}_1$ given by Equation* (27) *is*:

$$\hat{p}(x) = \begin{cases} \dfrac{\exp(-(1/(2m))(\lambda+\mu)(x+k_2)^2)}{\int_{-\infty}^{-\lambda_2/\mu}\exp(-(1/(2m))(\lambda+\mu)(x+k_2)^2)\,dx}, & \text{if } x \leq -\dfrac{\lambda_2}{\mu}, \\ 0, & \text{else.} \end{cases} \tag{30}$$

4.1.1. Insensitivity of the steady state distribution of the number-in-system process in the original system. It is well-known that the steady state distribution of the number in system in the M/G/k/k loss system is independent of the service time distribution (see Gross and Harris [16], pp. 245–247). It is also known that the steady state distribution of the number in system for the on-off source model is independent of the distribution of on and off times (see Cohen [7]). We will now focus on the system with both subscribers and an exogenous pay-per-use stream and show that the steady state distribution of the number in system of the prelimit process is insensitive to the distribution of on times, off times, and the service times, i.e., it depends only on the means of these distributions, when hold times are identically distributed as off times. This insensitivity result does not hold when the hold time and off time distributions are different and when the exogenous stream does not arrive as a Poisson process.

Before stating the result, we shall introduce some notation. Let F and G denote the off time and on time distributions, respectively, with $1/\lambda$ and $1/\mu$ denoting their respective means. Note that G is also the distribution of service times for the Poisson arrivals. We will assume that $F(0) = G(0) = 0$, and F and G are strictly increasing on $[0, \infty)$. Let the total number of subscribers in the system be n, the arrival rate of the Poisson stream be λ_p^n, and the number of servers be k. Let $J^s(t)$ be a set whose elements represent the subscribers in the on state at time t, where each subscriber is numbered from 1 to n and $U = \{1, 2, \ldots, n\}$ denotes the set of all subscribers. Let $J^p(t)$ denote the number of customers from the Poisson stream in service at time t. Let $R^s(t) \in \mathbb{R}_+^n$ denote the residual times for the subscribers at time t, i.e., for subscriber $u \in J^s(t)$, $R_u^s(t)$ denotes the amount of on time remaining and for subscriber $u \notin J^s(t)$, it denotes the amount of off time remaining. Let $R^p(t)$ denote the residual times for the customers in service from the Poisson stream. It can be verified that $\{X(t) = (J^s(t), R^s(t), J^p(t), R^p(t))\colon t \geq 0\}$ is a Markov process. The number-in-system process is given by $\{(|J^s(t)|, J^p(t))\colon t \geq 0\}$, where for a set A, $|A|$ denotes its cardinality. As this process is not Markovian, its invariant distribution is not well-defined. Thus, we shall work with the steady state probability of the number-in-system process. We are now ready to state the main insensitivity result.

Proposition 4.6. *The steady state distribution of the number-in-system process* $\{(|J^s(t)|, J^p(t))\colon t \geq 0\}$ *depends on the distributions F and G only through their means.*

This insensitivity result proves that to asymptotically characterize the invariant distribution for such systems, it is sufficient to analyze a Markovian model. This allows us to circumvent the issue of obtaining the process limits for these systems, which are known to be measure valued and fairly complicated, and provides a further justification for using Markovian models.

Note that we can actually characterize precisely the invariant distribution of the prelimit process (see the proof of Proposition 4.6). Thus, one can arrive at the result in Proposition 4.5 by evaluating the limit of this invariant distribution. However, our aim is to prove the convergence of the invariant distributions for the general system of §4, which does not follow directly. Further, the methodology developed to prove this convergence allows us to extract the asymptotic invariant distribution for the process in §4.1 relatively cheaply.

5. Convergence to diffusion limits. This section is concerned with the proof of Theorem 4.1(a). If we take a closer look at Proposition 4.2, we note that $R^n(t)$, which can be thought of as a reflection direction to keep the underlying process within the feasible region, depends on the state at time t as well as on n. This dependence complicates establishing regularity properties, such as Lipschitz continuity, on the reflection map $\Phi_{\kappa^n}^n$. To circumvent this issue, we introduce an intermediate system that is defined on the same probability space and differs from the original system only in the reflection direction. For this intermediate system, we choose the reflection direction

$$\tilde{R} = -(m, \lambda_1, -m)', \tag{31}$$

which by applying Proposition 4.3(a) and the continuous mapping theorem (Theorem 2.7 in Billingsley [3]) is the weak limit of R^n, i.e., $R^n \Rightarrow \tilde{R}$.

We shall prove that this intermediate system is indistinguishable from the system under consideration at both the fluid and diffusion scale. It will then be sufficient to derive the diffusion limit for the intermediate system alone. However, to prove the equivalence of these two systems, we shall require the existence of a weak limit for the intermediate system. Thus, we shall first obtain the diffusion limit for the intermediate system and then establish that this intermediate system is indistinguishable from the system under consideration at the fluid and diffusion scale. This will complete the proof.

We shall now define the intermediate system. We shall use the same notation for the intermediate system as we did for the original system, albeit with a modifier $(\tilde{\cdot})$ to differentiate between the two, i.e., $\tilde{Q}^n \in D_{\mathbb{R}^3}[0, \infty)$ represents the number-in-system process for this system. Before defining the intermediate system, analogous to Equations (17) and (21) we introduce a mapping $\tilde{\Phi}_a\colon D_{\mathbb{R}^3}[0, \infty) \to D_{\mathbb{R}^3}[0, \infty)$ such that

$$\tilde{\Phi}_a(X) \equiv X + \tilde{R}\tilde{Y}, \tag{32}$$

where $\tilde{R}$ is given by Equation (31) and for $t \geq 0$

$$\tilde{Y}(t) = -\frac{\sup_{0\leq s\leq t}(X_1(s) + X_2(s) - a)^+}{\hat{n}'\tilde{R}}$$

with $\hat{n} = (1, 1, 0)'$. Note that the mapping $\tilde{\Phi}_a$ is independent of n. Further, given $X \in D_{\mathbb{R}^3}[0, \infty)$, $\tilde{Y}$ is well-defined, and thus the mapping $\tilde{\Phi}_a$ is well-defined for each a.

We now define the number-in-system process for the intermediate system as follows:

$$\tilde{Q}^n(t) = \tilde{\Phi}_{k^n}(\tilde{U}^n)(t), \tag{33}$$

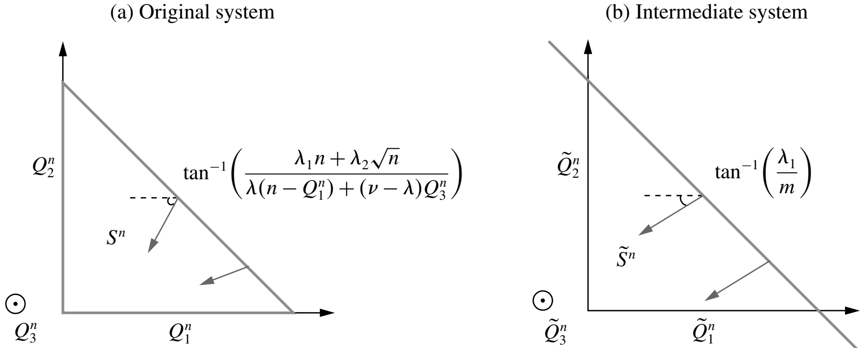

FIGURE 2. Reflection directions in the two systems.

where

$$\tilde{U}^n(t) = Q^n(0) + N^a\left(\int_0^t \tilde{\lambda}^n(\tilde{Q}^n(u))du\right) - N^d\left(\int_0^t \tilde{\mu}^n(\tilde{Q}^n(u))\,du\right)$$
$$+ N^r\left(\int_0^t (n \wedge \tilde{Q}_3^n(u))\nu\,du\right)(1, 0, -1)', \tag{34}$$

where for $(x, y, z) \in \mathbb{R}^3$, we have

$$\tilde{\lambda}^n(x, y, z) = ([(n-x-z)\wedge n]\lambda, \lambda_p^n, 0), \qquad \text{and} \tag{35}$$
$$\tilde{\mu}^n(x, y, z) = (x^+\mu, (y \wedge (k_1 + |k_2|)n)^+\mu, 0)'. \tag{36}$$

This system is similar to our original system as defined in Equation (5) and has the same initial state, i.e., $\tilde{Q}^n(0) = Q^n(0)$. However, unlike our original system, this system has an unbounded state space and the components of $\tilde{Q}^n$ can take negative values as well. Thus, this intermediate system has no "real" interpretation. Note that we modify the rate functions λ^n and μ^n so that the process is "well-behaved." This is useful in obtaining the uniqueness of the solution of Equations (33)–(34) in the following result:

PROPOSITION 5.1. *There exists a unique strong solution $\tilde{Q}^n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ to Equations (33)–(34). Further, the state space of $\tilde{Q}^n$ is $\tilde{S}^n = \{(x, y, z)\colon x + y \le k^n,\ x + z \le n,\ z \ge 0\}$.*

Figure 2 provides a two-dimensional cross-section of the state space along with the reflection direction of the number-in-system process in the two systems.

We now write out the dynamics for the scaled process $\tilde{Q}^n/n$ as follows.

$$\tilde{q}^n(t) \equiv \frac{\tilde{Q}^n(\cdot)}{n} = \tilde{\Phi}_{\kappa^n}(\tilde{X}^n)(t) = \tilde{X}^n(t) + \tilde{R}\tilde{Y}^n(t), \tag{37}$$

where

$$\tilde{X}^n(t) = \tilde{q}^n(0) + \frac{1}{n}\int_0^t \tilde{\theta}^n(n\tilde{q}^n(u))\,du + \tilde{\alpha}^n(t), \tag{38}$$

$$\tilde{Y}^n(t) = -\frac{\sup_{0\le s\le t}(\tilde{X}_1^n(s) + \tilde{X}_2^n(s) - \kappa^n)^+}{\hat{n}'\tilde{R}}, \tag{39}$$

and we have the following analog to Equations (11)–(15).

$$\tilde{\theta}^n(x, y, z) = \tilde{\lambda}^n(x, y, z) + \tilde{\nu}^n(x, y, z) - \tilde{\mu}^n(x, y, z), \tag{40}$$

$$\tilde{\nu}^n(x, y, z) = ((z \wedge n)\nu, 0, -(z \wedge n)\nu)', \tag{41}$$

$$\tilde{M}^{a,n}(t) = \bar{N}^a\left(\int_0^t \tilde{\lambda}^n(n\tilde{q}^n(u))\,du\right) + \bar{N}^r\left(\int_0^t n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right)(1, 0, 0)', \tag{42}$$

$$\tilde{M}^{d,n}(t) = \bar{N}^d\left(\int_0^t \tilde{\mu}^n(n\tilde{q}^n(u))\,du\right) + \bar{N}^r\left(\int_0^t n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right)(0, 0, 1)', \tag{43}$$

$$\widetilde{M}^n = \frac{1}{\sqrt{n}}(\widetilde{M}^{a,n}(t) - \widetilde{M}^{d,n}(t)), \tag{44}$$

$$\widetilde{\alpha}^n = \frac{1}{n}(\widetilde{M}^{a,n} - \widetilde{M}^{d,n}). \tag{45}$$

We now obtain the asymptotic limits for the intermediate system. Defining $Q^{*n}(\cdot) = \sqrt{n}(\widetilde{q}^n(\cdot) - \bar{q}) - (k_2, 0, 0)'$, where $\bar{q} = (\lambda/(\lambda+\mu), \lambda_1/\mu, 0)'$, we have the following asymptotic results.

Proposition 5.2. *If $k_1 = \lambda/(\lambda+\mu) + \lambda_1/\mu$ and $\widetilde{q}^n(0) \Rightarrow \bar{q}$, then*
(a) $\|\widetilde{q}^n - \bar{q}\|_T \Rightarrow 0$ *for any* $T \geq 0$.
(b) *If $Q^{*n}(0) \Rightarrow Q^*(0)$, then $Q^{*n} \Rightarrow Q^*$, the unique strong solution to*

$$Q^*(t) = \widetilde{\Phi}_0(Z^*)(t), \tag{46}$$

$$Z_1^*(t) = Q_1^*(0) + \int_0^t [(\nu - \lambda)Q_3^*(u) - (\lambda+\mu)(Q_1^*(u) + k_2)]\,du + \sqrt{2m}B_1(t), \tag{47}$$

$$Z_2^*(t) = Q_2^*(0) + \int_0^t (\lambda_2 - \mu Q_2^*(u))\,du + \sqrt{2\lambda_1}B_2(t), \tag{48}$$

$$Z_3^*(t) = Q_3^*(0) - \int_0^t \nu Q_3^*(u)\,du, \tag{49}$$

where B_1 and B_2 are two independent standard Brownian motions that are independent of $Q^*(0)$.

We shall now prove that the intermediate system is asymptotically equivalent to the original system at the $\sqrt{n}$-scale, which implies that both these systems have identical diffusion limits. Although, this equivalence result holds in far more generality, we shall prove it only for the case we are interested in.

Proposition 5.3. *For any $T > 0$, if $k_1 = \lambda/(\lambda+\mu) + \lambda_1/\mu$ and $q^n(0) \Rightarrow \bar{q}$, then $\|\widehat{Q}^n - Q^{*n}\|_T \Rightarrow 0$.*

This, along with Proposition 5.2, completes the proof of Theorem 4.1(a).

6. Convergence of invariant distributions. The subject of this section is the proof of Theorem 4.1(b). We shall prove the convergence of the invariant distributions as well as uniqueness of the invariant distribution of the limiting diffusion process. The proof employs a Lyapunov function argument as in Gamarnik and Zeevi [15]. However, as the processes we consider have state-dependent drift and Gamarnik and Zeevi [15] prove the limit interchange for Jackson networks (with state-independent drift), their main results are not directly applicable here. We shall begin with the following definitions for a Markov chain $(Q(t)\colon t \geq 0)$ with a complete, metrizable state space $\mathscr{S}$ as in Gamarnik and Zeevi [15]:

Definition 6.1. A function $f\colon \mathscr{S} \to \mathbb{R}_+$ is said to be a Lyapunov function for the process Q with drift size parameter $-\gamma$, where $\gamma > 0$, drift time parameter $t_0 > 0$, and exception parameter K if

$$\sup_{\{x \in \mathscr{S}: f(x) > K\}} \{\mathbb{E}_x f(Q(t_0)) - f(x)\} \leq -\gamma.$$

A function $f\colon \mathscr{S} \to \mathbb{R}_+$ is said to be a geometric Lyapunov function with a geometric drift size $0 < \gamma < 1$, drift time $t_0 > 0$, and exception parameter K if

$$\sup_{\{x \in \mathscr{S}: f(x) > K\}} \{(f(x))^{-1}\mathbb{E}_x f(Q(t_0))\} \leq \gamma.$$

Define $\phi(t) = \sup_{x \in \mathscr{S}}(f(x))^{-1}\mathbb{E}_x f(Q(t))$ and for any given $\beta > 0$,

$$L_1(\beta, t) \equiv \sup_{x \in \mathscr{S}} \mathbb{E}_x\big[\exp(\beta(f(Q(t)) - f(x)))\big],$$
$$L_2(\beta, t) \equiv \sup_{x \in \mathscr{S}} \mathbb{E}_x\big[(f(Q(t)) - f(x))^2 \exp(\beta(f(Q(t)) - f(x))^+)\big],$$

for $t \geq 0$.

Let $\widehat{q}^n = n(\bar{q} - q^n) - (k_2\sqrt{n}, 0, 0)'$. Recall that $q^n = Q^n/n$ and $\bar{q} = (\lambda/(\lambda+\mu), \lambda_1/\mu, 0)'$. Differing from Gamarnik and Zeevi [15], we shall prove that the function $f(z) = e'|z|$ is a Lyapunov function, where $|z| = (|z_1|, |z_2|, |z_3|)'$. The fact that the centered and scaled process can take negative values leads us to this choice. We shall use $\mathscr{S} = \mathbb{R}^2 \times \mathbb{R}_+$ equipped with the sup norm.

PROPOSITION 6.1. *For all sufficiently large n, the function $f(z) = e'|z|$ is a Lyapunov function for the process $\hat{q}^n$ with drift size parameter $-\sqrt{n}$, drift time parameter t_0, and exception parameter $c_0\sqrt{n}$. In addition, there exits $\beta_0 > 0$ such that the following hold.*

$$\limsup_{n\to\infty} L_1(\beta_0/\sqrt{n}, t_0) < \infty, \tag{50}$$

$$\limsup_{n\to\infty} \frac{1}{n} L_2(\beta_0/\sqrt{n}, t_0) < \infty. \tag{51}$$

We now establish the required tightness of the invariant distributions $\{\hat{\pi}^n\}$. Note that the existence and uniqueness of these invariant distributions follows from the fact that their corresponding Markov chains are irreducible and are defined on a finite state space, and thus are positive recurrent.

PROPOSITION 6.2. *There exist constants C_1, c_1 such that for all sufficiently large n, the sequence of invariant distributions $\hat{\pi}^n$ satisfies*

$$P_{\hat{\pi}^n}(n^{-1/2}e'|\hat{q}^n(0)| > s) \le C_1 \exp(-c_1 s),$$

for all $s > 0$.

A consequence of this proposition is the following result.

PROPOSITION 6.3. *Let $\hat{q}^n(0)/\sqrt{n}$ be distributed according to $\hat{\pi}^n$. Then, the sequence of random vectors $\{\hat{q}^n(0)/\sqrt{n}\}$ is tight.*

The following result combined with Proposition 6.3 completes the proof of Theorem 4.1(b).

PROPOSITION 6.4. *Any weak limit of $\{\hat{\pi}^n\}$ is invariant for the diffusion process $\hat{Q}(\cdot)$ given by Equations* (22)–(24). *Further, $\hat{Q}(\cdot)$ has a unique invariant distribution.*

7. Discussion. In this paper, we introduce a subscriber-based method of modeling customers, wherein each customer is associated with a three-state Markov chain that determines the subscriber's behavior. This model is descriptively quite different from a Poisson arrival model and this difference persists even as the number of customers grows without bound. To illustrate this effect, we compare the two systems: one with customers modeled as subscribers and the other with customers arriving as a Poisson stream using the denial rate as our performance metric. The denial rate is the steady state at which customer requests are denied. We shall vary the number of servers in this study and set the retrial rate of the subscribers to $\nu = \lambda$. For a fair comparison of the two streams, we shall equate the offered load by the customers. Hence, we shall choose the arrival rate of the Poisson stream to be $(\lambda\mu/(\lambda+\mu))n$.

We shall set the number of servers in each system at $(\lambda/(\lambda+\mu))n + k_2\sqrt{n}$ and then vary k_2. This is because setting $k_1 < \lambda/(\lambda+\mu)$ creates an $O(n)$ difference in the two systems, with the denial rate in the subscriber model being significantly higher. We choose $\lambda = \mu = 2$ for our computations. The limiting denial rates for the two systems can be computed as a limit of $Y^n(t)/t$ as $t \to \infty$ as in Proposition 6 in Randhawa [33]. Figure 3 plots the limiting denial rates as a function of k_2.

We observe that if k_2 is small, the denial rate is lower in the exogenous arrival model: however, as k_2 increases, the subscriber model has a lower denial rate, which demonstrates the difference in the two models. Figure 3 can

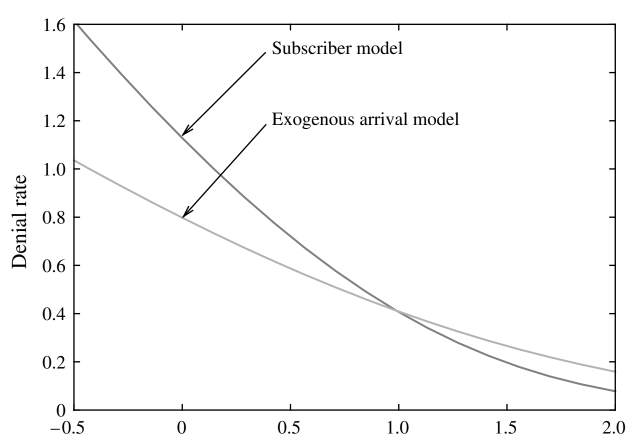

FIGURE 3. Comparison of denial rates in the subscriber and exogenous arrival models.

be explained loosely as follows. As the number of subscribers that are on increases, the number of subscribers attempting decreases. Consequently, when the capacity is large, the denial rate seen by the subscribers is smaller than that seen by the exogenous stream whose attempt rate is independent of the state of the system. When the capacity is small, the number of on subscribers is small due to server unavailability. Consequently, the number of subscribers attempting increases.

Appendix A. Auxiliary results and lemmas.

Lemma A.1 (Strong Approximation, Lemma 3.1, Kurtz [25]). *A standard (rate 1) Poisson process $N(t)$ can be realized on the same probability space as a standard Brownian motion $B(t)$ in such a way that the positive random variable X given by*

$$X \equiv \sup_{t\geq 0} \frac{|N(t)-t-B(t)|}{\log(2\vee t)} \tag{A1}$$

satisfies $\mathbb{E}e^{\eta X}<\infty$, $\forall\,\eta>0$ sufficiently small. In particular, $\mathbb{E}X<\infty$.

Lemma A.2 (cf. Lemma 3.2 in Kurtz [25], p. 14 of McKean [29]). *For any standard Brownian motion B and any $\epsilon>0$, $n\in\mathbb{N}$, and $T>0$*

$$\overline{M} \equiv \sup_{u,\,v\leq n\epsilon T} \frac{|B(u)-B(v)|}{\sqrt{|u-v|(1+\log(n\epsilon T/|u-v|))}} < \infty, \quad a.s. \tag{A2}$$

Further, the distribution of $\overline{M}$ does not dependent on n.

Lemma A.3 (Theorem 6, Gamarnik and Zeevi [15]). *Suppose a Markov process $Q(\cdot)$ defined on a complete, metrizable state space $\mathcal{S}$ possesses an invariant distribution π, and suppose f is a Lyapunov function with parameters γ, t_0, K. Further, assume there exists $\beta>0$ such that*

$$\beta L_2(\beta, t_0)\leq \gamma. \tag{A3}$$

Then, $e^{\beta f(\cdot)}$ is a geometric Lyapunov function with geometric drift size parameter $(1-\gamma\beta/2)$, drift time parameter t_0, and exception parameter $e^{\beta K}$. Consequently, for every $s>K$

$$P_\pi(f(Q(0))>s)\leq (1-\gamma\beta/2)^{-1}L_1(\beta,t_0)e^{-\beta(s-K)}.$$

Lemma A.4. *If Q is a Markov process with an invariant distribution π, for any sequence of time instants $\{t^m\}$ such that $t^m\uparrow\infty$, π is invariant for the discrete time Markov chain $Q^m\equiv Q(t^m)$. Further, if Q^m is ψ-irreducible, then π is the unique invariant distribution of Q.*

Proof. The first part of the claim follows trivially. For the second part, we apply Proposition 10.1.1 and Theorem 10.0.1 in Meyn and Tweedie [30] to conclude that π is the unique invariant distribution of Q^m. This immediately implies that π must be the unique invariant distribution of Q. □

Lemma A.5. *For a function $f\colon \mathbb{R}\to\mathbb{R}$, if $|f(x)-f(y)|\leq C|x-y|$ for $x, y\in\mathbb{R}$ and some constant $C>0$, then for any $a>0$, $|f(x)\wedge a - f(y)\wedge a|\leq C|x-y|$.*

Appendix B. Proof of results. For brevity, we omit the proofs of Propositions 3.1 and 4.4 as these results are a special case of Proposition 4.3 and Theorem 4.1 in §4 and can be proved in an identical fashion. Proposition 3.1 follows from Theorem 4.1 by setting $\lambda_1=\lambda_2=0$ and Proposition 4.4 follows by setting $Q_3^n(\cdot)=0$, as the system is equivalent to one in which subscribers turn off immediately when denied service.

B.1. Proof of results in §4.

B.1.1. Proof of Proposition 4.1. For convenience, we shall fix $n>0$ and drop it from the notation. We shall argue on the same lines as in Theorem 9.2 in Mandelbaum et al. [28]. Let $\overline{Q}^l=\{\overline{Q}^l(t)\colon t\geq 0\}$ for $l=0,1,\ldots,$ where $\overline{Q}^0(t)=Q(0)$ for $t\geq 0$ and for $l>0$, $\overline{Q}^l$ is given by

$$\begin{aligned}\overline{Q}^l(t) \equiv{}& Q(0)+N^a\bigg(\int_0^{t\wedge T^l} 1_{\{\overline{Q}_1^{l-1}(u)+\overline{Q}_2^{l-1}(u)<k\}}\lambda(\overline{Q}^{l-1}(u))\,du\bigg)-N^d\bigg(\int_0^{t\wedge T^l}\mu(\overline{Q}^{l-1}(u))\,du\bigg)\\ &+N^r\bigg(\int_0^{t\wedge T^l} 1_{\{\overline{Q}_1^{l-1}(u)+\overline{Q}_2^{l-1}(u)<k\}}\nu\overline{Q}_3^{l-1}(u)\,du\bigg)(1,0,-1)'\\ &+(0,0,\Delta N_1^a(\overline{Q}_1^{l-1},\overline{Q}_2^{l-1},\overline{Q}_3^{l-1})(t\wedge T^l))',\end{aligned} \tag{B1}$$

where the time for l events to occur T^l is given by

$$T^l = \inf\bigg\{t\colon N^a\bigg(\int_0^t 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\lambda(\bar{Q}^{l-1}(u))\,du\bigg)'e + N^d\bigg(\int_0^t \mu(\bar{Q}^{l-1}(u))\,du\bigg)'e$$
$$+ N^r\bigg(\int_0^t 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\nu\bar{Q}_3^{l-1}(u)\,du\bigg) + \Delta N_1^a(\bar{Q}_1^{l-1}, \bar{Q}_2^{l-1}, \bar{Q}_3^{l-1})(t) = l\bigg\}.$$

Note that given $\bar{Q}^{l-1}$, $\bar{Q}^l$ is well-defined for each l.

To complete the proof, we only need to show

(a) $\bar{Q}^l(t) = \bar{Q}^{l-1}(t)$ for all $0 \le t < T^l$.

(b) $\lim_{l\to\infty} T^l = \infty$ a.s.

The solution to Equation (5) Q can be constructed by setting

$$Q(t) = \bar{Q}^{l-1}(t) \quad \text{for all } 0 \le t < T^l.$$

Uniqueness then follows by using induction on r and by noting that Equation (B1) implies that the uniqueness of $\bar{Q}^l$ follows from uniqueness of $\bar{Q}^{l-1}$.

Note that the second claim follows trivially as $\bar{Q}^l$ is bounded. We prove the first claim using induction on l. The case $l = 1$ holds trivially as $\bar{Q}^1(t) = \bar{Q}^0(t) = Q(0)$ for $t < T^1$. Now, assume that $\bar{Q}^l(t) = \bar{Q}^{l-1}(t)$ for all $0 \le t < T^l$ for some $r \in \mathbb{N}$, then

$$N^a\bigg(\int_0^{t\wedge T^l} 1_{\{\bar{Q}_1^l(u)+\bar{Q}_2^l(u)<k\}}\lambda(\bar{Q}^l(u))\,du\bigg) = N^a\bigg(\int_0^{t\wedge T^l} 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\lambda(\bar{Q}^{l-1}(u))\,du\bigg)$$
$$N^d\bigg(\int_0^{t\wedge T^l} \mu(\bar{Q}^l(u))\,du\bigg) = N^d\bigg(\int_0^{t\wedge T^l} \mu(\bar{Q}^{l-1}(u))\,du\bigg),$$
$$N^r\bigg(\int_0^{t\wedge T^l} 1_{\{\bar{Q}_1^l(u)+\bar{Q}_2^l(u)<k\}}\nu\bar{Q}_3^l(u)\,du\bigg) = N^r\bigg(\int_0^{t\wedge T^l} 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\nu\bar{Q}_3^{l-1}(u)\,du\bigg),$$
$$\Delta N_1^a(\bar{Q}_1^l, \bar{Q}_2^l, \bar{Q}_3^l)(t\wedge T^l) = \Delta N_1^a(\bar{Q}_1^{l-1}, \bar{Q}_2^{l-1}, \bar{Q}_3^{l-1})(t\wedge T^l)$$

for $0 \le t \le T^l$, which implies that $\bar{Q}^{l+1}(t) = \bar{Q}^l(t)$ for $0 \le t \le T^l$.

For $T^l \le t < T^{l+1}$,

$$N^a\bigg(\int_0^t 1_{\{\bar{Q}_1^l(u)+\bar{Q}_2^l(u)<k\}}\lambda(\bar{Q}_1^l(u))\,du\bigg) = N^a\bigg(\int_0^{T^l} 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\lambda(\bar{Q}_1^{l-1}(u))\,du\bigg)$$
$$N^d\bigg(\int_0^t \mu(\bar{Q}^l(u))\,du\bigg) = N^d\bigg(\int_0^{T^l} \mu(\bar{Q}^{l-1}(u))\,du\bigg).$$

Similarly,

$$N^r\bigg(\int_0^t 1_{\{\bar{Q}_1^l(u)+\bar{Q}_2^l(u)<k\}}\nu\bar{Q}_3^l(u)\,du\bigg) = N^r\bigg(\int_0^{t\wedge T^l} 1_{\{\bar{Q}_1^{l-1}(u)+\bar{Q}_2^{l-1}(u)<k\}}\nu\bar{Q}_3^{l-1}(u)\,du\bigg),$$
$$\Delta N_1^a(\bar{Q}_1^l, \bar{Q}_2^l, \bar{Q}_3^l)(t) = \Delta N_1^a(\bar{Q}_1^{l-1}, \bar{Q}_2^{l-1}, \bar{Q}_3^{l-1})(t\wedge T^l).$$

Hence, $\bar{Q}^{l+1}(t) = \bar{Q}^l(t) = \bar{Q}^l(T^l)$ for $T^l \le t < T^{l+1}$. Therefore, $\bar{Q}^{l+1}(t) = \bar{Q}^l(t)$ for $0 \le t < T^{l+1}$, and the inductive hypothesis holds. ☐

B.1.2. Proof of Proposition 4.3(a). We shall prove the result for the case $k_1 = \lambda/(\lambda+\mu) + \lambda_1/\mu$; the proof for the case $k_1 > \lambda/(\lambda+\mu) + \lambda_1/\mu$ follows in a similar fashion.

Using Proposition 4.2 and the fact that $\bar{q} = (\lambda/(\lambda+\mu), \lambda_1/\mu, 0)'$ and $k_1 = \lambda/(\lambda+\mu) + \lambda_1/\mu$, we obtain

$$(q^n(0)-\bar{q}) + \bigg(0, \frac{\lambda_2}{\sqrt{n}}t, 0\bigg)' + \int_0^t \theta(q^n(u))\,du + \alpha^n(t) + \delta^n(t) - (1,1,0)' \sup_{0\le s\le t}(X_1^n(s)+X_2^n(s)-\kappa^n)^+$$
$$\le (q^n(t)-\bar{q})$$
$$\le (q^n(0)-\bar{q}) + \bigg(0, \frac{\lambda_2}{\sqrt{n}}t, 0\bigg)' + \int_0^t \theta(q^n(u))\,du + \alpha^n(t) + \delta^n(t) + (0,0,1)' \sup_{0\le s\le t}(X_1^n(s)+X_2^n(s)-\kappa^n)^+, \quad \text{(B2)}$$

where

$$\theta(x,y,z)=(\lambda-(\lambda+\mu)x+(\nu-\lambda)z,\lambda_1-\mu y,-\nu z)',\quad \text{for } (x,y,z)\in\mathbb{R}^3. \tag{B3}$$

Noting that $\theta(\bar{q})=0$, we have

$$\begin{aligned}
&\|q^n-\bar{q}\|_t\\
&\le \|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\theta(q^n(u))-\theta(\bar{q})|\,du+\|\alpha^n\|_t+\|\delta^n\|_t+\sup_{0\le s\le t}(X_1^n(s)+X_2^n(s)-\kappa^n)^+\\
&\le \|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\theta(q^n(u))-\theta(\bar{q})|\,du+\|\alpha^n\|_t+\|\delta^n\|_t\\
&\quad+[\|X_1^n+X_2^n-(\bar{q}_1+\bar{q}_2)\|_t+|\kappa^n-(\bar{q}_1+\bar{q}_2)|]\\
&\le \|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\theta(q^n(u))-\theta(\bar{q})|\,du+\|\alpha^n\|_t+\|\delta^n\|_t+2\|X^n-\bar{q}\|_t+\frac{k_2}{\sqrt{n}}.
\end{aligned} \tag{B4}$$

Note that θ defined in Equation (B3) is Lipschitz continuous with constant $K_1=2\lambda+\mu+\nu$.

Thus, we can rewrite Equation (B4) as

$$\begin{aligned}
\|q^n-\bar{q}\|_t\le{}&\|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|q^n-\bar{q}\|_u\,du+\|\alpha^n\|_t+\|\delta^n\|_t\\
&+2\|X^n-\bar{q}\|_t+\frac{k_2}{\sqrt{n}}.
\end{aligned}$$

We can repeat the argument in Equations (B2) and (B4) for X^n to obtain

$$\|X^n-\bar{q}\|_t\le\|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|q^n-\bar{q}\|_u\,du+\|\alpha^n\|_t+\|\delta^n\|_t.$$

Thus, we have

$$\|q^n-\bar{q}\|_t\le 3\left(\|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|q^n-\bar{q}\|_u\,du+\|\alpha^n\|_t+\|\delta^n\|_t+\frac{k_2}{\sqrt{n}}\right).$$

Using Gronwall's lemma, we obtain

$$\|q^n-\bar{q}\|_T\le D_1\left(\|q^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}T+\|\alpha^n\|_T+\|\delta^n\|_T+\frac{k_2}{\sqrt{n}}\right)e^{D_2T}, \tag{B5}$$

where D_1 and D_2 are constants. Noting that $q^n(0)\Rightarrow\bar{q}$ and $\bar{q}$ is a deterministic constant, we have $\|q^n(0)-\bar{q}\|\xrightarrow{p}0$ and thus $\|q^n(0)-\bar{q}\|\Rightarrow 0$. Hence, we only need to show that $\|\alpha^n\|_T$, $\|\delta^n\|_T\Rightarrow 0$ to complete the result. Note that we have the following bound on the rate functions: $\sup_{x\in S^n}\max_i\lambda_i^n(x)+\sup_{x\in S^n}\max_i\mu_i^n(x)+\sup_{x\in S^n}\max_i\nu_i^n(x)\le Cn$, where $C<\infty$ is some constant. Thus, for $i=1,2$, we have $\|\bar{N}_i^a(\int_0^\cdot\lambda^n(nq^n(u))\,du)\|_T\le\|\bar{N}_i^a(Cn\cdot)\|_T$. Using the functional law of large numbers, as $\|(1/n)\bar{N}_i^a(n\cdot)\|_T\Rightarrow 0$, we have $\|(1/n)\bar{N}_i^a(Cn\cdot)\|_T\Rightarrow 0$. Thus, $(1/n)\|\bar{N}_i^a(\int_0^\cdot\lambda^n(nq^n(u))\,du)\|_T\Rightarrow 0$. Similarly,

$$\frac{1}{n}\left\|\bar{N}_i^d\left(\int_0^\cdot\mu^n(nq^n(u))\,du\right)\right\|_T,\qquad\frac{1}{n}\left\|\bar{N}^r\left(\int_0^\cdot n\tilde{q}_3^n(u)\nu\,du\right)\right\|_T\;\Rightarrow\;0.$$

Hence, $\|\alpha^n\|_T$, $\|\delta^n\|_T\Rightarrow 0$. $\square$

B.1.3. Proof of Proposition 4.3(b). The Skorohod representation theorem allows us to construct random variables $r^n(0)$ defined on a different probability space $(\widehat{\Omega}, \widehat{\mathscr{F}}, \widehat{\mathbb{P}})$ such that $r^n(0) \overset{d}{=} q^n(0)$ and $r^n(0) \to \bar{\bar{q}}$ a.s. Let $nr^n(\cdot)$ be the analog of $Q^n(\cdot)$ on this space, i.e.,

$$
\begin{aligned}
nr_1^n(t) &= nr_1^n(0) + \widehat{N}_1^a\left(\int_0^t 1_{\{nr_1^n(u)+nr_2^n(u)<k^n\}}(n - nr_1^n(u) - nr_3^n(u))\lambda\, du\right) - \widehat{N}_1^d\left(\int_0^t \mu n r_1^n(u)\, du\right)\\
&\quad + \widehat{N}^r\left(\int_0^t 1_{\{nr_1^n(u)+nr_2^n(u)<k^n\}}\nu n r_3^n(u)\, du\right),\\
nr_2^n(t) &= nr_2^n(0) + \widehat{N}_2^a\left(\lambda_p^n\int_0^t 1_{\{nr_1^n(u)+nr_2^n(u)<k^n\}}\, du\right) - \widehat{N}_2^d\left(\int_0^t \mu n r_2^n(u)\, du\right),\\
nr_3^n(t) &= nr_3^n(0) - \widehat{N}^r\left(\int_0^t 1_{\{nr_1^n(u)+nr_2^n(u)<k^n\}}\nu n r_3^n(u)\, du\right) + \widehat{N}_1^a\left(\int_0^t (n - nr_1^n(u) - nr_3^n(u))\lambda\, du\right)\\
&\quad - \widehat{N}_1^a\left(\int_0^t 1_{\{nr_1^n(u)+nr_2^n(u)<k^n\}}(n - nr_1^n(u) - nr_3^n(u))\lambda\, du\right),
\end{aligned}
$$

for $t \geq 0$, where $\widehat{N}_i^a(\cdot)$ for $i = 1, 2$, $N_j^d(\cdot)$ for $j = 1, 2$, and $\widehat{N}^r(\cdot)$ are five independent unit rate Poisson processes defined on $(\widehat{\Omega}, \widehat{\mathscr{F}}, \widehat{\mathbb{P}})$. Then, $r^n(\cdot) \overset{d}{=} q^n(\cdot)$.

For the following analysis, pick a sample path $\omega \in \widehat{\Omega}$ such that $r^n(0) \to \bar{\bar{q}}$. We shall prove that $r^n \to \bar{\bar{q}}$. We use an algebraic rearrangement analogous to Proposition 4.2 to obtain the following relation for $r^n(\cdot)$:

$$
\begin{aligned}
r_1^n(t) &= r_1^n(0) + \int_0^t [\lambda - (\lambda+\mu)r_1^n(u) + (\nu-\lambda)r_3^n(u)]\, du + \hat{\alpha}_1^n(t)\\
&\quad - \int_0^t ((1 - r_1^n(u))\lambda + (\nu-\lambda)r_3^n(u))1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du, \qquad \text{(B6)}
\end{aligned}
$$

$$
r_2^n(t) = r_2^n(0) + \int_0^t [\lambda_1 - \mu r_2^n(u)]\, du + \frac{\lambda_2}{\sqrt{n}}t + \hat{\alpha}_2^n(t) - \int_0^t\left(\lambda_1 + \frac{\lambda_2}{\sqrt{n}}\right)1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du, \qquad \text{(B7)}
$$

$$
r_3^n(t) = r_3^n(0) - \int_0^t \nu r_3^n(u)\, du + \hat{\delta}_3^n(t) + \int_0^t ((1 - r_1^n(u))\lambda + (\nu-\lambda)r_3^n(u))1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du, \qquad \text{(B8)}
$$

where $\hat{\alpha}^n$ and $\hat{\delta}^n$ are defined analogous to Equations (14) and (15), respectively. Note that we must have

$$
\int_0^t (r_1^n(u) + r_2^n(u) - \kappa^n)1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du = 0 \quad \text{for all } t \geq 0. \qquad \text{(B9)}
$$

Define:

$$
y^n(t) \equiv \int_0^t 1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du, \quad \text{for } t \geq 0. \qquad \text{(B10)}
$$

Thus, $y^n \in C_{\mathbb{R}}[0,\infty)$, the space of $\mathbb{R}$-valued continuous functions and we can rewrite Equation (B9) as

$$
\int_0^t (r_1^n(u) + r_2^n(u) - \kappa^n)\, dy^n(u) = 0, \quad \text{for all } t \geq 0. \qquad \text{(B11)}
$$

Further, define $\tilde{y}^n(t) \equiv \int_0^t ((1 - r_1^n(u))\lambda + (\nu-\lambda)r_3^n(u))1_{\{r_1^n(u)+r_2^n(u)=\kappa^n\}}\, du \in C_{\mathbb{R}}[0,\infty)$, and note that for $0 \leq s < t$, we have:

$$
|\tilde{y}^n(t) - \tilde{y}^n(s)| \leq \int_s^t \|(1 - r_1^n(\cdot))\lambda + (\nu-\lambda)r_3^n(\cdot)\|_u\, du \leq (2\lambda+\nu)|t-s|
$$

as $0 \leq r_1^n(u), r_3^n(u) \leq 1$ for all $u \geq 0$. This implies that $\{\tilde{y}^n\colon [0,T] \to [0,T]\}$ is a family of equicontinuous and pointwise bounded functions. Thus, $\{y^n\colon [0,T] \to [0,T]\}$ and $\{\tilde{y}^n\colon [0,T] \to [0,T]\}$ are two families of equicontinuous and pointwise bounded functions defined on a separable space. Using the Ascoli-Arzelá theorem (see Corollary 41 in §7.10 of Royden [37]), we obtain the existence of a uniformly convergent subsequence for each family, which we denote by $\{y^{n_k}\}$ and $\{\tilde{y}^{n_k}\}$, respectively. Let y and $\tilde{y}$ denote the corresponding limiting functions (which exist as the space $[0,T]$ equipped with the uniform metric is complete), i.e., $y^{n_k} \to y$ and $\tilde{y}^{n_k} \to \tilde{y}$. For convenience, we shall abuse notation and denote the subsequence n_k simply by n.

Define $\bar{r}$ to be any solution to

$$\begin{aligned}\bar{r}_1(t) &= \bar{\bar{q}}_1 + \int_0^t [\lambda - (\lambda+\mu)\bar{r}_1(u) + (\nu-\lambda)\bar{r}_3(u)]\,du - \tilde{y}(t)\\ \bar{r}_2(t) &= \bar{\bar{q}}_2 + \int_0^t [\lambda_1 - \mu\bar{r}_2(u)]\,du - \lambda_1 y(t)\\ \bar{r}_3(t) &= \bar{\bar{q}}_3 - \nu\int_0^t \bar{r}_3(u)\,du + \tilde{y}(t).\end{aligned} \tag{B12}$$

Note that $\bar{r}$ must be continuous (as y and $\tilde{y}$ are continuous).

We now show that the subsequence $r^n \to \bar{r}$. Note that we can write

$$\begin{aligned}\|r^n - \bar{r}\|_t \le{}& \|r^n(0) - \bar{\bar{q}}\| + (2\lambda+\mu+\nu)\int_0^t \|r^n - \bar{r}\|_u\,du + \lambda_2 t/\sqrt{n} + \|\hat{\alpha}^n\|_t\\ &+ \|\hat{\delta}^n\|_t + \|\tilde{y}^n - \tilde{y}\|_t + \lambda_1\|y^n - y\|_t + \lambda_2 y^n(t)/\sqrt{n}.\end{aligned} \tag{B13}$$

Thus, applying Gronwall's lemma we obtain, for any $T>0$,

$$\begin{aligned}\|r^n - \bar{r}\|_T \le{}& C_1(\|r^n(0) - \bar{\bar{q}}\| + \lambda_2 t/\sqrt{n} + \|\hat{\alpha}^n\|_T + \|\hat{\delta}^n\|_T + \|\tilde{y}^n - \tilde{y}\|_T\\ &+ \lambda_1\|y^n - y\|_T + \lambda_2 y^n(t)/\sqrt{n})e^{C_2 T}\end{aligned}$$

for some constants C_1, C_2. Proceeding as in the proof of Proposition 4.3(a), we can show that $\|\hat{\alpha}^n\|_T$, $\|\hat{\delta}^n\|_T \to 0$. Further, using the convergence of $r^n(0)$ and those of y^n and $\tilde{y}^n$ proved earlier, we can conclude that $r^n \to \bar{r}$ uniformly on $[0,T]$, almost surely (a.s.).

As $r^n(t) \in \{(x,y,z)\in\mathbb{R}_+^3\colon x+y\le\kappa^n,\ x+z\le 1\}$, we must have $\bar{r}(t)\in\{(x,y,z)\in\mathbb{R}_+^3\colon x+y\le k_1,\ x+z\le 1\}$. Further, noting that $-(k_1+|k_2|)\le(r_1^n(t)+r_2^n(t)-\kappa^n)\le 0$, we apply Lemma 2.4 in Dai and Williams [9] with $f(x)\equiv x\wedge 0\vee -(k_1+|k_2|)$ for $x\in\mathbb{R}$ to obtain

$$\begin{aligned}\int_0^{\cdot}(r_1^n(u)+r_2^n(u)-\kappa^n)\,dy^n(u) &= \int_0^{\cdot} f(r_1^n(u)+r_2^n(u)-\kappa^n)\,dy^n(u) \to \int_0^{\cdot} f(\bar{r}_1(u)+\bar{r}_2(u)-k_1)\,dy(u)\\ &= \int_0^{\cdot}(\bar{r}_1(u)+\bar{r}_2(u)-k_1)\,dy(u),\end{aligned}$$

where the last equality follows by observing that $-(k_1+|k_2|)\le(\bar{r}_1(u)+\bar{r}_2(u)-k_1)\le 0$. Thus, from Equation (B11) we have:

$$\int_0^{\cdot}(\bar{r}_1(u)+\bar{r}_2(u)-k_1)\,dy(u)=0. \tag{B14}$$

Another similar application of Lemma 2.4 in Dai and Williams [9] gives us $\int_0^{\cdot}(\lambda(1-r_1^n(u))+(\nu-\lambda)\cdot r_3^n(u))\,dy^n(u)\to\int_0^{\cdot}((1-\bar{r}_1(u))\lambda+(\nu-\lambda)\bar{r}_3(u))\,dy(u)$. Noting that we can write $\tilde{y}^n(t)=\int_0^t((1-r_1^n(u))\lambda+(\nu-\lambda)r_3^n(u))dy^n(u)$, this implies that $\tilde{y}^n\to\int_0^{\cdot}((1-\bar{r}_1(u))\lambda+(\nu-\lambda)\bar{r}_3(u))\,dy(u)$. However, we know that $\tilde{y}^n\to\tilde{y}$ and thus $\tilde{y}(t)=\int_0^t((1-\bar{r}_1(u))\lambda+(\nu-\lambda)\bar{r}_3(u))\,dy(u)$.

We now prove that $\bar{r}(t)=\bar{\bar{q}}$ for $t\in[0,T]$. To do so, we prove that there exists some $\epsilon>0$ such that $\bar{r}(t)=\bar{\bar{q}}$ for $t\in[0,\epsilon]$. The result then follows by using the time homogeneity of $\bar{r}(\cdot)$ and repeating the argument for intervals $[\epsilon,2\epsilon]$, $[2\epsilon,3\epsilon]$, … .

We first prove that $\bar{r}$ is differentiable almost everywhere (a.e.) on $[0,T]$. Observe that y^n defined in Equation (B10) satisfies $|y^n(t)-y^n(s)|\le|t-s|$ for all $s,t\ge 0$ and $n\in\mathbb{N}$, and thus its limit y is Lipschitz continuous. Noting that a function that is Lipschitz continuous must be absolutely continuous as well, it is almost everywhere differentiable (cf. Corollary 12 in §5.4 of Royden [37]). This allows us to take derivatives in Equation (B12) to obtain

$$\begin{aligned}\dot{\bar{r}}_1(t) &= \lambda-(\lambda+\mu)\bar{r}_1(t)+(\nu-\lambda)\bar{r}_3(t)-((1-\bar{r}_1(t))\lambda+(\nu-\lambda)\bar{r}_3(t))\dot{y}(t)\\ \dot{\bar{r}}_2(t) &= \lambda_1-\mu\bar{r}_2(t)-\lambda_1\dot{y}(t)\\ \dot{\bar{r}}_3(t) &= -\nu\bar{r}_3(t)+((1-\bar{r}_1)\lambda+(\nu-\lambda)\bar{r}_3(t))\dot{y}(t)\end{aligned} \tag{B15}$$

a.e. on $[0,T]$.

We now prove that there exists $\epsilon_1 > 0$ such that $\bar{r}_1(t) + \bar{r}_2(t) = k_1$ for all $0 \le t \le \epsilon_1$. Suppose for some $x > 0$, $\bar{r}_1(x) + \bar{r}_2(x) < k_1$ (if such an x does not exist, then the result trivially follows). Then, if $\dot{y}(x)$ exists, it must equal zero by Equation (B14). Thus, if $\dot{\bar{r}}(x)$ exists, we must have

$$\begin{aligned}\dot{\bar{r}}_1(x) &= \lambda - (\lambda+\mu)\bar{r}_1(x) + (\nu-\lambda)\bar{r}_3(x)\\ \dot{\bar{r}}_2(x) &= \lambda_1 - \mu\bar{r}_2(x).\end{aligned} \tag{B16}$$

Note that the right-hand sides in the above relations are strictly positive at $x = 0$. Thus, as $\bar{r}$ is continuous, there exists some $\epsilon_1 > 0$ such that $\dot{\bar{r}}_i(t) \ge 0$ for $i = 1, 2$ and $t \in [0, \epsilon_1]$, and thus $\dot{\bar{r}}_1(t) + \dot{\bar{r}}_2(t) \ge 0$ on $[0, \epsilon_1]$. Arguing as in Lemma 1 of Dai and Prabhakar [8], we obtain $\bar{r}_1(t) + \bar{r}_2(t) = k_1$ for $t \in [0, \epsilon_1]$.

We now prove that $\bar{r}(t) = \bar{\bar{q}}$ on $[0, \epsilon]$ for some $\epsilon > 0$. As we have $\bar{r}_1(t) + \bar{r}_2(t) = k_1$ for $t \in [0, \epsilon_1]$, we must have $\dot{\bar{r}}_1(t) + \dot{\bar{r}}_2(t) = 0$ a.e. on $[0, \epsilon_1]$. Using this condition in Equation (B15), we obtain

$$\dot{y}(t) = 1 - \frac{\mu k_1}{(1-\bar{r}_1(t))\lambda + (\nu-\lambda)\bar{r}_3(t) + \lambda_1}, \quad \text{a.e. on } [0, \epsilon_1]. \tag{B17}$$

Using Equation (B17), we can rewrite Equation (B15) as

$$\dot{\bar{r}}(t) = \psi(\bar{r}(t)), \quad \text{a.e. on } [0, \epsilon_1] \tag{B18}$$

with the initial condition $\bar{r}(0) = \bar{\bar{q}}$, where for $x \in \mathbb{R}^3$, the function ψ: $\mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$\begin{aligned}\psi_1(x) &= -\mu x_1 + \mu k_1\left(1 - \frac{\lambda_1}{(1-x_1)\lambda + (\nu-\lambda)x_3 + \lambda_1}\right),\\ \psi_2(x) &= -\mu x_2 + \lambda_1\mu k_1 \frac{1}{(1-x_1)\lambda + (\nu-\lambda)x_3 + \lambda_1},\\ \psi_3(x) &= (1-x_1-x_3)\lambda - \mu k_1\left(1 - \frac{\lambda_1}{(1-x_1)\lambda + (\nu-\lambda)x_3 + \lambda_1}\right).\end{aligned} \tag{B19}$$

Define $B \equiv \{x \in \mathbb{R}^3_+ : x_1 + x_3 \le 1\}$ (recall that $\bar{r}(t) \in \{(x, y, z) \in \mathbb{R}^3_+ : x + y \le k_1,\ x + z \le 1\}$). Then, for any $x \in B$, we have:

$$\frac{1}{(1-x_1)\lambda + (\nu-\lambda)x_3 + \lambda_1} = \frac{1}{(1-(x_1+x_3))\lambda + \nu x_3 + \lambda_1} \le \frac{1}{\lambda_1}. \tag{B20}$$

Using this relation, it follows in a straightforward fashion that ψ is Lipschitz continuous on B. Using the continuity of $\bar{r}$ and arguing as in p. 42 of Coddington and Levinson [6], the differential equation (B18) can be assumed to hold everywhere on $[0, \epsilon_1]$ without loss of generality. A further application of Theorem 5 in Chapter 6 of Coddington [5] implies that Equation (B18) can have at most one solution on $[0, \epsilon_2]$ for some $0 < \epsilon_2 \le \epsilon_1$. Thus, noting that $\bar{r}(t) = \bar{\bar{q}}$ is a solution to Equation (B18), it must be the unique solution. Defining $\epsilon = \epsilon_2$, we obtain that $\bar{r}(t) = \bar{\bar{q}}$ on $[0, \epsilon]$.

We can now repeat this argument over the intervals $[i\epsilon, (i+1)\epsilon]$ for $i = 1, 2, \ldots, K$, where $K > T/\epsilon - 1$ to obtain $\bar{r}(t) = \bar{\bar{q}}$ on $[0, T]$. As the limit $\bar{r}$ is independent of the convergent subsequence and T is arbitrary, we obtain $r^n \to \bar{\bar{q}}$ uniformly on compact sets, a.s. □

B.1.4. Proof of Proposition 4.5. We shall prove Proposition 4.5(a); the proof of Proposition 4.5(b) follows in a similar manner. Consider the process $Q^{n,\infty} \in D_{\mathbb{R}^2}[0, \infty)$ defined analogous to Q^n but with an infinite number of servers. Let $\pi^{n,\infty}$ be the corresponding invariant distribution. Then, for any $x, y \in \mathbb{Z}_+$, we can write $\pi^{n,\infty}(x, y) = \pi_1^{n,\infty}(x)\pi_2^{n,\infty}(y)$, where for $i = 1, 2$, $\pi_i^{n,\infty}$ is the invariant distribution of $Q_i^{n,\infty}$ and can be obtained using elementary analysis (see Equations (42) and (52) in Das and Srikant [10]). Define $\hat{Q}^{n,\infty} \equiv \sqrt{n}(Q^{n,\infty}/n - \bar{q}) - (k_2, 0, 0)'$ and let $\hat{\pi}_i^{n,\infty}$ denote the corresponding invariant distribution of $\hat{Q}_i^{n,\infty}$. Then, an application of the central limit theorem as used in proving Fact 1 in Appendix I and Fact 7 in Appendix II in Das and Srikant [10] gives us $\hat{\pi}_i^{n,\infty} \Rightarrow \Phi_i$, where for any Borel measurable set A, $\Phi_1(A) = (1/\sqrt{2\pi})\sqrt{(\lambda+\mu)/m}\int_A \exp(-(1/2)(x+k_2)^2(\lambda+\mu)/m)\,dx$ and $\Phi_2(A) = (1/\sqrt{2\pi})\sqrt{\mu/\lambda_1}\int_A \exp(-(1/2)(y-\lambda_2/\mu)^2\mu/\lambda_1)\,dy$. Thus, we obtain the invariant distribution of $\hat{Q}^{n,\infty}$, $\hat{\pi}^{n,\infty} \Rightarrow \hat{\pi}^\infty$, where for any $A \in$ Borel σ-field on $\mathbb{R}^2$, $\hat{\pi}^\infty = \int_A d\Phi_1(x)d\Phi_2(y)$.

Note that the process $\hat{Q}^n$ is a truncation of $\hat{Q}^{n,\infty}$ to the set $S = \{(x, y) \in \mathbb{R}^2 : x + y < 0\}$ (as in §1.6 of Kelly [20]). We can verify that $\hat{\pi}^{n,\infty}$ satisfies the conditions of Theorem 1.2 in Kelly [20], namely a detailed balance relation, and thus $\hat{Q}^{n,\infty}$ is a reversible Markov process. It follows from Corollary 1.10 in Kelly [20] that $\hat{\pi}^{n,\infty}/\hat{\pi}^{n,\infty}(S)$ is an invariant distribution for the process $\hat{Q}^n$. Noting that $\hat{Q}$ has continuous sample paths a.s., we have $D(\widehat{Q}) \equiv \{t \geq 0: \mathbb{P}(\widehat{Q}(t) = \widehat{Q}(t-)) = 1\} = [0, \infty)$. Thus, we obtain the weak convergence of the finite dimensional distributions of $\widehat{Q}^n$ to that of $\widehat{Q}$ by Theorem 7.8 in Chapter 3 of Ethier and Kurtz [14]. Then, the finite dimensional distributions of $\widehat{Q}^n$ with $\widehat{Q}^n(0)$ distributed as $\widehat{\pi}^{n,\infty}/\widehat{\pi}^{n,\infty}(S)$ converge weakly to that of $\widehat{Q}$ with $\widehat{Q}(0)$ distributed as $\widehat{\pi}^{\infty}/\widehat{\pi}^{\infty}(S)$. As $\widehat{Q}^n$ with $\widehat{Q}^n(0)$ distributed as $\widehat{\pi}^{n,\infty}/\widehat{\pi}^{n,\infty}(S)$ is stationary, so is $\widehat{Q}$ with $\widehat{Q}(0)$ distributed as $\widehat{\pi}^{\infty}/\widehat{\pi}^{\infty}(S)$. Thus, we obtain that $\widehat{\pi}^{\infty}/\widehat{\pi}^{\infty}(S)$ is an invariant distribution for the diffusion $\widehat{Q}$. The uniqueness of the invariant distribution established in Proposition 4.4(c) completes the proof. □

B.1.5. Proof of Proposition 4.6. We begin by proving the following irreducibility property of the Markov chain $X(\cdot)$.

LEMMA B.1. *For any sequence of time instants $\{t^m\}$ with $t^m \uparrow \infty$, the Markov chain $X^m \equiv X(t^m)$ is ψ-irreducible.*

PROOF. The proof of this result is straightforward, and we only provide a sketch of the argument. Define the state space $S = 2^U \times \mathbb{R}_+^n \times \{0, 1, \ldots, k\} \times \mathbb{R}_+^k$ and its corresponding σ-field $\mathcal{S} = 2^{2^U} \times \mathcal{B}(\mathbb{R}_+^n) \times 2^{\{0,1,\ldots,k\}} \times \mathcal{B}(\mathbb{R}_+^k)$, where $\mathcal{B}(\mathbb{R}_+^l)$ denotes the Borel σ-field on $\mathbb{R}_+^l$. The result follows by Theorem 4.0.1 in Meyn and Tweedie [30] if we prove that for every $(a, x, i, y) \in 2^U \times \mathbb{R}_+^n \times \{0, 1 \ldots, k\} \times \mathbb{R}_+^i$ and $A = A_1 \times A_2 \times A_3 \times A_4 \in \mathcal{S}$ of positive measure, there exists some w such that $\mathbb{P}(X^w \in A \mid X^0 = (a, x, i, y)) > 0$.

One can explicitly construct a set of paths of positive probability along which $X^w \in A$, conditioned on $X^0 = (a, x, i, y)$, for any w such that $t^w > \max(\max_l x_l, \max_l y_l)$. A sketch of the construction is as follows: For each subscriber whose residual clock runs out, if the subscriber's desired state (in service or not) at time t^w differs from that at zero, we consider paths where the corresponding on or off time is chosen so that at time t^w, the residual time will lie in the desired set. If the subscriber's desired state is the same, we choose a small on or off time and, upon completion of this time, draw the corresponding on or off time so that the residual time at t^w lies in the desired set. Using the fact that both distribution functions F and G are strictly increasing on $[0, \infty)$, we obtain that the probability of all such paths is strictly positive and the result follows. □

We will prove the result using the concept of reallocatable generalized semi-Markov processes (RGSMP) as described in Miyazawa [31]. We proceed analogous to Example 5.1 in Miyazawa [31] and the reader is advised to have a copy of this paper on hand. Let $G' = 2^U \times \{0, 1, \ldots, k\}$ denote the set of macrostates (note that we use the qualifier $'$ to avoid confusion with the distribution G), $S = U \cup \{0, 1, 2, \ldots, k\}$ the set of sites, and $D = \{e, f, g\}$ the set of distribution indices. For each macrostate $g' \in G'$, a finite subset $A(g') \subseteq S$ and a function $\gamma_{g'}$ from $A(g')$ into D are defined. Elements of $A(g')$ are called active sites. Each active site has a clock that counts the remaining lifetime of the present site. $\gamma_{g'}(s)$ denotes the index of the lifetime distribution of site $s \in A(g')$ under macrostate g'. In this case, for $(j^s, j^p) \in G'$, $\gamma_{(j^s, j^p)}(0) = e$ denotes the exponential distribution with rate λ_p^n, $\gamma_{(j^s, j^p)}(s^p) = g$ for $1 \leq s^p \leq j^p$, and $\gamma_{(j^s, j^p)}(u) = g$ if $u \in j^s$, and f otherwise, where f and g denote the distributions F and G, respectively.

We number the servers $1, 2, \ldots, k$ and define the service discipline as follows: If there are j Poisson stream customers in the system, they occupy servers numbered $1, 2, \ldots, j$. Then, an arriving Poisson stream customer who finds j Poisson customers in the system is taken into service by server s with probability $\delta_{s, j+1} = 1/(j+1)$ for $s = 1, \ldots, j+1$ and customers being served by servers $s, s+1, \ldots, j$ move to servers $s+1, s+2, \ldots, j+1$, respectively. All customers in service are served at a unit rate and have service requirements i.i.d. according to the distribution G. If a Poisson customer being served by server s leaves the system, customers in positions $s+1, s+2, \ldots, j$ move to servers $s, s+1, \ldots, j-1$.

The transition probabilities of this system are as follows:

$$q((j^s, j^p, 0), (j^s, j^p+1, s^p)) = \delta_{s^p, j^p+1}, \quad (|j^s| + j^p < k, j^p \geq 0, 1 \leq s^p \leq j^p + 1),$$
$$q((j^s, j^p, 0), (j^s, j^p, 0)) = 1, \quad (|j^s| + j^p = k, j^p \geq 0),$$
$$q((j^s, j^p, s^p), (j^s, j^p - 1, \phi)) = 1, \quad (|j^s| + j^p \leq k, j^p > 0, 1 \leq s^p \leq j^p),$$
$$q((j^s, j^p, u), (j^s \cup \{u\}, j^p, u)) = 1, \quad (|j^s| + j^p < k, j^p \geq 0, u \notin j^s),$$
$$q((j^s, j^p, u), (j^s, j^p, u)) = 1, \quad (|j^s| + j^p = k, j^p \geq 0, u \notin j^s),$$
$$q((j^s, j^p, u), (j^s \setminus \{u\}, j^p, u)) = 1, \quad (|j^s| + j^p \leq k, |j^s| > 0, u \in j^s).$$

Consider the "macrostate process" $\{Z(t) = (J^s(t), J^p(t)): t \geq 0\}$. Assume for now that all distributions are exponential. Then, $\{Z(t)\}$ is clearly an irreducible, ergodic Markov chain (as it is defined on a finite state space), and thus possesses a unique steady state distribution that we denote by π_e. In fact, a straightforward calculation yields

$$\pi_e(j^s, j^p) = \begin{cases} C\left(\dfrac{\lambda}{\mu}\right)^{|j^s|} \dfrac{1}{j^p!}\left(\dfrac{\lambda_p^n}{\mu}\right)^{j^p}, & \text{if } |j^s| + j^p \le k, \quad j^p \ge 0, \\ 0, & \text{else}, \end{cases} \tag{B21}$$

where $C > 0$ is chosen to ensure π_e is a valid distribution. The distribution π_e possesses the following balance properties: For any j^s, j^p such that $|j^s| + j^p < n$ and $j^p \ge 0$, we have

$$\pi_e(j^s, j^p)\lambda = \pi_e(j^s \cup \{u\}, j^p)\mu, \quad \text{for any } u \notin j^s$$
$$\pi_e(j^s, j^p)\lambda_p^n = \pi_e(j^s, j^p + 1)\mu(j^p + 1).$$

These conditions allow us to use Theorem 4.1 in Miyazawa [31] to obtain the density corresponding to a stationary distribution for the RGSMP $\{Z(t)\}$ as

$$\hat{\pi}(j^s, x^s, j^p, x^p) = \pi_e(j^s, j^p) \prod_{u \in j^s} \mu \bar{G}(x_u^s) \prod_{u \notin j^s} \lambda \bar{F}(x_u^s) \prod_{i=1}^{j^p} \mu \bar{G}(x_i^p). \tag{B22}$$

The associated stationary distribution is given by

$$\pi(A) = \sum_{j^s \in A_1,\, j^p \in A_3} \int_{A_2 \times A_4} \hat{\pi}(j^s, x^s, j^p, x^p) \prod_{l=1}^{n} dx_l^s \prod_{m=1}^{j^p} dx_m^p$$

for any $A_1 \times A_2 \in A_3 \times A_4 \in 2^{2^U} \times$ Borel σ-field on $\mathbb{R}^n \times 2^{\{0,1,\ldots,k\}} \times$ Borel σ-field on $\mathbb{R}^k$.

The uniqueness follows by applying Lemmas B.1 and A.4. □

B.2. Proof of results in §5.

B.2.1. Proof of Proposition 5.1. We shall proceed as in the proof of Proposition 4.1. For convenience, we shall fix $n > 0$ and drop it from the notation. Let $\bar{Q}^l = \{\bar{Q}^l(t)\colon t \ge 0\}$ for $l = 0, 1, \ldots,$ where $\bar{Q}^0(t) = \tilde{Q}(0)$ for $t \ge 0$ and for $l > 0$, $\bar{Q}^l$ is given by

$$\begin{aligned} \bar{Q}^l(t) &\equiv \tilde{\Phi}_{k^n}(\bar{U}^l)(t), \\ \bar{U}^l(t) &\equiv Q(0) + N^a\left(\int_0^{t\wedge T^l} \tilde{\lambda}(\bar{Q}^{l-1}(u))\, du\right) - N^d\left(\int_0^{t\wedge T^l} \tilde{\mu}(\bar{Q}^{l-1}(u))\, du\right) \\ &\quad + N^r\left(\int_0^{t\wedge T^l} (n \wedge \bar{Q}_3^{l-1}(u))\nu\, du\right)(1, 0, -1)', \end{aligned} \tag{B23}$$

where the time for l events to occur T^l is given by

$$T^l = \inf\left\{t\colon N^a\left(\int_0^t \tilde{\lambda}(\bar{Q}^{l-1}(u))\, du\right)' e + N^d\left(\int_0^t \tilde{\mu}(\bar{Q}^{l-1}(u))\, du\right)' e + N^r\left(\int_0^t (\bar{Q}_3^{l-1}(u) \wedge n)\nu\, du\right) = l\right\}.$$

Note that given $\bar{Q}^{l-1}$, $\bar{U}^l$ is well-defined and then using the mapping $\tilde{\Phi}_k$ defined in Equation (32), we obtain $\bar{Q}^l$. Thus, to complete the proof, we only need to show

(a) $\bar{Q}^l(t) = \bar{Q}^{l-1}(t)$ for all $0 \le t < T^l$.
(b) $\bar{Q}^l(t) \in \{(x, y, z)\colon x + y \le k^n,\ x + z \le n,\ z \ge 0\}$ for all $t \ge 0$ a.s.
(c) $\lim_{l\to\infty} T^l = \infty$ a.s.

The solution to Equations (33)–(34) $\tilde{Q}$ can then be constructed by setting

$$\tilde{Q}(t) = \bar{Q}^{l-1}(t) \quad \text{for all } 0 \le t < T^l.$$

This gives us the existence of a solution. Uniqueness then follows by using induction on l and noting that Equation (B23) implies that the uniqueness of $\bar{Q}^l$ follows from uniqueness of $\bar{Q}^{l-1}$.

We prove the first claim using induction on l. The case $l=1$ holds trivially as $\bar{Q}^1(t)=\bar{Q}^0(t)=Q(0)$ for $t<T^1$. Now, assume that $\bar{Q}^l(t)=\bar{Q}^{l-1}(t)$ for all $0\le t<T^l$ for some $l\in\mathbb{N}$, then

$$\begin{aligned}
N^a\left(\int_0^{t\wedge T^l}\tilde{\lambda}(\bar{Q}^l(u))\,du\right)&=N^a\left(\int_0^{t\wedge T^l}\tilde{\lambda}(\bar{Q}^{l-1}(u))\,du\right),\\
N^d\left(\int_0^{t\wedge T^l}\tilde{\mu}(\bar{Q}^l(u))\,du\right)&=N^d\left(\int_0^{t\wedge T^l}\tilde{\mu}(\bar{Q}^{l-1}(u))\,du\right),\\
N^r\left(\int_0^{t\wedge T^l}(\bar{Q}^l_3(u)\wedge n)\nu\,du\right)&=N^r\left(\int_0^{t\wedge T^l}(\bar{Q}^{l-1}_3(u)\wedge n)\nu\,du\right),
\end{aligned}\tag{B24}$$

for $0\le t\le T^l$, which implies that $\bar{Q}^{l+1}(t)=\bar{Q}^l(t)$ for $0\le t\le T^l$.

For $T^l\le t<T^{l+1}$,

$$\begin{aligned}
N^a\left(\int_0^{t}\tilde{\lambda}(\bar{Q}^l(u))\,du\right)&=N^a\left(\int_0^{T^l}\tilde{\lambda}(\bar{Q}^{l-1}(u))\,du\right),\\
N^d\left(\int_0^{t}\tilde{\mu}(\bar{Q}^l(u))\,du\right)&=N^d\left(\int_0^{T^l}\tilde{\mu}(\bar{Q}^{l-1}(u))\,du\right).
\end{aligned}$$

Similarly,

$$N^r\left(\int_0^{t}(\bar{Q}^l_3(u)\wedge n)\nu\,du\right)=N^r\left(\int_0^{t\wedge T^l}(\bar{Q}^{l-1}_3(u)\wedge n)\nu\,du\right).$$

Hence, $\bar{Q}^{l+1}(t)=\bar{Q}^l(t)=\bar{Q}^l(T^l)$ for $T^l\le t<T^{l+1}$. Therefore, $\bar{Q}^{l+1}(t)=\bar{Q}^l(t)$ for $0\le t<T^{l+1}$, and the inductive hypothesis holds.

We now show that $\bar{Q}^l(t)\in\tilde{S}^n=\{(x,y,z): x+y\le k^n,\ x+z\le n,\ z\ge 0\}$ for all $t\ge 0$, a.s. Noting that for any process $A\in D_{\mathbb{R}^3}[0,\infty)$, $(1,1,0)\tilde{\Phi}_{k^n}(A)(\cdot)\le k^n$, we obtain $\bar{Q}^l_1(t)+\bar{Q}^l_2(t)\le k^n$ for all $t\ge 0$. We now show that $\bar{Q}^l_1(t)+\bar{Q}^l_3(t)\le n$ for all $t\ge 0$ a.s. Using Equations (32), (B23), and (B24) and the fact that $\tilde{R}_1+\tilde{R}_3=0$, for any $t\ge 0$ we can write

$$\begin{aligned}
\bar{Q}^l_1(t)+\bar{Q}^l_3(t)&=\bar{U}^l_1(t)+\bar{U}^l_3(t)\\
&=\tilde{Q}_1(0)+\tilde{Q}_3(0)+N^a_1\left(\int_0^{t\wedge T^l}[(n-(\bar{Q}^l_1(u)+\bar{Q}^l_3(u)))\wedge n]\lambda\,du\right)\\
&\quad-N^d_1\left(\int_0^{t\wedge T^l}\bar{Q}^l_1(u)^+\mu\,du\right).
\end{aligned}$$

Thus, using $\tilde{Q}_1(0)+\tilde{Q}_3(0)\le n$, we have $\bar{Q}^l_1(t)+\bar{Q}^l_3(t)\le n$ for $t\ge 0$ a.s. Finally, $\bar{Q}^l_3(t)\ge 0$ a.s. follows as Equation (B23) implies that $\bar{U}^l_3(t)\ge 0$, and noting that $(\tilde{\Phi}_{k^n}(\bar{U}^l)(t)-\bar{U}^l(t))_3\ge 0$.

Now, all that is left to show is that $\lim_{l\to\infty}T^l=\infty$ a.s. Note that we have the following bound on the rate functions: $\sup_{x\in\tilde{S}^n}\max_i|\tilde{\lambda}_i(x)|+\sup_{x\in\tilde{S}^n}\max_i|\tilde{\mu}_i(x)|+\sup_{y\ge 0}(y\wedge n)\nu\le Cn$, where $C<\infty$ is some constant. Thus, $T^l\ge\bar{T}^l$, where $\bar{T}^l$ is defined using the upper bound Cn on the rate functions as follows:

$$\bar{T}^l=\inf\{t: N^a(Cnt)'e+N^d(Cnt)'e+N^r(Cnt)=l\}.$$

Clearly, $\bar{T}^l\to\infty$ a.s., as $l\to\infty$, and the result follows. □

B.2.2. Proof of Proposition 5.2(a). This proof follows in an identical fashion to that of Proposition 4.3(a). The only difference is that, in this case, the rate functions are slightly different and there is no term corresponding to δ^n. We reproduce the entire argument for completeness.

Using Equations (37)–(45) and the fact that $\bar{q}=(\lambda/(\lambda+\mu),\lambda_1/\mu,0)'$ and $k_1=\lambda/(\lambda+\mu)+\lambda_1/\mu$, we obtain:

$$\begin{aligned}
&(\tilde{q}^n(0)-\bar{q})+\left(0,\frac{\lambda_2}{\sqrt{n}}t,0\right)'+\int_0^t\tilde{\theta}(\tilde{q}^n(u))\,du+\tilde{\alpha}^n(t)-(1,1,0)'\sup_{0\le s\le t}(\tilde{X}^n_1(s)+\tilde{X}^n_2(s)-\kappa^n)^+\\
&\quad\le(\tilde{q}^n(t)-\bar{q})\\
&\quad\le(\tilde{q}^n(0)-\bar{q})+\left(0,\frac{\lambda_2}{\sqrt{n}}t,0\right)'+\int_0^t\tilde{\theta}(\tilde{q}^n(u))\,du+\tilde{\alpha}^n(t)+(0,0,1)'\sup_{0\le s\le t}(\tilde{X}^n_1(s)+\tilde{X}^n_2(s)-\kappa^n)^+,
\end{aligned}\tag{B25}$$

where

$$\tilde{\theta}(x,y,z)=(\lambda[(1-x-z)\wedge 1]-\mu[x]^{+}+\nu[z\wedge 1],\lambda_1-\mu[y\wedge(k_1+|k_2|)]^{+},-\nu[z\wedge 1])',\quad \text{for } (x,y,z)\in\mathbb{R}^3. \quad \text{(B26)}$$

Noting that $\tilde{\theta}(\bar{q})=0$, we have

$$\begin{aligned}
&\|\tilde{q}^n-\bar{q}\|_t\\
&\quad\le \|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\tilde{\theta}(\tilde{q}^n(u))-\tilde{\theta}(\bar{q})|\,du+\|\tilde{\alpha}^n\|_t+\sup_{0\le s\le t}(\widetilde{X}_1^n(s)+\widetilde{X}_2^n(s)-\kappa^n)^{+}\\
&\quad\le \|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\tilde{\theta}(\tilde{q}^n(u))-\tilde{\theta}(\bar{q})|\,du+\|\tilde{\alpha}^n\|_t\\
&\qquad+[\|\widetilde{X}_1^n+\widetilde{X}_2^n-(\bar{q}_1+\bar{q}_2)\|_t+|\kappa^n-(\bar{q}_1+\bar{q}_2)|]\\
&\quad\le \|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+\int_0^t|\tilde{\theta}(\tilde{q}^n(u))-\tilde{\theta}(\bar{q})|\,du+\|\tilde{\alpha}^n\|_t+2\|\widetilde{X}^n-\bar{q}\|_t+\frac{k_2}{\sqrt{n}}. \qquad \text{(B27)}
\end{aligned}$$

Note that $\tilde{\theta}$ defined in Equation (B26) is Lipschitz continuous with constant $K_1=2\lambda+\mu+\nu$. To see this, note that for $(x_1,x_2,x_3),(y_1,y_2,y_3)\in\mathbb{R}^3$, we can write:

$$\begin{aligned}
&|\tilde{\theta}_1(x_1,x_2,x_3)-\tilde{\theta}_1(y_1,y_2,y_3)|\\
&\quad\le\lambda|[(1-x_1-x_3)\wedge 1]-[(1-y_1-y_3)\wedge 1]|+\mu|(x_1\wedge 1)^{+}-(y_1\wedge 1)^{+}|+\nu|(x_3\wedge 1)-(y_3\wedge 1)|\\
&\quad\le(2\lambda+\mu+\nu)\|x-y\|,
\end{aligned}$$

where the last inequality holds by using Lemma A.5 and the fact that for any function $f\colon\mathbb{R}\to\mathbb{R}$, if $|f(x)-f(y)|\le C|x-y|$ for any $x,y\in\mathbb{R}$ and constant $C>0$, then $|f(x)^{+}-f(y)^{+}|\le C|x-y|$. Similarly, we can show that the same upper bound holds for $|\tilde{\theta}_i(x_1,x_2,x_3)-\tilde{\theta}_i(y_1,y_2,y_3)|$, $i=2,3$.

Thus, we can rewrite Equation (B27) as

$$\begin{aligned}
\|\tilde{q}^n-\bar{q}\|_t\le \|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|\tilde{q}^n-\bar{q}\|_u\,du+\|\tilde{\alpha}^n\|_t\\
+2\|\widetilde{X}^n-\bar{q}\|_t+\frac{k_2}{\sqrt{n}}.
\end{aligned}$$

Note that we can repeat the argument in Equations (B25) and (B27) for $\widetilde{X}^n$ to obtain

$$\|\widetilde{X}^n-\bar{q}\|_t\le \|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|\tilde{q}^n-\bar{q}\|_u\,du+\|\tilde{\alpha}^n\|_t.$$

Thus, we have

$$\|\tilde{q}^n-\bar{q}\|_t\le 3\left(\|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}t+(2\lambda+\mu+\nu)\int_0^t\|\tilde{q}^n-\bar{q}\|_u\,du+\|\tilde{\alpha}^n\|_t+\frac{k_2}{\sqrt{n}}\right).$$

Using Gronwall's lemma, we obtain:

$$\|\tilde{q}^n-\bar{q}\|_T\le D_1\left(\|\tilde{q}^n(0)-\bar{q}\|+\frac{\lambda_2}{\sqrt{n}}T+\|\tilde{\alpha}^n\|_T+\frac{k_2}{\sqrt{n}}\right)e^{D_2T},$$

where D_1 and D_2 are constants. Noting that $\tilde{q}^n(0)\Rightarrow\bar{q}$ and $\bar{q}$ is a deterministic constant, we have $\|\tilde{q}^n(0)-\bar{q}\|\xrightarrow{p}0$, and thus $\|\tilde{q}^n(0)-\bar{q}\|\Rightarrow 0$. Hence, we only need to show that $\|\tilde{\alpha}^n\|_T\Rightarrow 0$ to complete the result. Note that we have the following bound on the rate functions: $\sup_{x\in\tilde{S}^n}\max_i\tilde{\lambda}_i^n(x)+\sup_{x\in\tilde{S}^n}\max_i\tilde{\mu}_i^n(x)+\sup_{y\ge 0}(y\wedge n)\nu\le Cn$, where $C<\infty$ is some constant. Thus, for $i=1,2$, we have $\|\bar{N}_i^a(\int_0^{\cdot}\tilde{\lambda}^n(n\tilde{q}^n(u))\,du)\|_T\le\|\bar{N}_i^a(Cn\cdot)\|_T$. Using the functional law of large numbers, as $\|(1/n)\bar{N}_i^a(n\cdot)\|_T\Rightarrow 0$, we have $\|(1/n)\bar{N}_i^a(Cn\cdot)\|_T\Rightarrow 0$. Thus, $(1/n)\|\bar{N}_i^a(\int_0^{\cdot}\tilde{\lambda}^n(n\tilde{q}^n(u))\,du)\|_T\Rightarrow 0$. Similarly,

$$\frac{1}{n}\left\|\bar{N}_i^d\left(\int_0^{\cdot}\tilde{\mu}^n(n\tilde{q}^n(u))\,du\right)\right\|_T,\qquad \frac{1}{n}\left\|\bar{N}^r\left(\int_0^{\cdot}n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right)\right\|_T\ \Rightarrow\ 0.$$

Hence, $\|\tilde{\alpha}^n\|_T\Rightarrow 0$. □

B.2.3. Proof of Proposition 5.2(b). We shall prove this result on similar lines as Theorem 7.2 in Mandelbaum and Pats [27]. First, we shall rewrite Q^{*n} in the following manner.

$$Q^{*n} = \sqrt{n}\left(\widetilde{\Phi}_{\kappa^n}\left(\bar{q} + \frac{\widetilde{Z}^n + (k_2, 0, 0)'}{\sqrt{n}}\right) - \widetilde{\Phi}_{\kappa^n}(\bar{q}\mathbf{e})\right) - (k_2, 0, 0)', \tag{B28}$$

where

$$\widetilde{Z}^n(t) = Q^{*n}(0) + \widetilde{D}^n_{\tilde{\theta}}(t) + \widetilde{M}^n(t), \tag{B29}$$

$$\widetilde{D}^n_{\tilde{\theta}}(t) = \sqrt{n}\int_0^t \left(\tilde{\theta}(\bar{q}^n(u)) - \tilde{\theta}(\bar{q}) + \left(0, \frac{\lambda_2}{\sqrt{n}}, 0\right)'\right) du \tag{B30}$$

with $\tilde{\theta}$ as in Equation (B26), $\widetilde{M}^n$ as in Equation (44), and $\mathbf{e}(t) = (1,1,1)'$ for $t \geq 0$.

Using the fact that for any vector $z \in \mathbb{R}^3$, $\widetilde{\Phi}_{a+z_1+z_2}(X+z) = \widetilde{\Phi}_a(X) + z$ and for any $b \in \mathbb{R}_+$, $b\widetilde{\Phi}_0(X) = \widetilde{\Phi}_0(bX)$, we can rewrite Q^{*n} as

$$Q^{*n} = \widetilde{\Phi}_0(\widetilde{Z}^n). \tag{B31}$$

The result then follows by applying Lemmas B.2–B.8 below.

LEMMA B.2. *The mapping $\widetilde{\Phi}_a$ is Lipschitz continuous in the uniform metric for each $a \in \mathbb{R}$, i.e, there exists $K \geq 0$ such that for every $T > 0$ and $Z^1, Z^2 \in D_{\mathbb{R}^3}[0,\infty)$, $\|\widetilde{\Phi}_a(Z^1) - \widetilde{\Phi}_a(Z^2)\|_T \leq K\|Z^1 - Z^2\|_T$.*

PROOF. Using Equation (32), we have for any process $Z \in D_{\mathbb{R}^3}[0,\infty)$

$$\widetilde{\Phi}_a(Z)(t) = Z(t) + \widetilde{R}Y(t), \tag{B32}$$

where Y is a nonnegative, nondecreasing process such that

$$\hat{n}'\widetilde{R}Y(t) = -\sup_{0\leq s\leq t}(Z_1(s) + Z_2(s) - a)^+. \tag{B33}$$

Using this representation of $\widetilde{\Phi}$, we obtain

$$\begin{aligned}
&\|\widetilde{\Phi}_a(Z^1) - \widetilde{\Phi}_a(Z^2)\|_T \\
&\quad \leq \|Z^1 - Z^2\|_T + \frac{\max(m,\lambda_1)}{m+\lambda_1}\left\|\sup_{0\leq s\leq t}(Z_1^2 + Z_2^2 - a)^+(s) - \sup_{0\leq s\leq t}(Z_1^1 + Z_2^1 - a)^+(s)\right\|_T \\
&\quad \leq K\|Z^1 - Z^2\|_T,
\end{aligned}$$

where the second inequality follows by noting that for any two real functions f, g, we have $|\sup_{0\leq s\leq t} f(s) - \sup_{0\leq s\leq t} g(s)| \leq \sup_{0\leq s\leq t}|f(s) - g(s)|$ and $\|Z_1^2 + Z_2^2 - (Z_1^1 + Z_2^1)\|_T \leq 2\|Z^1 - Z^2\|_T$. Hence, the mapping is Lipschitz continuous with the constant $K = 1 + 2(\max(m,\lambda_1)/(m+\lambda_1))$. □

LEMMA B.3. *$\widetilde{M}^n \Rightarrow W$, where $W_1(\cdot) = \sqrt{2m}B_1(\cdot)$, $W_2(\cdot) = \sqrt{2\lambda_1}B_2(\cdot)$, and $W_3(\cdot) = 0$.*

PROOF. The proof involves using the strong approximation result to bound the error between the centered Poisson processes and the corresponding Brownian motions, with appropriate time changes. We then use the property that a spatial scaling of a Brownian motion remains a Brownian motion with an appropriate time change. Finally, we use the fluid level convergence established in Proposition 5.2(a) to complete the argument.

We associate the independent unit rate Poisson processes $N_i^a(\cdot)$, $i = 1,2$, $N_j^d(\cdot)$, $j = 1,2$, and $N^r(\cdot)$ with a family of independent standard Brownian motions $B_i^a(\cdot)$, $i = 1,2$, $B_j^d(\cdot)$, $j = 1,2$, and $B^r(\cdot)$ such that the strong approximation result in Equation (A1) of Lemma A.1 holds. In particular, there exist random variables X_i^z for $z \in \{a, d\}$ and $i = 1,2$ and X^r, such that

$$X_i^z \equiv \sup_{t\geq 0}\frac{|N_i^z(t) - t - B_i^z(t)|}{\log(2\vee t)} < \infty,$$

and

$$X^r \equiv \sup_{t\geq 0}\frac{|N^r(t) - t - B^r(t)|}{\log(2\vee t)} < \infty.$$

Moreover, X_i^z and X^r are i.i.d and have finite means. Let $X \equiv X_1^a + X_2^a + X_1^d + X_2^d + X^r$.

We shall first focus on the term $\widetilde{M}_1^n$. Recall from Equations (42)–(44),

$$\widetilde{M}_1^n = \frac{1}{\sqrt{n}}(\widetilde{M}_1^{a,n}(t) - \widetilde{M}_1^{d,n}(t)) = \frac{1}{\sqrt{n}}\bar{N}_1^a\left(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right) + \frac{1}{\sqrt{n}}\bar{N}^r\left(\int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du\right) - \frac{1}{\sqrt{n}}\bar{N}_1^d\left(\mu n\int_0^t \tilde{q}_1^n(u)^+\,du\right).$$

Using the strong approximation result, we obtain

$$\sup_{t\ge 0}\frac{1}{\sqrt{n}}\frac{|\bar{N}_1^a(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du) - B_1^a(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du)|}{\log(2\vee\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du)} \le \frac{X_1^a}{\sqrt{n}}, \tag{B34}$$

$$\sup_{t\ge 0}\frac{1}{\sqrt{n}}\frac{|\bar{N}^r(\int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du) - B^r(\int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du)|}{\log(2\vee\int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du)} \le \frac{X^r}{\sqrt{n}}, \tag{B35}$$

$$\sup_{t\ge 0}\frac{1}{\sqrt{n}}\frac{|\bar{N}_1^d(\mu n\int_0^t \tilde{q}_1^n(u)^+\,du) - B_1^d(\mu n\int_0^t \tilde{q}_1^n(u)^+\,du)|}{\log(2\vee \mu n\int_0^t \tilde{q}_1^n(u)^+\,du)} \le \frac{X_1^d}{\sqrt{n}}. \tag{B36}$$

Defining $K\equiv(\lambda+\mu+\nu)$, we note that $Knt \ge \max(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du, \int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du, \mu n\int_0^t \tilde{q}_1^n(u)^+\,du)$. Thus, combining Equations (B34)–(B36), we obtain

$$\sup_{t\ge 0}\frac{|\widetilde{M}_1^n(t) - \widetilde{V}^n(t)|}{\log(2\vee Knt)} \le \frac{1}{\sqrt{n}}X, \tag{B37}$$

where

$$\widetilde{V}^n(t) \equiv \frac{1}{\sqrt{n}}B_1^a\left(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right) + \frac{1}{\sqrt{n}}B^r\left(\int_0^t \nu n(\tilde{q}_3^n(u)\wedge 1)\,du\right) - \frac{1}{\sqrt{n}}B_1^d\left(\mu n\int_0^t \tilde{q}_1^n(u)^+\,du\right) \tag{B38}$$

for $t\ge 0$ and K is some constant. Note that by definition, the distribution of X is independent of n. Using Equation (B37), we obtain

$$\|\widetilde{M}_1^n - \widetilde{V}^n\|_T \le \log(2\vee KnT)\sup_{T\ge t\ge 0}\frac{|\widetilde{M}_1^n(t)-\widetilde{V}^n(t)|}{\log(2\vee Knt)} \le \log(2\vee KnT)\sup_{t\ge 0}\frac{|\widetilde{M}_1^n(t)-\widetilde{V}^n(t)|}{\log(2\vee Knt)} \le \frac{\log(2\vee KnT)}{\sqrt{n}}X.$$

As the distribution of X is independent of n and $\lim_{n\to\infty}(\log(2\vee KnT)/\sqrt{n}) = 0$, we obtain $\|\widetilde{M}_1^n - \widetilde{V}^n\|_T \Rightarrow 0$. Thus, using Theorem 3.1 in Chapter 1 of Billingsley [3], to prove the claim, it suffices to prove the convergence of $\widetilde{V}^n$. We first focus on the term $(1/\sqrt{n})B_1^a(\int_0^{\cdot} \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du)$. Note that we can write

$$\frac{1}{\sqrt{n}}\left\|B_1^a\left(\int_0^{\cdot} \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right) - B_1^a(nm\cdot)\right\|_T = \frac{1}{\sqrt{n}}\left\|B_1^a\left(\int_0^{\cdot} \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right) - B_1^a\left(\int_0^{\cdot} \lambda n[(1-\bar{q}_1-\bar{q}_3)\wedge 1]\,du\right)\right\|_T \tag{B39}$$

by noting that for any $t\ge 0$, we can write $nmt = \int_0^t \lambda n[(1-\bar{q}_1-\bar{q}_3)\wedge 1]\,du$ (the minimum with 1 is redundant but we leave it in for clarity of the next step). We now argue as in Lemma 3.2 of Kurtz [25]. Define $\gamma^n(t)\equiv (1/n)|\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du - \int_0^t \lambda n[(1-\bar{q}_1^n-\bar{q}_3^n)\wedge 1]\,du|$ and $\bar{\gamma}^n \equiv \sup_{0\le t\le T}\gamma^n(t)$. Note that $\gamma^n(t)\le 2\lambda t$, and thus $\bar{\gamma}^n$ is well-defined. Using Lemma A.2, we have for $0\le t\le T$:

$$\begin{aligned}\frac{1}{\sqrt{n}}&\left|B_1^a\left(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right) - B_1^a\left(\int_0^t \lambda n[(1-\bar{q}_1-\bar{q}_3)\wedge 1]\,du\right)\right| \\ &= \frac{1}{\sqrt{n}}\frac{|B_1^a(\int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du) - B_1^a(\int_0^t \lambda n[(1-\bar{q}_1-\bar{q}_3)\wedge 1]\,du)|}{\sqrt{n\gamma^n(t)(1+\log(\lambda T/\gamma^n(t))}} \\ &\quad\times\sqrt{n\gamma^n(t)(1+\log(\lambda T/\gamma^n(t))} \\ &\le \bar{M}\sqrt{\bar{\gamma}^n(1+\log(\lambda T/\bar{\gamma}^n))},\end{aligned} \tag{B40}$$

where the last inequality follows from the fact that $\sqrt{x(1+\log(\lambda T/x))}$ is increasing in x.

Combining this with Equation (B39), we obtain

$$\frac{1}{\sqrt{n}}\left\|B_1^a\left(\int_0^{\cdot}\lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right)-B_1^a(nm\cdot)\right\|_T\le \bar{M}\sqrt{\bar{\gamma}^n(1+\log(\lambda T/\bar{\gamma}^n))}.$$

Further, applying Lemma A.5, we have:

$$\bar{\gamma}^n=\left\|\int_0^{\cdot}\lambda([(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]-[(1-\bar{q}_1^n-\bar{q}_3^n)\wedge 1])\,du\right\|_T\le 2\lambda T\|\tilde{q}^n-\bar{q}\|_T.$$

Thus, Proposition 5.2(a) implies that $\bar{\gamma}^n\Rightarrow 0$. Noting that $\lim_{x\to 0}\sqrt{x(1+\log(\lambda T/x))}=0$ and that the distribution of $\bar{M}$ is independent of n, we obtain $\bar{M}\sqrt{\bar{\gamma}^n(1+\log(\lambda T/\bar{\gamma}^n))}\Rightarrow 0$, and thus:

$$\frac{1}{\sqrt{n}}\left\|B_1^a\left(\int_0^{\cdot}\lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du\right)-B_1^a(nm\cdot)\right\|_T\ \Rightarrow\ 0. \tag{B41}$$

Arguing similarly, we obtain:

$$\frac{1}{\sqrt{n}}\left\|B^r\left(\int_0^{\cdot}\nu n(\tilde{q}_3^n(u)\wedge 1)\,du\right)\right\|_T\ \Rightarrow\ 0,$$

$$\frac{1}{\sqrt{n}}\left\|B_1^d\left(\mu n\int_0^{\cdot}\tilde{q}_1^n(u)^+\,du\right)-B_1^d(nm\cdot)\right\|_T\ \Rightarrow\ 0.$$

Thus, $\|\widetilde{M}_1^n-(1/\sqrt{n})B_1^a(nm\cdot)-(1/\sqrt{n})B_1^d(nm\cdot)\|_T\Rightarrow 0$. In a similar fashion, we can prove that:

$$\left\|\widetilde{M}_2^n-\frac{1}{\sqrt{n}}B_2^a(n\lambda_1\cdot)-\frac{1}{\sqrt{n}}B_2^d(n\lambda_1\cdot)\right\|_T\ \Rightarrow\ 0,$$

and $\|\widetilde{M}_3^n\|_T\Rightarrow 0$.

Using the fact that for any Brownian motion B and constant $C>0$, $B(nC\cdot)/\sqrt{n}\overset{d}{=}B(C\cdot)$ for all $n>0$ and Theorem 3.1 in Chapter 1 of Billingsley [3], the result follows. □

Lemma B.4. *The sequence $\{Q^{*n}\}$ adheres to the compact containment condition*

$$\lim_{l\uparrow\infty}\limsup_{n}\mathbb{P}\{\|Q^{*n}\|_T>l\}=0.$$

Proof. Using the Lipschitz continuity of $\widetilde{\Phi}_0$ in Equation (B31), we have:

$$\|Q^{*n}\|_t\le K\|\widetilde{Z}^n\|_t,\quad 0\le t\le T, \tag{B42}$$

where K is the Lipschitz constant derived in Lemma B.2.

Using the Lipschitz continuity of $\tilde{\theta}$ with constant $K_1=2\lambda+\mu+\nu$ (derived in the proof of Proposition 5.2(a)) in Equations (B29)–(B30), we obtain

$$\|\widetilde{Z}^n\|_t\le\|Q^{*n}(0)\|+K_1\int_0^t\|Q^{*n}\|_u\,du+\lambda_2 t+\|\widetilde{M}^n\|_t.$$

Combining the above with Equation (B42), and using Gronwall's lemma, we get:

$$\|Q^{*n}\|_T\le K(\|Q^{*n}(0)+(k_2,0,0)'\|+\|\widetilde{M}^n\|_T+\lambda_2 T)e^{K_2T},$$

where K_2 is another constant. Now, as $\{Q^{*n}(0)\}$ converges weakly and $\{\widetilde{M}^n\}$ converges weakly to a process W that is continuous almost surely, $\|Q^{*n}(0)\|$ and $\|\widetilde{M}^n\|_T$ converge weakly as well by an application of the continuous mapping theorem. Thus, the result follows. □

Lemma B.5. *The sequence $\{(Q^{*n},\widetilde{Z}^n,\widetilde{D}_{\tilde{\theta}}^n,\widetilde{M}^n)\}$ is C-tight.*

Proof. We shall use the same notion of C-tightness used in Chen and Zhang [4]. Let $\{X^n\}$ be a sequence of stochastic processes such that $X^n \in D_{\mathbb{R}^m}[0,\infty)$. Then, $\{X^n\}$ is C-tight if $\{X^n(0)\}$ is tight, and for any $T>0$, for each $\epsilon, \eta > 0$, there exist $\delta > 0$ and $N < \infty$ such that for all $n > N$:

$$\mathbb{P}\left(\sup_{0\le s,\, t\le T,\, |s-t|<\delta} \|X^n(s)-X^n(t)\| \ge \epsilon\right) \le \eta.$$

Note that using this notion of C-tightness, to prove that $\{(Q^{*n}, \tilde{Z}^n, \tilde{D}^n_{\tilde{\theta}}, \tilde{M}^n)\}$ is C-tight, it suffices to prove that the individual sequences $\{Q^{*n}\}$, $\{\tilde{Z}^n\}$, $\{\tilde{D}^n_{\tilde{\theta}}\}$, and $\{\tilde{M}^n\}$ are C-tight. We now establish this in the reverse order. First, note that the sequence $\{M^n\}$ converges to W (cf. Lemma B.3): by applying Lemma 4.2(ii) of Chen and Zhang [4], we obtain that $\{M^n\}$ is C-tight. We now focus on $\{\tilde{D}^n_{\tilde{\theta}}\}$ given by Equation (B30). Note that $\{\tilde{D}^n_{\tilde{\theta}}(0)\}$ is trivially tight. Further, we can use the Lipschitz continuity of $\tilde{\theta}$ (derived in the proof of Proposition 5.2(a)) in Equation (B30) to obtain the relation

$$\|\tilde{D}^n_{\tilde{\theta}}(t) - \tilde{D}^n_{\tilde{\theta}}(s)\| \le (C\|Q^{*n}\|_T + \lambda_2)(t-s), \quad 0\le s\le t\le T. \tag{B43}$$

Applying Lemma B.4 in Equation (B43), we obtain the required C-tightness of $\{\tilde{D}^n_{\tilde{\theta}}\}$. Using the tightness of $\{Q^{*n}(0)\}$ and the C-tightness of $\{\tilde{D}^n_{\tilde{\theta}}\}$ and $\{\tilde{M}^n\}$, we can apply Lemma 4.2(i) of Chen and Zhang [4] to obtain that $\{\tilde{Z}^n\}$ given in Equation (B29) is C-tight. Finally, using Equations (B32) and (B33) and arguing as in the proof of Lemma 4.3 of Chen and Zhang [4], we obtain $\|Q^{*n}(t) - Q^{*n}(s)\| \le \tilde{K}\sup_{s\le u\le t}\|\tilde{Z}^n(u) - \tilde{Z}^n(s)\|$ for $0\le s\le t\le T$ and some constant $\tilde{K}>0$, and thus using the tightness of $\tilde{Z}^n$, we obtain that $\{Q^{*n}\}$ is C-tight. Thus, the sequence $\{(Q^{*n}, \tilde{Z}^n, \tilde{D}^n_{\tilde{\theta}}, \tilde{M}^n)\}$ is C-tight. □

Lemma B.6. *Let* $(Q^*, Z^*, \tilde{D}_{\tilde{\theta}}, W)$ *denote the limit of any weakly convergent subsequence of* $\{(Q^{*n}, \tilde{Z}^n, \tilde{D}^n_{\tilde{\theta}}, \tilde{M}^n)\}$. *Then,* (Q^*, Z^*) *satisfies Equations* (46)–(49) *with* $B_1 = W_1/\sqrt{2m}$, $B_2 = W_2/\sqrt{2\lambda_1}$.

Proof. We shall use the Skorohod representation theorem (as in Theorem 1.8 in Chapter 3 of Ethier and Kurtz [14] and noting that $(D_{\mathbb{R}^m}[0,\infty), d)$ is separable by Theorem 5.6 in Chapter 3 of Ethier and Kurtz [14]). Let $\{n_k\}$ denote the index of the converging subsequence. Then, let $\{(Q'^{*n_k}, \tilde{Z}'^{n_k}, \tilde{D}'^{n_k}_{\tilde{\theta}}, \tilde{M}'^{n_k})\}$, $(Q'^*, Z'^*, \tilde{D}'_{\tilde{\theta}}, W')$ be defined on a different probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ such that $(Q'^{*n_k}, \tilde{Z}'^{n_k}, \tilde{D}'^{n_k}_{\tilde{\theta}}, \tilde{M}'^{n_k}) \stackrel{d}{=} (Q^{*n_k}, \tilde{Z}^{n_k}, \tilde{D}^{n_k}_{\tilde{\theta}}, \tilde{M}^{n_k})$, $(Q'^*, Z'^*, \tilde{D}'_{\tilde{\theta}}, W') \stackrel{d}{=} (Q^*, Z^*, \tilde{D}_{\tilde{\theta}}, W)$, and $(Q'^{*n_k}, \tilde{Z}'^{n_k}, \tilde{D}'^{n_k}_{\tilde{\theta}}, \tilde{M}'^{n_k}) \to (Q'^*, Z'^*, \tilde{D}'_{\tilde{\theta}}, W')$ a.s.

As the original sequence $\{(Q^{*n}, \tilde{Z}^n, \tilde{D}^n_{\tilde{\theta}}, \tilde{M}^n)\}$ is C-tight, the weak limit of any subsequence must be continuous (cf. Proposition 4.1 in Chen and Zhang [4]). Thus, the convergence in Skorohod topology is equivalent to convergence in the uniform on compact sets topology. For any fixed $T>0$, this gives us $\|Z'^{n_k} - Z'^*\|_T \to 0$ a.s., or $\tilde{Z}'^{n_k} \to Z'^*$ uniformly on $[0,T]$, a.s., and we can use the continuity of $\tilde{\Phi}_0$ proved in Lemma B.2 to obtain $Q^{*n_k} = \tilde{\Phi}_0(\tilde{Z}^{n_k}) \to \tilde{\Phi}_0(Z^*) = Q^*$ uniformly on $[0,T]$, a.s. Noting the convergence of $\tilde{M}^n$ in Lemma B.3 and comparing Equations (B31) and (B29) with Equations (46)–(49), the result follows if

$$\tilde{D}'^{n_k}_{\tilde{\theta}}(\cdot) \to \int_0^{\cdot} (-(\lambda+\mu)(Q'^*_1(u)+k_2) + (\nu-\lambda)Q'^*_3(u), \lambda_2 - \mu Q'^*_2(u), -\nu Q'^*_3(u))\, du$$
$$\text{uniformly on } [0,T], \text{ a.s.} \tag{B44}$$

To see this relation, define $\tilde{q}'^{n_k} = \bar{q} + (Q'^{*n_k} + (k_2, 0, 0)')/\sqrt{n_k}$. Then, by the convergence of Q'^{*n_k}, it follows that $\tilde{q}'^{n_k} \to \bar{q}$ uniformly on $[0,T]$, a.s. Thus, for almost every sample path $\omega \in \Omega'$, there exists $N<\infty$ such that for all $n>N$, on $[0,T]$, we have $1 - \tilde{q}'^{n_k}_1 - \tilde{q}'^{n_k}_3 \le 1$, $\tilde{q}'^{n_k}_1 \ge 0$, $\tilde{q}'^{n_k}_2 \le k_1 + |k_2|$, and $\tilde{q}'^{n_k}_3 \le 1$. For all such sample paths ω, we have for $n_k > N$

$$\sqrt{n_k}(\tilde{\theta}(\tilde{q}'^{n_k}) - \tilde{\theta}(\bar{q})) = ((\lambda+\mu)\sqrt{n_k}(\bar{q}_1 - \tilde{q}'^{n_k}_1) + (\nu-\lambda)\sqrt{n_k}\tilde{q}'^{n_k}_3, \mu\sqrt{n_k}(\bar{q}_2 - \tilde{q}'^{n_k}_2), -\nu\sqrt{n_k}\tilde{q}'^{n_k}_3)'.$$

Thus, $\sqrt{n_k}(\tilde{\theta}(\tilde{q}'^{n_k}) - \tilde{\theta}(\bar{q})) \to (-(\lambda+\mu)(Q'^*_1 + k_2) + (\nu-\lambda)Q'^*_3, -\mu Q'^*_2, -\nu Q'^*_3)'$ uniformly on $[0,T]$ a.s., which then immediately implies

$$\int_0^{\cdot} \sqrt{n_k}(\tilde{\theta}(\tilde{q}'^{n_k}(u)) - \tilde{\theta}(\bar{q}))\, du \to \int_0^{\cdot} (-(\lambda+\mu)(Q'^*_1(u)+k_2) + (\nu-\lambda)Q'^*_3(u), -\mu Q'^*_2(u), -\nu Q'^*_3(u))'\, du$$

uniformly on $[0,T]$, a.s.

Hence, Equation (B44) holds. □

LEMMA B.7. *Equations* (46)–(49) *have a unique strong solution.*

PROOF. Note that the reflection $\widetilde{\Phi}_0$ is well-defined in the sense that for any $X \in D_{\mathbb{R}^3}[0,\infty)$, $\widetilde{\Phi}_0(X) \in S = \{(x,y,z): x+y \le 0,\, z \ge 0\}$. Thus, using the Lipschitz continuity of the reflection map $\widetilde{\Phi}_0$ (from Lemma B.2) and the fact that the drift terms are linear, we obtain the existence of a unique strong solution to Equations (46)–(49) by applying Theorem 2.1 in Atar et al. [1]. □

LEMMA B.8. *We have* $(Q^{*n}, \widetilde{Z}^n) \Rightarrow (Q^*, Z^*)$, *the solution to Equations* (46)–(49).

PROOF. This result follows immediately from Lemma B.6 in light of Lemma B.7. □

B.2.4. Proof of Proposition 5.3. We shall prove that $\|\widehat{Q}^n - Q^{*n}\|_T \Rightarrow 0$ by proving that for any sequence $\{\|\widehat{Q}^n - Q^{*n}\|_T\}$, there exists a subsequence denoted by n_l such that $\|\widehat{Q}^{n_l} - Q^{*n_l}\|_T \Rightarrow 0$ as $l \to \infty$. Propositions 4.3(a) and 5.2(a) imply $\|q^n - \bar{q}\|_T$, $\|\tilde{q}^n - \bar{q}\|_T \xrightarrow{p} 0$. Let n_k denote the common subsequence such that $\|q^{n_k}(0) - \bar{q}\|$, $\|q^{n_k} - \bar{q}\|_T$, $\|\tilde{q}^{n_k} - \bar{q}\|_T \to 0$ a.s., as $k \to \infty$. (The existence of such a sequence follows by Theorem 20.5 in Billingsley [2].) For convenience, we shall abuse notation to denote n_k by n. For this subsequence, we can write

$$\widehat{Q}^n(t) - Q^{*n}(t) = \sqrt{n}(X^n(t) - \widetilde{X}^n(t)) + \sqrt{n}\left(\int_0^t R^n(u)\,dY^n(u) - \widetilde{R}\widetilde{Y}^n(t)\right). \tag{B45}$$

Consider the second term. As $\|\tilde{q}^n - \bar{q}\|_T \to 0$, we have $\|R^n(\cdot) - \widetilde{R}\|_T \to 0$. Thus, for every $\epsilon > 0$ small enough, there exists $n^* \in \mathbb{N}$ such that for all $n \ge n^*$,

$$\|R^n(\cdot) - \widetilde{R}\|_T \le \frac{\epsilon}{2}. \tag{B46}$$

This implies that for $0 \le t \le T$

$$\begin{aligned}\int_0^t \left[\widetilde{R} - \frac{\epsilon}{2}(1,1,1)'\right] dY^n(u) - \widetilde{R}\widetilde{Y}^n(t) &\le \left(\int_0^t R^n(u)\,dY^n(u) - \widetilde{R}\widetilde{Y}^n(t)\right)\\ &\le \int_0^t \left[\widetilde{R} + \frac{\epsilon}{2}(1,1,1)'\right] dY^n(u) - \widetilde{R}\widetilde{Y}^n(t),\end{aligned}$$

and thus

$$\begin{aligned}\left[\widetilde{R} - \frac{\epsilon}{2}(1,1,1)'\right][Y^n(t) - \widetilde{Y}^n(t)] - \frac{\epsilon}{2}(1,1,1)'\widetilde{Y}^n(t) &\le \left(\int_0^t R^n(u)\,dY^n(u) - \widetilde{R}\widetilde{Y}^n(t)\right)\\ &\le \left[\widetilde{R} + \frac{\epsilon}{2}(1,1,1)'\right][Y^n(t) - \widetilde{Y}^n(t)] + \frac{\epsilon}{2}(1,1,1)'\widetilde{Y}^n(t).\end{aligned}$$

Hence, we obtain

$$\left\| \int_0^t R^n(u)\,dY^n(u) - \widetilde{R}\widetilde{Y}^n(t) \right\|_T \le (m + \lambda_1 + \epsilon)\|Y^n - \widetilde{Y}^n\|_T + \epsilon\|\widetilde{Y}^n\|_T. \tag{B47}$$

Note that Equation (B46) implies that

$$(m + \lambda_1 - \epsilon)Y^n(t) \le -(1,1,0).\int_0^t R^n(u)\,dY^n(u) \le (m + \lambda_1 + \epsilon)Y^n(t),$$

which combined with Equation (21) implies

$$\frac{\sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+}{m + \lambda_1 + \epsilon} \le Y^n(t) \le \frac{\sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+}{m + \lambda_1 - \epsilon}. \tag{B48}$$

Using Equation (39), we can write $\widetilde{Y}^n$ as follows

$$\widetilde{Y}^n(t) = \frac{\sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+}{m+\lambda_1}. \tag{B49}$$

Thus, using Equations (B48) and (B49) in Equation (B47), we obtain

$$\begin{aligned}
&\left\| \int_0^t R^n(u)\, dY^n(u) - \tilde{R}\widetilde{Y}^n(t) \right\|_T \\
&\le (m+\lambda_1+\epsilon) \left\| \frac{\sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+}{m+\lambda_1-\epsilon} - \frac{\sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+}{m+\lambda_1} \right\|_T \\
&\quad + (m+\lambda_1+\epsilon) \left\| \frac{\sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+}{m+\lambda_1+\epsilon} - \frac{\sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+}{m+\lambda_1} \right\|_T + \epsilon\|\widetilde{Y}^n\|_T \\
&\le \frac{m+\lambda_1+\epsilon}{m+\lambda_1-\epsilon} \left\| \sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+ - \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T \\
&\quad + (m+\lambda_1+\epsilon) \left| \frac{1}{m+\lambda_1-\epsilon} - \frac{1}{m+\lambda_1} \right| \left\| \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T \\
&\quad + \left\| \sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+ - \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T \\
&\quad + (m+\lambda_1+\epsilon) \left| \frac{1}{m+\lambda_1} - \frac{1}{m+\lambda_1+\epsilon} \right| \left\| \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T + \epsilon\|\widetilde{Y}^n\|_T \\
&\overset{(a)}{=} \frac{m+\lambda_1+\epsilon}{m+\lambda_1-\epsilon} \left\| \sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+ - \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T \\
&\quad + \frac{\epsilon(m+\lambda_1+\epsilon)}{m+\lambda_1-\epsilon}\|\widetilde{Y}^n\|_T + \left\| \sup_{0\le s\le t}(X_1^n(s) + X_2^n(s) - \kappa^n)^+ - \sup_{0\le s\le t}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - \kappa^n)^+ \right\|_T \\
&\quad + 2\epsilon\|\widetilde{Y}^n\|_T \\
&\le C\|X^n - \widetilde{X}^n\|_T + \gamma(\epsilon)\|\widetilde{Y}^n\|_T,
\end{aligned}$$

where C is a constant and $\gamma(\epsilon) \equiv \epsilon(m+\lambda_1+\epsilon)/(m+\lambda_1-\epsilon) + 2\epsilon = \epsilon((3(m+\lambda_1)-\epsilon)/(m+\lambda_1-\epsilon))$. The relation (a) follows by using Equation (B49), and the last inequality follows by the observation that for any two real functions f, g we have $|\sup_{0\le s\le t} f(s) - \sup_{0\le s\le t} g(s)| \le \sup_{s\le t}|f(s) - g(s)|$, $\|X_1^n + X_2^n - (\widetilde{X}_1^n + \widetilde{X}_2^n)\|_T \le 2\|X^n - \widetilde{X}^n\|_T$, and $C \ge 2(m+\lambda_1+\epsilon)/(m+\lambda_1-\epsilon) + 2$ ($C = 5$ suffices for ϵ small enough).

Using the above relation in Equation (B45), we obtain

$$\|\widehat{Q}^n - Q^{*n}\|_T \le (C+1)\|\sqrt{n}(X^n - \widetilde{X}^n)\|_T + \gamma(\epsilon)\|\sqrt{n}\widetilde{Y}^n\|_T. \tag{B50}$$

Now, consider the term $\sqrt{n}(X^n - \widetilde{X}^n)$. Using the definition of X^n and $\widetilde{X}^n$ in, Equations (18) and (38), respectively, we obtain

$$\begin{aligned}
\sqrt{n}(X^n - \widetilde{X}^n) &= \frac{1}{\sqrt{n}} \int_0^t [\tilde{\theta}^n(nq^n(u)) - \tilde{\theta}^n(n\tilde{q}^n(u))]\, du + \sqrt{n}(\alpha^n(t) - \tilde{\alpha}^n(t)) + \sqrt{n}\delta^n(t) \\
&= \sqrt{n} \int_0^t [\tilde{\theta}(q^n(u)) - \tilde{\theta}(\tilde{q}^n(u))]\, du + \sqrt{n}(\alpha^n(t) - \tilde{\alpha}^n(t)) + \sqrt{n}\delta^n(t),
\end{aligned}$$

where we use the fact that $\tilde{\theta}^n(nq^n(t)) = \theta^n(nq^n(t))$ for all $t \ge 0$ as $q_i^n(\cdot) \ge 0$ for $i = 1, 2, 3$, $0 \le q_1^n(\cdot) + q_3^n(\cdot) \le 1$ and $q_2^n(\cdot) \le k^n/n \le (k_1 + |k_2|)$, and $\tilde{\theta}^n(nx) = n\tilde{\theta}(x)$ for $x \in \mathbb{R}^3$ with $\tilde{\theta}$ defined in Equation (B26). Hence, using the Lipschitz continuity of $\tilde{\theta}$ (derived in the proof of Proposition 5.2(a)), we obtain

$$\|\sqrt{n}(X^n-\widetilde{X}^n)\|_T \le \int_0^t (2\lambda+\mu+\nu)\|\sqrt{n}(q^n-\tilde{q}^n)\|_u\,du + \|\sqrt{n}(\alpha^n-\tilde{\alpha}^n)\|_T + \|\sqrt{n}\delta^n\|_T.$$

Combining this relation with Equation (B50), we obtain

$$\|\widehat{Q}^n-Q^{*n}\|_T \le (C+1)(2\lambda+\mu+\nu)\int_0^T \|\widehat{Q}^n-Q^{*n}\|_u\,du + (C+1)\|\sqrt{n}(\alpha^n-\tilde{\alpha}^n)\|_T + (C+1)\|\sqrt{n}\delta^n\|_T + \gamma(\epsilon)\|\sqrt{n}\widetilde{Y}^n\|_T.$$

Applying Gronwall's lemma, we obtain

$$\|\widehat{Q}^n-Q^{*n}\|_T \le C_1(\|\sqrt{n}(\alpha^n-\tilde{\alpha}^n)\|_T + \|\sqrt{n}\delta^n\|_T + \gamma(\epsilon)\|\sqrt{n}\widetilde{Y}^n\|_T)\mathrm{e}^{C_2T}, \tag{B51}$$

where C_1, C_2 are constants.

We now prove that $\|\sqrt{n}(\alpha^n-\tilde{\alpha}^n)\|_T, \|\sqrt{n}\delta^n\|_T \Rightarrow 0$. Note that we have:

$$\|\sqrt{n}(\alpha^n-\tilde{\alpha}^n)\|_T \le \left\|\frac{1}{\sqrt{n}}(M^{a,n}-\widetilde{M}^{a,n})\right\|_T + \left\|\frac{1}{\sqrt{n}}(M^{d,n}-\widetilde{M}^{d,n})\right\|_T.$$

We use the family of Brownian motions defined in Lemma B.3 associated with the independent unit rate Poisson processes $N_i^a(\cdot)$, $i=1,2$, $N_j^d(\cdot)$, $j=1,2$, and $N^r(\cdot)$ to obtain

$$\sup_{t\ge 0}\frac{|(1/\sqrt{n})(M_1^{a,n}(t)-\widetilde{M}_1^{a,n}(t))-(V^n(t)-\widetilde{V}^n(t))|}{\log(2\vee \widetilde{C}_1 nt)} \le \frac{\widetilde{C}_2}{\sqrt{n}}X,$$

where X is the random variable, with finite mean, defined in the proof of Lemma B.3, $\widetilde{C}_1$ and $\widetilde{C}_2$ are some constants, and

$$V^n(t) \equiv \frac{1}{\sqrt{n}}B_1^a\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda_1^n(nq^n(u))\,du\right) + \frac{1}{\sqrt{n}}B^r\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\nu n q_3^n(u)\,du\right),$$

$$\widetilde{V}^n(t) \equiv \frac{1}{\sqrt{n}}B_1^a\left(\int_0^t \tilde{\lambda}_1^n(n\tilde{q}^n(u))\,du\right) + \frac{1}{\sqrt{n}}B^r\left(\int_0^t n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right),$$

where λ^n and $\tilde{\lambda}^n$ are defined in Equations (7) and (35), respectively. Thus, as in the proof of Lemma B.3, to prove $\|(1/\sqrt{n})(M_1^{a,n}-\widetilde{M}_1^{a,n})\|_T \Rightarrow 0$, it suffices to prove $\|V^n-\widetilde{V}^n\|_T \Rightarrow 0$. We can write:

$$\|V^n-\widetilde{V}^n\|_T = \frac{1}{\sqrt{n}}\left\|B_1^a\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda_1^n(nq^n(u))\,du\right) - B_1^a\left(\int_0^t \tilde{\lambda}_1^n(n\tilde{q}^n(u))\,du\right)\right\|_T + \frac{1}{\sqrt{n}}\left\|B^r\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\nu n q_3^n(u)\,du\right) - B^r\left(\int_0^t n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right)\right\|_T.$$

We first prove the convergence of the first term of the right-hand side of the above expression. Noting that $\lambda_1^n(nq^n(t)) = \tilde{\lambda}_1^n(nq^n(t))$ for all $t\ge 0$ as $0\le q_1^n(t)+q_3^n(t)\le 1$, we proceed analogous to Equation (B40). Defining $\gamma^n(t) \equiv (1/n)|\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda n[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du - \int_0^t \lambda n[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du|$ and $\bar{\gamma}^n \equiv \sup_{0\le t\le T}\gamma^n(t)$, we have for $0\le t\le T$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left|B_1^a\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda n[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du\right) - B_1^a\left(\int_0^t \lambda n[(1-\tilde{q}_1(u)-\tilde{q}_3(u))\wedge 1]\,du\right)\right| \\
&\quad= \frac{|B_1^a(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda n[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du) - B_1^a(\int_0^t \lambda n[(1-\tilde{q}_1(u)-\tilde{q}_3(u))\wedge 1]\,du)|}{\sqrt{n}\sqrt{n\gamma^n(t)(1+\log(\lambda T/\gamma^n(t))}} \\
&\qquad\times\sqrt{n\gamma^n(t)(1+\log(\lambda T/\gamma^n(t))} \\
&\quad\le \overline{M}\sqrt{\bar{\gamma}^n(1+\log(\lambda T/\bar{\gamma}^n))},
\end{aligned}\tag{B52}$$

where $\bar{M}$ is given in Equation (A2) and the last inequality follows from the fact that $\sqrt{x(1+\log(\lambda T/x))}$ is increasing in x. Note that we can write

$$\begin{aligned}\bar{\gamma}^n &\le \left\| \int_0^{\cdot} \lambda[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du - \int_0^{t} \lambda[(1-\tilde{q}_1^n(u)-\tilde{q}_3^n(u))\wedge 1]\,du \right\|_T \\ &\quad + \left\| \int_0^{\cdot} \lambda[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du - \int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du \right\|_T \\ &\le 2\lambda T\|q^n-\tilde{q}^n\|_T + \left\| \int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)=\kappa^n\}}\lambda[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du \right\|_T,\end{aligned}$$

where the last inequality follows by applying Lemma A.5. Clearly, the first term above converges to zero a.s. For the second term, we can write

$$\begin{aligned}&\left\| \int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)=\kappa^n\}}\lambda[(1-q_1^n(u)-q_3^n(u))\wedge 1]\,du \right\|_T = \left\| \int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)=\kappa^n\}}\lambda(1-q_1^n(u)-q_3^n(u))\,du \right\|_T \\ &\le \left\| \int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)=\kappa^n\}}[\lambda(1-q_1^n(u)-q_3^n(u))+\nu q_3^n(u)]\,du \right\|_T \\ &\le \|q^n-q^n(0)\|_T + \frac{1}{n}\left\| \int_0^{\cdot} [\lambda_1^n(nq^n(u))+\nu_1^n(nq^n(u))-\mu_1^n(nq^n(u))]\,du \right\|_T \\ &\quad + \frac{1}{n}\left\| \bar{N}_1^a\left(\int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda_1^n(nq^n(u))\,du\right)\right\|_T + \frac{1}{n}\left\| \bar{N}_1^d\left(\int_0^{\cdot} \mu_1^n(nq^n(u))\,du\right)\right\|_T \\ &\quad + \frac{1}{n}\left\| \bar{N}^r\left(\int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\nu nq_3^n(u)\,du\right)\right\|_T + \frac{1}{n}\|\delta^n\|_T,\end{aligned}$$

where the last inequality follows from Equation (6) by dividing both sides by n and rearranging terms and noting that $\zeta_1^n(nq^n(u)) = \lambda(n - nq_1^n(u) - nq_3^n(u)) + \nu nq_3^n(u)$. As $\|q^n-\bar{q}\|_T, \|q^n(0)-\bar{q}\| \to 0$, we obtain $\|q^n-q^n(0)\|_T \to 0$. Repeating the argument in the proof of Proposition 4.3(a), we obtain that the third, fourth, fifth, and sixth terms converge to zero almost surely as well. The second term can be written as

$$\begin{aligned}&\frac{1}{n}\left\| \int_0^{\cdot} [\lambda_1^n(nq^n(u))+\nu_1^n(nq^n(u))-\mu_1^n(nq^n(u))]\,du \right\|_T \\ &= \left\| \int_0^{\cdot} [\lambda(1-q_1^n(u)-q_3^n(u))+\nu q_3^n(u)-\mu q_1^n(u)]\,du \right\|_T \\ &= \left\| \int_0^{\cdot} [(\lambda+\mu)(\bar{q}_1-q_1^n(u))+(\nu-\lambda)q_3^n(u)]\,du \right\|_T \\ &\le (2\lambda+\mu+\nu)T\|q^n-\bar{q}\|_T.\end{aligned}$$

As $\|q^n-\bar{q}\|_T \to 0$, this term converges to zero almost surely. Thus, $\bar{\gamma}^n \to 0$ and using the fact that $\lim_{x\to 0}\sqrt{x(1+\log(\lambda T/x))} = 0$ in Equation (B52), we obtain

$$\frac{1}{\sqrt{n}}\left\| B_1^a\left(\int_0^{t} 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\lambda_1^n(nq^n(u))\,du\right) - B_1^a\left(\int_0^{t} \tilde{\lambda}_1^n(n\tilde{q}^n(u))\,du\right)\right\|_T \Rightarrow 0. \tag{B53}$$

Similarly,

$$\frac{1}{\sqrt{n}}\left\| B^r\left(\int_0^{\cdot} 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}}\nu nq_3^n(u)du\right) - B^r\left(\int_0^{\cdot} n(\tilde{q}_3^n(u)\wedge 1)\nu\,du\right)\right\|_T \Rightarrow 0,$$

and, hence,

$$\left\| \frac{1}{\sqrt{n}}(M_1^{a,n} - \widetilde{M}_1^{a,n})\right\|_T \Rightarrow 0.$$

In a similar manner, we can show that

$$\left\| \frac{1}{\sqrt{n}}(M_i^{a,n} - \widetilde{M}_i^{a,n}) \right\|_T \Rightarrow 0, \quad i = 2, 3.$$

Similarly, we obtain $\|(1/\sqrt{n})(M_i^{d,n} - \widetilde{M}_i^{d,n})\|_T \Rightarrow 0$, $i = 1, 2, 3$. Hence,

$$\left\| \frac{1}{\sqrt{n}}(\alpha^n - \tilde{\alpha}^n) \right\|_T \Rightarrow 0.$$

We can mimic the argument used to derive Equation (B53) from Equation (B52) to obtain

$$\|\sqrt{n}\delta^n\|_T = \frac{1}{\sqrt{n}}\left| B_1^a\left(\int_0^t 1_{\{q_1^n(u)+q_2^n(u)<\kappa^n\}} \lambda_1^n(nq^n(u))\,du\right) - B_1^a\left(\int_0^t \lambda_1^n(nq^n(u))\,du\right)\right| \Rightarrow 0.$$

Using these results in Equation (B51), we obtain for any $\xi > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}(\|\widehat{Q}^n - Q^{*n}\|_T \geq \xi) \\
&\quad\leq \limsup_{n\to\infty} \mathbb{P}(\|\sqrt{n}(\alpha^n - \tilde{\alpha}^n)\|_T + \|\sqrt{n}\delta^n\|_T + \gamma(\epsilon)\|\sqrt{n}\widetilde{Y}^n\|_T \geq C_3\xi) \\
&\quad\leq \limsup_{n\to\infty} \mathbb{P}(\|\sqrt{n}(\alpha^n - \tilde{\alpha}^n)\|_T \geq C_3\xi/3) + \limsup_{n\to\infty} \mathbb{P}(\|\sqrt{n}\delta^n\|_T \geq C_3\xi/3) \\
&\qquad + \limsup_{n\to\infty} \mathbb{P}\left(\|\sqrt{n}\widetilde{Y}^n\|_T \geq \frac{C_3\xi}{3\gamma(\epsilon)}\right) \\
&\quad\leq \limsup_{n\to\infty} \mathbb{P}\left(\|\sqrt{n}\widetilde{Y}^n\|_T \geq \frac{C_3\xi}{3\gamma(\epsilon)}\right),
\end{aligned}$$

where $C_3 = e^{-C_2T}/C_1$ and the weak convergence of the terms $\|\sqrt{n}(\alpha^n - \tilde{\alpha}^n)\|_T$ and $\|\sqrt{n}\delta^n\|_T$ implies $\limsup_{n\to\infty} \mathbb{P}(\|\sqrt{n}(\alpha^n - \tilde{\alpha}^n)\|_T \geq C_3\xi/3) = 0$ and $\limsup_{n\to\infty} \mathbb{P}(\|\sqrt{n}\delta^n\|_T \geq C_3\xi/3) = 0$. Further, noting that ϵ is arbitrary, we have:

$$\limsup_{n\to\infty} \mathbb{P}(\|\widehat{Q}^n - Q^{*n}\|_T \geq \xi) \leq \lim_{\epsilon\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\|\sqrt{n}\widetilde{Y}^n\|_T \geq \frac{C_3\xi}{3\gamma(\epsilon)}\right).$$

Now, comparing $\widetilde{X}^n$ (defined in Equations (37)–(38)) with $\widetilde{Z}^n$ defined in Equations (B28)–(B29), we have $\sqrt{n}(\widetilde{X}^n - \bar{q}) - (k_2, 0, 0)' = \widetilde{Z}^n$, and thus we can write

$$\sqrt{n}\widetilde{Y}^n(t) = \sup_{0\leq s\leq t} \frac{\sqrt{n}(\widetilde{X}_1^n(s) + \widetilde{X}_2^n(s) - (k_1 + k_2/\sqrt{n}))^+}{m + \lambda_1} = \sup_{0\leq s\leq t} \frac{(\widetilde{Z}_1^n(s) + \widetilde{Z}_2^n(s))^+}{m + \lambda_1}.$$

Further, by Lemma B.8, $\widetilde{Z}^n \Rightarrow Z^*$, where Z^* is defined in Equations (46)–(49), and thus using the continuous mapping theorem, we obtain $\sqrt{n}\|\widetilde{Y}^n\|_T = \sqrt{n}\widetilde{Y}^n(T) \Rightarrow \sup_{0\leq s\leq T}(Z_1^*(s) + Z_2^*(s))^+/(m + \lambda_1)$. This implies that

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\|\sqrt{n}\widetilde{Y}^n\|_T \geq \frac{C_3\xi}{3\gamma(\epsilon)}\right) = 0$$

and we obtain $\limsup_{n\to\infty} \mathbb{P}(\|\widehat{Q}^n - Q^{*n}\|_T \geq \xi) = 0$. Thus, we have $\|\widehat{Q}^n - Q^{*n}\|_T \Rightarrow 0$ and the result follows. □

B.3. Proof of results in §6.

B.3.1. Proof of Proposition 6.1. We shall prove that there exist constants t_0, $c_0 > 0$ that are independent of n such that for all sufficiently large n

$$\sup_{z\in\mathbb{R}^2\times\mathbb{R}_+:\, e'|z|\geq c_0\sqrt{n}} \{\mathbb{E}[e'|\hat{q}^n(t_0)| \mid \hat{q}^n(0) = z] - e'|z|\} \leq -\sqrt{n}.$$

In addition, there exists $\beta_0 > 0$ such that

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} \mathbb{E}[\exp(n^{-1/2}\beta_0(e'|\hat{q}^n(t_0)| - e'|z|)^+) \mid \hat{q}^n(0) = z] < \infty \tag{B54}$$

and

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} n^{-1}\mathbb{E}[(e'|\hat{q}^n(t_0)| - e'|z|)^2 \exp(n^{-1/2}\beta_0(e'|\hat{q}^n(t_0)| - e'|z|)^+) \mid \hat{q}^n(0) = z] < \infty, \tag{B55}$$

where $S = \mathbb{R}^2 \times \mathbb{R}_+$. Definition 6.1 then immediately implies the result.

We begin this proof by defining a fluid trajectory $\bar{q}^n(\cdot)$ that starts at $q^n(0)$ and then obtaining bounds on the deviations of the process q^n from this fluid trajectory. Formally, $\bar{q}^n$ is defined as follows.

$$\bar{q}^n = \Phi^n_{\kappa^n}(\bar{x}^n),$$
$$\bar{x}^n = \bar{q}^n(0) + \int_0^t \tilde{\theta}(\bar{q}^n(u))\,du,$$

where Φ is given by Equation (17) and $\tilde{\theta}(x, y, z) = (-(\lambda+\mu)x - (\lambda-\nu)z, -\mu y, -\nu z)'$. $\bar{q}^n$ satisfies the following ordinary differential equation:

$$\frac{d\bar{q}^n_1}{dt} = \lambda - (\lambda+\mu)\bar{q}^n_1 - (\lambda-\nu)\bar{q}^n_3,$$
$$\frac{d\bar{q}^n_2}{dt} = \lambda_1 - \mu\bar{q}^n_2,$$
$$\frac{d\bar{q}^n_3}{dt} = -\nu\bar{q}^n_3,$$
$$\bar{q}^n(0) = \bar{q} - \left(\frac{k_2}{\sqrt{n}}, 0\right)' - \frac{z}{n},$$

which is solved by

$$\begin{aligned}
\bar{q}^n_1(t) &= \bar{q}_1 - \left(\frac{z_1 + k_2\sqrt{n}}{n} + \frac{z_3}{n(\lambda+\mu-\nu)}\right)\exp(-(\lambda+\mu)t) \\
&\quad - \frac{z_3}{n(\lambda+\mu-\nu)}\exp(-\nu t), \\
\bar{q}^n_2(t) &= \bar{q}_2 - \frac{z_2}{n}\exp(-\mu t), \\
\bar{q}^n_3(t) &= \frac{z_3}{n}\exp(-\nu t),
\end{aligned} \tag{B56}$$

if $\nu \neq \lambda+\mu$. For the case $\nu = \lambda+\mu$, the solution is given by

$$\bar{q}^n(t) = \bar{q} - \left(\left(\frac{z_1 + k_2\sqrt{n}}{n} + t\frac{z_3}{n}\right)\exp(-(\lambda+\mu)t), \frac{z_2}{n}\exp(-\mu t), \frac{z_3}{n}\exp(-\nu t)\right)'.$$

The following result will be useful.

Lemma B.9. *The following results hold.*

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} n^{1/2}\mathbb{E}[\|\bar{q}^n - q^n\|_{t_0} \mid \hat{q}^n(0) = z] < \infty, \tag{B57}$$

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} \mathbb{E}[\exp(n^{1/2}\beta_0\|\bar{q}^n - q^n\|_{t_0}) \mid \hat{q}^n(0) = z] < \infty, \tag{B58}$$

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} n\mathbb{E}[\|\bar{q}^n - q^n\|_{t_0}{}^2 \exp(n^{1/2}\beta_0\|\bar{q}^n - q^n\|_{t_0}) \mid \hat{q}^n(0) = z] < \infty. \tag{B59}$$

Proof. We will begin by demonstrating that Equation (B58) implies both Equation (B57) and Equation (B59). To see this, for a fixed $\beta_1 > 0$, choose C large enough so that $\exp(\beta_1 x) > x^2 > x$ for $x > C$. Replacing $n^{1/2}\|\bar{q}^n - q^n\|_{t_0}$ by $\max(C, n^{1/2}\|\bar{q}^n - q^n\|_{t_0})$ to obtain that Equations (B59) and (B57) hold provided the following holds:

$$\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} \mathbb{E}[\exp(2\beta_0 \max(C, n^{1/2}\|\bar{q}^n - q^n\|_{t_0})) \mid \hat{q}^n(0) = z] < \infty,$$

which holds if and only if Equation (B58) holds. Hence, we will focus on proving Equation (B58).

For $i = 1, 2, 3$, we have

$$\begin{aligned}\mathbb{P}\Big(\sup_{0\le t\le t_0} \exp(\beta n^{1/2}|\bar{q}_i^n(t) - q_i^n(t)|) \ge u \mid \hat{q}^n(0) = z\Big)\\ = \mathbb{P}\Big(\sup_{0\le t\le t_0} |\bar{q}_i^n(t) - q_i^n(t)| \ge \beta^{-1} n^{-1/2}\log u \mid \hat{q}^n(0) = z\Big).\end{aligned} \tag{B60}$$

Note that

$$\begin{aligned}\sup_{0\le t\le t_0} |\bar{q}_i^n(t) - q_i^n(t)| &\le \|\bar{q}^n - q^n\|_{t_0} \le \|\bar{q}^n - \bar{q}\|_{t_0} + \|q^n - \bar{q}\|_{t_0}\\ &\le a(t_0)\frac{\lambda_2}{\sqrt{n}} + b(t_0)\|\alpha^n\|_{t_0} + c(t_0)\|\delta^n\|_{t_0}\end{aligned}$$

for some constants $a(t_0), b(t_0), c(t_0) \ge 0$. We use Equation (B5) to bound $\|q^n - \bar{q}\|_{t_0}$ and an analogous argument to bound $\|\bar{q}^n - \bar{q}\|_{t_0}$. Using this bound in Equation (B60), we obtain:

$$\begin{aligned}&\mathbb{P}\Big(\sup_{0\le t\le t_0} \exp(\beta n^{1/2}|\bar{q}_i^n(t) - q_i^n(t)|) \ge u \mid \hat{q}^n(0) = z\Big)\\ &\le \mathbb{P}\Big(a(t_0)\frac{\lambda_2}{\sqrt{n}} + b(t_0)\|\alpha^n\|_{t_0} + c(t_0)\|\delta^n\|_{t_0} \ge \beta^{-1} n^{-1/2}\log u \mid \hat{q}^n(0) = z\Big).\end{aligned}$$

We will now use the strong approximation result to bound the terms $\|\alpha^n\|_{t_0}$ and $\|\delta^n\|_{t_0}$. First, define the independent and identically distributed random variables $X_j^i = \sup_{t>0}(|N_j^i(t) - t - B_j^i(t)|/\log(2\vee t))$, for $i \in \{a, d\}$ and $j = 1, 2$, and $X^r = \sup_{t>0}(|N^r(t) - t - B^r(t)|/\log(2\vee t))$. Lemma A.1 ensures that $\mathbb{E}e^{\eta X} < \infty$ for η small enough, where $X = X_j^i, X^r$.

Observe that the definition of α^n in Equation (14), the fact that q_i^n is bounded by definition, and the strong approximation result imply that there exists a constant $\gamma > 0$ such that $\|n^{1/2}\alpha^n\|_{t_0} \le (1/\sqrt{n})\sum_{i=1}^2(\|B_i^a(\gamma n\cdot)\|_{t_0} + \|B_i^d(\gamma n\cdot)\|_{t_0}) + (1/\sqrt{n})\|B^r(\gamma n\cdot)\|_{t_0} + (\sum_{i,j} X_j^i + X^r)\log(2\vee\gamma n t_0)/\sqrt{n}$ and $\|n^{1/2}\delta^n\|_{t_0} \le (1/\sqrt{n})\|B_1^a(\gamma n\cdot)\|_{t_0} + 2X_1^a\log(2\vee\gamma n t_0)/\sqrt{n}$. This combined with the above relation suggests the following

$$\begin{aligned}&\mathbb{P}\Big(\sup_{0\le t\le t_0} \exp(\beta n^{1/2}|\bar{q}_i^n(t) - q_i^n(t)|) \ge u \mid \hat{q}^n(0) = z\Big)\\ &\le \mathbb{P}\Big(\frac{1}{\sqrt{n}}(b(t_0) + c(t_0))\Big(\sum_{i=1}^2(\|B_i^a(\gamma n\cdot)\|_{t_0} + \|B_i^d(\gamma n\cdot)\|_{t_0}) + \|B^r(\gamma n\cdot)\|_{t_0}\Big) \ge \frac{\beta^{-1}\log u - \lambda_2 a(t_0)}{2}\Big)\\ &\quad + \mathbb{P}\Big((b(t_0) + 2c(t_0))\Big(\sum_{i=1}^2(X_i^a + X_i^d) + X^r\Big) \ge \frac{\beta^{-1}\log u - \lambda_2 a(t_0)}{2}\frac{\sqrt{n}}{\log(2\vee\gamma n t_0)}\Big)\\ &\le 5\mathbb{P}\Big((b(t_0) + c(t_0))\|W(\gamma\cdot)\|_{t_0} \ge \frac{\beta^{-1}\log u - \lambda_2 a(t_0)}{10}\Big)\\ &\quad + 5\mathbb{P}\Big((b(t_0) + 2c(t_0))X_1^a \ge \frac{\beta^{-1}\log u - \lambda_2 a(t_0)}{10}\frac{\sqrt{n}}{\log(2\vee\gamma n t_0)}\Big)\\ &\le C_1\exp(-c_1(\beta^{-1}\log u - \zeta)^2) + K_1\exp\Big(-\eta(\beta^{-1}\log u - \zeta)\frac{\sqrt{n}}{\log(2\vee\gamma n t_0)}\Big),\end{aligned}$$

where $\zeta = \lambda_2 a(t_0)$ and K_1, C_1, c_1 are constants. $W(\cdot)$ is an independent, standard Brownian motion. The last relation holds using tail bounds on the maximum of the Brownian motion and using $\mathbb{E}e^{\eta X_1^a} < \infty$ for some $\eta > 0$. Now,

$$\int_2^\infty C_1\exp(-c_1(\beta^{-1}\log u - \zeta)^2)\,du = \int_2^\infty C_2 u^{-c_1\beta^{-1}(\beta^{-1}\log u - 2\zeta)}\,du < \infty,$$

where C_2 is a constant and the finiteness follows as for $u > u_0 \equiv \exp(2\zeta\beta + 2\beta^2/c_1)$, the exponent of u in the above relation is less than -2. We also observe that

$$\int_{e^{\zeta\beta}}^{\infty} K_1 \exp\left(-\eta(\beta^{-1}\log u - \zeta)\frac{\sqrt{n}}{\log(2\vee \gamma n t_0)}\right) du < \infty.$$

This implies

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} \mathbb{E}\left[\sup_{0\le t\le t_0} \exp(n^{1/2}\beta_0(\bar{q}_i^n(t) - q_i^n(t))^+) \mid \hat{q}^n(0) = z\right]\\
&\le e^{\zeta\beta} + \limsup_{n\to\infty} \int_{e^{\zeta\beta}}^{\infty} K_1 \exp\left(-\eta(\beta^{-1}\log u - \zeta)\frac{\sqrt{n}}{\log(2\vee\gamma n t_0)}\right) du + 2\\
&\quad + \limsup_{n\to\infty} \sup_{z\in\mathbb{R}^2\times\mathbb{R}_+} \int_2^{\infty} \mathbb{P}\left(\sup_{0\le t\le t_0} (\bar{q}_i^n(t) - q_i^n(t))^+ \ge \beta^{-1} n^{1/2}\log u \mid \hat{q}^n(0) = z\right) du\\
&< \infty,
\end{aligned}$$

and, hence,

$$\limsup_{n\to\infty} \sup_{\{z:\, e'z\ge 0\}} \mathbb{E}[\exp(n^{1/2}\beta_0(\|\bar{q}_i^n - q_i^n\|_{t_0})^+) \mid \hat{q}^n(0) = z] < \infty,$$

which concludes the proof of Equation (B58). □

For the remainder of the proof, we shall assume $\nu \ne \lambda + \mu$. The case $\nu = \lambda + \mu$ can be handled in an analogous manner. Using Equation (B57), we obtain the existence of a constant $C_2 > 0$ such that

$$\begin{aligned}
&\mathbb{E}[|[e'|\hat{q}^n(t_0)| - e'|n(\bar{q} - \bar{q}^n(t_0)) - (k_2\sqrt{n}, 0)'|]| \mid \hat{q}^n(0) = z]\\
&\le 2n\mathbb{E}[\|\bar{q}^n - q^n\|_{t_0} \mid \hat{q}^n(0) = z]\\
&\le C_2\sqrt{n}.
\end{aligned}$$

Combining this with Equation (B56), we obtain that

$$\begin{aligned}
\mathbb{E}[e'|\hat{q}^n(t_0)| \mid \hat{q}^n(0) = z] &\le C_2\sqrt{n} + |z_1|\exp(-(\lambda+\mu)t_0) + |z_2|\exp(-\mu t_0)\\
&\quad + z_3\left(\exp(-\nu t_0) + \frac{1}{|\lambda+\mu-\nu|}(\exp(-(\lambda+\mu)t_0) + \exp(-\nu t_0))\right)\\
&\quad + |k_2|\sqrt{n}(1 - \exp(-(\lambda+\mu)t_0))\\
&= e'|z| + \widetilde{C}_2\sqrt{n} - |z_1|(1 - \exp(-(\lambda+\mu)t_0)) - |z_2|(1 - \exp(-\mu t_0))\\
&\quad - |z_3|\left(1 - \left(\exp(-\nu t_0) + \frac{1}{|\lambda+\mu-\nu|}(\exp(-(\lambda+\mu)t_0) + \exp(-\nu t_0))\right)\right)\\
&\le e'|z| + \widetilde{C}_2\sqrt{n} - (|z_1| + |z_2|)(1 - \exp(-\mu t_0))\\
&\quad - |z_3|\left(1 - \left(\exp(-\nu t_0) + \frac{1}{|\lambda+\mu-\nu|}(\exp(-(\lambda+\mu)t_0) + \exp(-\nu t_0))\right)\right),
\end{aligned}$$

where $\widetilde{C}_2 = C_2 + |k_2|(1 - \exp(-(\lambda+\mu)t_0))$. Hence, choosing $c_0 = (\widetilde{C}_2 + 3)/2$ and t_0 such that $c_0\exp(-\mu t_0) \le 1$ and $c_0(\exp(-\nu t_0) + (1/(|\lambda+\mu-\nu|))(\exp(-(\lambda+\mu)t_0) + \exp(-\nu t_0))) \le 1$ and noting that $e'|z| > c_0\sqrt{n}$, we obtain

$$\mathbb{E}[e'|\hat{q}^n(t_0)| \mid \hat{q}^n(0) = z] - e'|z| \le -\sqrt{n}.$$

To prove Equations (B54) and (B55), note that for any $\beta > 0$, the following relation holds.

$$\begin{aligned}
n^{-1/2}\beta(e'|\hat{q}^n(t_0)| - e'|z|)^+ &\le n^{-1/2}\beta(e'|\hat{q}^n(t_0)| - e'|n(\bar{q} - \bar{q}^n(t_0)) - (k_2\sqrt{n}, 0, 0)'|)^+\\
&\quad + n^{-1/2}\beta(e'|n(\bar{q} - \bar{q}^n(t_0)) - (k_2\sqrt{n}, 0, 0)'| - e'|z|)^+\\
&\le 2n^{1/2}\beta\|\bar{q}^n - q^n\|_{t_0} + \beta|k_2|, \qquad \text{(B61)}
\end{aligned}$$

as Equation (B56) implies $(e'|n(\bar{q} - \bar{q}^n(t_0)) - (k_2\sqrt{n}, 0, 0)'| - e'|z|)^+ \le |k_2|\sqrt{n}$. This relation along with Lemma B.9 completes the proof. □

B.3.2. Proof of Proposition 6.2. Using Proposition 6.1 for $\beta_n = c\beta_0 n^{-1/2}$ where $0 < c < 1$ is a constant, we obtain

$$\limsup_{n\to\infty} n^{-1/2}\beta_n L_2(\beta_n, t_0) = \limsup_{n\to\infty} n^{-1} c\beta_0 L_2(\beta_n, t_0) \le 1$$

for c chosen sufficiently small.

Now, applying Lemma A.3 and using $\beta_n = c\beta_0 n^{-1/2}$, we obtain for every $s > 0$

$$P_{\hat{\pi}^n}(e'|\hat{q}^n(0)|/\sqrt{n} > s) \le \frac{L_1(\beta_n, t_0)}{1 - c\beta_0/2}\exp(-c\beta_0(s - c_0)).$$

Equation (50) implies that there exists a sufficiently large $n_0 > 0$ and a constant $c_2 > 0$ such that $L_1(\beta_n, t_0) < c_2$ for $n > n_0$. Then, for all $s > 0$ and all $n > n_0$:

$$P_{\hat{\pi}^n}(e'|\hat{q}^n(0)|/\sqrt{n} > s) \le \frac{c_2}{1 - c\beta_0/2}\exp(-c\beta_0(s - c_0)). \quad \square$$

B.3.3. Proof of Proposition 6.4. Noting that $\widehat{Q}$ has continuous sample paths a.s., we obtain the weak convergence of the finite dimensional distributions of $\widehat{Q}^n$ to that of $\widehat{Q}$ by Theorem 7.8 in Chapter 3 of Ethier and Kurtz [14]. Let $\{\hat{\pi}^{n'}\}$ be a subsequence such that $\hat{\pi}^{n'} \Rightarrow \hat{\pi}'$. Then, the finite dimensional distributions of $\widehat{Q}^n$ with $\widehat{Q}^{n'}(0)$ distributed as $\hat{\pi}^{n'}$ converge weakly to that of $\widehat{Q}$ with $\widehat{Q}(0)$ distributed as $\hat{\pi}'$. As $\widehat{Q}^{n'}$ with $\widehat{Q}^{n'}(0)$ distributed as $\hat{\pi}^{n'}$ is stationary, so is $\widehat{Q}$ with $\widehat{Q}(0)$ distributed as $\hat{\pi}'$, and the first part of the result follows.

Now, using Proposition 6.3, we obtain the existence of an invariant distribution for the diffusion process $\widehat{Q}$. To prove uniqueness, we will show that a discrete time version of $\widehat{Q}$ is ψ-irreducible by first proving a similar property for a related driftless process, and then arguing using Girsanov's theorem. The uniqueness then follows by applying Lemma A.4. The details are as follows.

Define $\widehat{Q}_i^R$ for $i = 1, 2, 3$ by

$$\widehat{Q}_1^R(t) = \widehat{Q}_1(0) + \sqrt{2m}B_1^R(t) - mY^R(t), \tag{B62}$$

$$\widehat{Q}_2^R(t) = \widehat{Q}_2(0) + \sqrt{2\lambda_1}B_2^R(t) - \lambda_1 Y^R(t), \tag{B63}$$

$$\widehat{Q}_3^R(t) = \widehat{Q}_3(0) - \int_0^t \nu \widehat{Q}_3^R(u)\,du + mY^R(t), \tag{B64}$$

where B_1^R and B_2^R denote two independent standard Brownian motions defined on $(\Omega, \mathcal{F}, \mathbb{P})$ along with a filtration $\{\mathcal{F}_t\}$ and Y^R is the nonnegative, nondecreasing, continuous process such that $\widehat{Q}_1^R(t) + \widehat{Q}_2^R(t) \le 0$ and $\int_0^t (\widehat{Q}_1^R(u) + \widehat{Q}_2^R(u))\,dY^R(u) = 0$, $\forall\, t \ge 0$, and $Y^R(0) = 0$. The fact that Equations (B62)–(B64) has a unique strong solution follows analogous to the proof of Lemma B.7, noting that the key difference between $\widehat{Q}^R$ and $\widehat{Q}$ is that $\widehat{Q}_i^R$, $i = 1, 2$ do not have any drift terms. Adding Equation (B62) and Equation (B63), and defining $B_3^R(\cdot) \equiv \sqrt{2m}B_1^R(\cdot) + \sqrt{2\lambda_1}B_2^R(\cdot)$, we note that

$$(m + \lambda_1)Y^R(t) = \sup_{0\le u\le t}(\bar{x}_1 + \bar{x}_2 + B_3^R(u))^+. \tag{B65}$$

Let λ_3 denote the Lebesgue measure on the measurable space $(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3))$, where $\mathcal{B}(\mathbb{R}^3)$ denotes the Borel σ-field on $\mathbb{R}^3$. For $A \in \mathcal{B}(\mathbb{R}^3)$, define $\varphi(A) = \lambda_3(A \cap \{(x, y, z)\colon x + y \le 0, z \ge 0\})$. Further, define $T_n = \inf\{t\colon |B_1^R(t)| \ge n\} \wedge \inf\{t\colon |B_2^R(t)| \ge n\} \wedge \inf\{t\colon Y^R(t) \ge n\}$. Finally, let $\bar{x}$ denote any point that lies in the state space, i.e., $\bar{x} \in \mathbb{R}^3$ and $\bar{x}_1 + \bar{x}_2 \le 0$, $\bar{x}_3 \ge 0$.

Lemma B.10. *For any sequence of time instants $\{t^m\}$ with $t^m \uparrow \infty$ and $A \in \mathcal{B}(\mathbb{R}^3)$, if $\varphi(A) > 0$, there exist n, $j \in \mathbb{N}$ such that $\mathbb{P}(\widehat{Q}^R(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) > 0$.*

Proof. Using the mapping $\Gamma\colon (x, y, z) \to (x, x+y, z)$ for any $x, y, z \in \mathbb{R}$, it suffices to prove that $\mathbb{P}(\Gamma\widehat{Q}^R(t^j) \in \Gamma A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) > 0$ if $\varphi(A) > 0$. Using Equation (B65), we shall write $Y^R(t) = (1/(m+\lambda_1))L^R(t)$, where $L^R(t) = \sup_{0\le u\le t}(\bar{x}_1 + \bar{x}_2 + B_3^R(u))^+$. Thus, given $\widehat{Q}(0) = \bar{x}$, $\Gamma\widehat{Q}^R(t) = (\bar{x}_1 + \sqrt{2m}B_1^R(t) - m/(m+\lambda_1)\cdot L^R(t), \bar{x}_1 + \bar{x}_2 + B_3^R(t) - L^R(t), \widehat{Q}_3^R(t))$, where $\widehat{Q}_3^R(t)$ is given in Equation (B64).

Assume $t^0 = 0$ without loss of generality. It suffices to consider the case where there exists $(z_1, z_2, z_3) \in \Gamma A$ such that $z_2 \le 0$, $z_3 \ge 0$, and for some $\delta > 0$, $\tilde{A} \equiv [z_1 - \delta, z_1 + \delta] \times [z_2 - \delta, z_2 + \delta] \times [z_3 - \delta, z_3 + \delta] \subset \Gamma A$. Note that we can write $\widehat{Q}_3^R(t) = \bar{x}_3 e^{-\nu t} + (m/(m+\lambda_1))(L^R(t) - \nu e^{-\nu t}\int_0^t L^R(s)e^{\nu s}\,ds)$. So, there exist $\epsilon, y_3 > 0$ with $\epsilon < \delta/2$ and $y_3 > \epsilon$, and $t^j > \epsilon$ such that $L^R(\epsilon) \in [y_3 - \epsilon, y_3 + \epsilon]$ and $L^R(t^j) \le$

$L^R(\epsilon)+\delta/2$ ensures $\widehat{Q}_3^R(t^j)\in[z_3-\delta, z_3+\delta]$. Choose any $n>(|\bar{x}_1|+|z_1|+y_3+3(\delta/2)+\epsilon)/\sqrt{2m}+(|\bar{x}_1|+|\bar{x}_2|+|z_2|+y_3+\delta+\epsilon)/(\sqrt{2m}+\sqrt{2\lambda_1})+(y_3+\delta+\epsilon)/(m+\lambda_1)$. So, we can write

$$\begin{aligned}
&\mathbb{P}(\Gamma\widehat{Q}^R(t^j)\in\tilde{A}, T_n\geq t^j \mid \widehat{Q}(0)=\bar{x})\\
&\geq \mathbb{P}\Big(\bar{x}_1+\sqrt{2m}B_1^R(t^j)\in[z_1-\delta+(y_3+\epsilon+\delta/2)m/(m+\lambda_1), z_1+\delta+(y_3-\epsilon)m/(m+\lambda_1)],\\
&\qquad \bar{x}_1+\bar{x}_2+B_3^R(t^j)\in[z_2-\delta/2+y_3+\epsilon, z_2+\delta+y_3-\epsilon], L^R(\epsilon)\in[y_3-\epsilon, y_3+\epsilon], L^R(t^j)\leq L^R(\epsilon)+\delta/2,\\
&\qquad \|B_1^R\|_{t^j}\leq n, \|B_2^R\|_{t^j}\leq n\Big)\\
&=\mathbb{P}\Big(\bar{x}_1+\sqrt{2m}B_1^R(t^j)\in[z_1-\delta+(y_3+\epsilon+\delta/2)m/(m+\lambda_1), z_1+\delta+(y_3-\epsilon)m/(m+\lambda_1)],\\
&\qquad \bar{x}_1+\bar{x}_2+B_3^R(t^j)\in[z_2-\delta/2+y_3+\epsilon, z_2+\delta+y_3-\epsilon], \sup_{0\leq u\leq\epsilon}(\bar{x}_1+\bar{x}_2+B_3^R(u))^+\in[y_3-\epsilon, y_3+\epsilon],\\
&\qquad \sup_{0\leq u\leq t^j}(\bar{x}_1+\bar{x}_2+B_3^R(u))^+\leq \sup_{0\leq u\leq\epsilon}(\bar{x}_1+\bar{x}_2+B_3^R(u))^+ +\delta/2, \|B_1^R\|_{t^j}\leq n, \|B_2^R\|_{t^j}\leq n\Big)\\
&\geq \mathbb{P}\Big(\bar{x}_1+\sqrt{2m}B_1^R(t^j)\in[\hat{z}_{11},\hat{z}_{12}], \bar{x}_1+\bar{x}_2+B_3^R(t^j)\in[\hat{z}_{21},\hat{z}_{22}], \bar{x}_1+\bar{x}_2+B_3^R(\epsilon)\in[y_3-\epsilon, y_3+\epsilon],\\
&\qquad \|\bar{x}_1+\bar{x}_2+B_3^R(\cdot)\|_\epsilon\leq y_3+\epsilon, \sup_{\epsilon\leq u\leq t^j}B_3^R(u)\leq B_3^R(\epsilon)+\delta/2, \|B_1^R\|_{t^j}\leq n, \|B_2^R\|_{t^j}\leq n\Big),
\end{aligned} \tag{B66}$$

where $\hat{z}_{11}=z_1-\delta+(y_3+\epsilon+\delta/2)m/(m+\lambda_1)$, $\hat{z}_{12}=z_1+\delta+(y_3-\epsilon)m/(m+\lambda_1)$, $\hat{z}_{21}=z_2-\delta/2+y_3+\epsilon$, $\hat{z}_2=z_2+\delta+y_3-\epsilon$. Then, there exist $h_1, h_2\in C_{\mathbb{R}}[0,\infty)$, the space of $\mathbb{R}$-valued continuous functions, and $\eta>0$ such that

$$\begin{gathered}
\bar{x}_1+\sqrt{2m}h_1(t^j)\in(\hat{z}_{11},\hat{z}_{12}), \qquad \bar{x}_1+\bar{x}_2+\sqrt{2m}h_1(t^j)+\sqrt{2\lambda_1}h_2(t^j)\in(\hat{z}_{21},\hat{z}_{22}),\\
\bar{x}_1+\bar{x}_2+\sqrt{2m}h_1(\epsilon)+\sqrt{2\lambda_1}h_2(\epsilon)\in(y_3-\epsilon, y_3+\epsilon), \qquad \|\bar{x}_1+\bar{x}_2+\sqrt{2m}h_1+\sqrt{2\lambda_1}h_2\|_\epsilon<y_3+\epsilon,\\
\sup_{\epsilon\leq u\leq t^j}(\sqrt{2m}h_1(u)+\sqrt{2\lambda_1}h_2(u))<\sqrt{2m}h_1(\epsilon)+\sqrt{2\lambda_1}h_2(\epsilon)+\delta/2, \quad \|h_1\|_{t^j}<n, \quad \|h_2\|_{t^j}<n,
\end{gathered}$$

and for any realization $\omega\in\Omega$ such that $\max_{i=1,2}\|B_i^R-h_i\|_{t^j}\leq\eta$, we have

$$\begin{gathered}
\bar{x}_1+\sqrt{2m}B_1^R(t^j)\in[\hat{z}_{11},\hat{z}_{12}], \qquad \bar{x}_1+\bar{x}_2+B_3^R(t^j)\in[\hat{z}_{21},\hat{z}_{22}], \qquad \bar{x}_1+\bar{x}_2+B_3^R(\epsilon)\in[y_3-\epsilon, y_3+\epsilon],\\
\|\bar{x}_1+\bar{x}_2+B_3^R(\cdot)\|_\epsilon\leq y_3+\epsilon, \qquad \sup_{\epsilon\leq u\leq t^j}B_3^R(u)\leq B_3^R(\epsilon)+\delta/2, \quad \|B_1^R\|_{t^j}\leq n, \quad \|B_2^R\|_{t^j}\leq n.
\end{gathered}$$

Combining this with Theorem 5.4 of Durrett [12] in Equation (B66), we obtain

$$\mathbb{P}(\Gamma\widehat{Q}^R(t^j)\in\tilde{A}, T_n\geq t^j\mid\widehat{Q}(0)=\bar{x})>0. \quad \square$$

LEMMA B.11. *For every $n=1,2,\ldots$, there exists a probability measure $\mathbb{P}_n'$ such that for the processes defined by*:

$$B_1(t)\equiv B_1^R(t\wedge T_n)-\frac{1}{\sqrt{2m}}\int_0^{t\wedge T_n}[(\nu-\lambda)\widehat{Q}_3^R(u)-(\lambda+\mu)(\widehat{Q}_1^R(u)+k_2)]\,du, \qquad \text{and} \tag{B67}$$

$$B_2(t)\equiv B_2^R(t\wedge T_n)-\frac{1}{\sqrt{2\lambda_1}}\int_0^{t\wedge T_n}(\lambda_2-\mu\widehat{Q}_2^R(u))\,du, \tag{B68}$$

$\{(B_1(t),B_2(t)),\mathcal{F}_t; t\geq 0\}$ is a two-dimensional standard Brownian motion stopped at T_n on $(\Omega,\mathcal{F}_{T_n},\mathbb{P}_n')$. Further, for each $n=1,2,\ldots$, the measures $\mathbb{P}_n'$ and $\mathbb{P}$ are mutually absolutely continuous.

PROOF. For convenience, denote $X_1^R(t)=(1/\sqrt{2m})[(\nu-\lambda)\widehat{Q}_3^R(t)-(\lambda+\mu)(\widehat{Q}_1^R(t)+k_2)]$ and $X_2^R(t)=(1/\sqrt{2\lambda_1})(\lambda_2-\mu\widehat{Q}_2^R(t))$. Define

$$Z(X^R, t\wedge T_n)\equiv\exp\bigg(\sum_{i=1}^2\int_0^{t\wedge T_n}X_i^R(s)\,dB_i^R-\frac{1}{2}\int_0^{t\wedge T_n}((X_1^R(s))^2+(X_2^R(s))^2)\,ds\bigg)$$

and $Z(X^R, T_n) = \lim_{t\to\infty} Z(X^R, t \wedge T_n)$. Define $\mathbb{P}'_n$ on $(\Omega, \mathcal{F}_{T_n})$ by

$$\frac{d\mathbb{P}'_n}{d\mathbb{P}} = Z(X^R, T_n).$$

As $Z(X^R, t \wedge T_n)$ is an L^2-bounded martingale, applying the Cameron-Martin-Girsanov theorem (cf. Theorem 38.5 of Rogers and Williams [36]) completes the proof. □

Lemma B.12. *For any sequence of time instants $\{t^m\}$ with $t^m \uparrow \infty$, the Markov chain $\widehat{Q}^m \equiv \widehat{Q}(t^m)$ is ψ-irreducible.*

Proof. Pick any set $A \in \mathcal{B}(\mathbb{R}^3)$ such that $\varphi(A) > 0$. We prove that there exists $j \in \mathbb{N}$ such that $\mathbb{P}(\widehat{Q}(t^j) \in A \mid \widehat{Q}(0) = \bar{x}) > 0$. Then, applying Theorem 4.0.1 of Meyn and Tweedie [30], the result follows.

Applying Lemma B.10, we obtain the existence of some $n, j \in \mathbb{N}$ such that $\mathbb{P}(\widehat{Q}^R(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) > 0$. Let $\mathbb{P}'_n$ be the measure obtained in Lemma B.11. Then, as $\mathbb{P}'_n$ and $\mathbb{P}$ are mutually absolutely continuous, we obtain

$$\mathbb{P}'_n(\widehat{Q}^R(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) > 0. \tag{B69}$$

Note that for $t \le T_n$ and B_1, B_2 as in Equations (B67)–(B68), we can rewrite Equations (B62)–(B63) as

$$\widehat{Q}^R_1(t) = \widehat{Q}_1(0) + \sqrt{2m}B_1(t) - \frac{1}{\sqrt{2m}}\int_0^t [(\nu-\lambda)\widehat{Q}^R_3(u) - (\lambda+\mu)(\widehat{Q}^R_1(u)+k_2)]\,du - mY^R(t),$$

$$\widehat{Q}^R_2(t) = \widehat{Q}_2(0) + \sqrt{2\lambda_1}B_2(t) + \frac{1}{\sqrt{2\lambda_1}}\int_0^t (\lambda_2 - \mu\widehat{Q}^R_2(u))\,du - \lambda_1 Y^R(t).$$

Now, using Lemma B.11 along with the fact that Equations (22)–(24) has a unique strong solution, this implies that $\{\widehat{Q}^R(t)\colon 0 \le t \le T_n\}$ under the measure $\mathbb{P}'_n$ has the same law as $\{\widehat{Q}(t)\colon 0 \le t \le T_n\}$ under the measure $\mathbb{P}$. Thus, we have $\mathbb{P}'_n(\widehat{Q}^R(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) = \mathbb{P}(\widehat{Q}(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x})$, which combined with Equation (B69) gives us $\mathbb{P}(\widehat{Q}(t^j) \in A, T_n \ge t^j \mid \widehat{Q}(0) = \bar{x}) > 0$, and the result follows. □

The desired uniqueness now follows by applying Lemmas B.12 and A.4. □

References

[1] Atar, R., A. Budhiraja, P. Dupuis. 2001. On positive recurrence of constrained diffusion processes. *Ann. Probab.* **29**(2) 979–1000.
[2] Billingsley, P. 1986. *Probability and Measure*. John Wiley and Sons, New York.
[3] Billingsley, P. 1995. *Convergence of Probability Measures*, 2nd ed. John Wiley and Sons, New York.
[4] Chen, H., H. Zhang. 2000. Diffusion approximations for some multiclass queueing networks with FIFO service disciplines. *Math. Oper. Res.* **25**(4) 679–707.
[5] Coddington, E. A. 1961. *An Introduction to Ordinary Differential Equations*. Prentice-Hall, Englewood Cliffs, NJ.
[6] Coddington, E. A., N. Levinson. 1955. *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.
[7] Cohen, J. W. 1957. The generalized engset formulae. *Philips Telecomm. Rev.* **18**(4) 158–170.
[8] Dai, J. G., B. Prabhakar. 2000. The throughput of data switches with and without speedup. *IEEE INFOCOM* **2** 556–564.
[9] Dai, J. G., R. J. Williams. 1995. Existence and uniqueness of semimartingale reflecting Brownian motions in convex polyhedrons. *Theory Probab. Appl.* **50** 3–53.
[10] Das, A., R. Srikant. 2000. Diffusion approximations for a single node accessed by congestion controlled sources. *IEEE Trans. Automatic Control* **45**(10) 1783–1799.
[11] de Véricourt, F., O. B. Jennings. 2008. Dimensioning large-scale membership services. *Oper. Res.* **56**(1) 173–187.
[12] Durrett, R. 1996. *Stochastic Calculus: A Practical Introduction*. CRC Press, Boca Raton, FL.
[13] Erlang, A. K. 1917. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Electroteknikeren* **13** 5–13.
[14] Ethier, S. N., T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, New York.
[15] Gamarnik, D., A. Zeevi. 2006. Validity of heavy traffic steady-state approximations in generalized Jackson networks. *Ann. Appl. Probab.* **16** 56–90.
[16] Gross, D., C. Harris. 1998. *Fundamentals of Queueing Theory*. John Wiley and Sons, New York.
[17] Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* **29** 567–588.
[18] Harrison, J. M., R. J. Williams. 1996. A multiclass closed queueing network with unconventional heavy traffic behavior. *Ann. Appl. Probab.* **6**(1) 1–47.
[19] Hunt, P. J., F. P. Kelly. 1989. On critically loaded loss networks. *Adv. Appl. Probab.* **21** 831–841.
[20] Kelly, F. P. 1979. *Reversibility and Stochastic Networks*. John Wiley and Sons, New York.
[21] Kleinrock, L. 1975. *Queueing Systems*, Vol. I: *Theory*. John Wiley and Sons, New York.
[22] Kogan, Y. A., R. S. Lipster, A. V. Smorodinskii. 1986. Gaussian diffusion approximations of closed Markov models of computer networks. *Problems Inform. Transmission* **22** 38–51.
[23] Krichagina, E. V., A. A. Puhalskii. 1997. A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. *Queueing Systems: Theory Appl.* **25** 235–280.

[24] Kumar, S. 2000. Two-server closed networks in heavy traffic: Diffusion limits and asymptotic optimality. *Ann. Appl. Probab.* **10**(3) 930–961.
[25] Kurtz, T. 1978. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.* **6** 223–240.
[26] Maglaras, C., A. Zeevi. 2003. Pricing and capacity sizing for systems with shared resources: Approximate solutions and scaling relations. *Management Sci.* **49** 1018–1038.
[27] Mandelbaum, A., G. Pats. 1998. State-dependent stochastic networks. Part I: Approximations and applications with continuous diffusion limits. *Ann. Appl. Probab.* **8**(2) 569–646.
[28] Mandelbaum, A., W. A. Massey, M. I. Reiman. 1998. Strong approximations for Markovian service networks. *Queueing Systems* **30** 149–201.
[29] McKean, H. P. J. 1969. *Stochastic Integrals*. Academic Press, London.
[30] Meyn, S. P., R. L. Tweedie. 1993. *Markov Chains and Stochastic Stability*. Springer-Verlag, London.
[31] Miyazawa, M. 1993. Insensitivity and product form decomposability of reallocatable GSMP. *Adv. Appl. Probab.* **25** 415–437.
[32] Puhalskii, A. A., M. I. Reiman. 1998. A critically loaded multirate link with trunk reservation. *Queueing Systems: Theory Appl.* **28** 157–190.
[33] Randhawa, R. S. 2006. Operations with subscribers. Ph.D. thesis, Stanford University, Palo Alto, CA.
[34] Randhawa, R. S., S. Kumar. 2008. Usage restriction and subscription services: Operational benefits with rational users. *Manufacturing Service Oper. Management* **10**(3) 429–447.
[35] Reiman, M. I. 1991. A critically loaded multiclass Erlang loss system. *Queueing Systems* **9** 65–82.
[36] Rogers, L. C. G., D. Williams. 2000. *Diffusions, Markov Processes and Martingales*, Vol. 2, 2nd ed. Cambridge University Press, New York.
[37] Royden, H. L. 1988. *Real Analysis*. Prentice Hall, Englewood Cliffs, NJ.
[38] Savin, S. V., M. A. Cohen, N. Gans, Z. Katalan. 2005. Capacity management in rental businesses with two customer bases. *Oper. Res.* **53**(4) 617–631.