Approximating the GI/GI/1+GI Queue with a Nonlinear Drift Diffusion: Hazard Rate Scaling in Heavy Traffic

J. E Reed
Amy R. Ward
School of Industrial and Systems Engineering
Georgia Institute of Technology

May 17, 2007

Abstract

We study a single server queue, operating under the FIFO service discipline, in which each customer independently abandons the queue if his service has not begun within a generally distributed amount of time. Under some mild conditions on the abandonment distribution, we identify a limiting heavy traffic regime in which the resulting diffusion approximation for both the offered waiting time process (the process that tracks the amount of time an infinitely patient arriving customer would wait for service) and the queue length process contains the entire abandonment distribution. In order to use a continuous mapping approach to establish our weak convergence results, we additionally develop existence, uniqueness, and continuity results for non-linear generalized regulator mappings that are of independent interest. We further perform a simulation study to evaluate the quality of the proposed approximations for the steady-state mean queue-length and the steady-state probability of abandonment suggested by the limiting diffusion process.

Keywords: Abandonment, Deadlines, Reneging, Customer Impatience, Diffusion Approximations, Generalized Regulator Mappings

1 Introduction

Although many standard queueing models assume that customers are infinitely patient while waiting for service, it is often the case that such an assumption is not reasonable. In particular, impatient customers faced with long waiting times often evidence their frustration by abandoning the system before completing service. For example, call center callers placed on hold frequently hang up while waiting for an agent to assist them, and web browser users often cancel their viewing requests in the face of long download times.

To the best of our knowledge, the first person to remark on the importance of incorporating customer abandonment in a queueing model was Palm [24], who witnessed the impatient behavior of telephone switchboard customers. More recent studies have established that Markovian

M/M/n+M abandonment models (where the final +M represents the abandonment distribution) are well-approximated by diffusion processes in heavy traffic, both when the number of servers grows large (see Garnett et al [9]), and when the number of servers remains fixed (see Ward and Glynn [33]). However, in reality, it is often the case that customer abandonment times are not exponentially distributed, as exhibited by the study of Brown et al [4] of a bank call center data set. Although stability results for models in very general frameworks exist (see, for example, Stanford [30], Baccelli, Boyer, and Hebuterne [2], Lillo and Martin [22] and Bambos and Ward [32]), the problem of rigorously establishing diffusion approximations for systems with general abandonment time distributions still remains an area of active research. Zeltyn and Mandelbaum [39] consider many server systems with generally distributed abandonment times, and Ward and Glynn [35] consider a single-server system with generally distributed abandonment times. Both works develop a heavy traffic asymptotic regime in which the limiting results depend on the abandonment distribution only through the value of its density at 0.

The value of the density of a distribution at a single point is not a very robust statistic. For example, the estimated hazard rate function associated with the abandonment times of a U.S. bank's call center customers displayed in Figure 12 in the Internet Supplement to Zeltyn and Mandelbaum [39] is unstable near the origin. (Note that because abandonment times are non-negative, the values of both the hazard rate function and density associated with the abandonment distribution coincide at the point 0.) Therefore, identifying a limiting regime that preserves more of the structure of the abandonment distribution is of interest.

The main contribution of this paper is to rigorously establish a heavy traffic regime for a single server queue operating under the FIFO service discipline with renewal arrivals, general service times, and general abandonment times in which the *entire* abandonment time distribution appears in the limiting diffusion approximation. We study both unbounded and bounded abandonment distributions, and develop diffusion approximations for both the offered waiting time process (the process that tracks the amount of time an infinitely patient arriving customer would wait for service) and the queue-length process. In so doing, we also prove results on the existence, uniqueness, and continuity of generalizations of the one-sided regulator mapping introduced in Skorokhod [29] and the two-sided regulator mapping having an explicit formula given in Kruk et al [17] to allow for a general, non-linear state dependence. Our key insight is to model customer abandonment times using the hazard rate function associated with the assumed abandonment time distribution, as suggested by Whitt [38].

To specify our proposed diffusion approximation for the offered waiting time and queue-length processes (for simplicity we assume mean service times are one so that the proposed approximations for these two processes are identical), consider a single-server, FIFO queue having renewal arrival and service processes with identical rates n[1], in which each customer independently abandons the system if his service has not begun within an amount of time having a distribution with hazard rate function h and cumulative hazard function $H \equiv \int_0^x h(y)dy$. Let

$$H_D^n(x) \equiv \int_0^x h\left(\frac{y}{\sqrt{n}}\right) dy = \sqrt{n} H\left(\frac{x}{\sqrt{n}}\right). \tag{1.1}$$

Then, our suggested diffusion approximation for the scaled queue-length process $n^{-1/2}Q^n(\cdot)$

[1]Our analysis does not use the assumption of perfect balance; however, having equal arrival and service rates eases the exposition in the Introduction.

p	E[queue-length] Simulated	Approximated	% Error	P[abandon] Simulated	Approximated	% Error
0.5	9.0093	8.418	6.57%	0.041292	0.043202	4.63%
2	84.911	86.835	2.27%	0.003367	0.003273	2.80%

Table 1: A comparison of the simulated mean queue-length and abandonment probability for a GI/GI/1-GI queue with Poisson arrivals at rate 2500 per unit, deterministic service with mean 1/2500, and abandonment times distributed according to a gamma distribution with scale and shape parameter p.

when n is large has infinitesimal drift $-H_D^n(x)$, where x is the state of the diffusion, constant infinitesimal variance that depends on the variance of the inter-arrival and service times, and is instantaneously reflected at the origin. When the distribution of abandonment times is bounded, our suggested approximating diffusion also has an upper reflecting barrier.

Table 1 displays the results of using our approximation to estimate the mean queue-length and abandonment probability in a queue with Poisson arrivals having rate $n = 2500$, deterministic service times having mean 1/2500, and abandonment times distributed according to $G(p)$, a mean 1 gamma distribution having both scale and shape parameters equalling p. The cumulative hazard function associated with $G(p)$ is

$$H(x) \equiv -\ln\left(1 - \frac{\Gamma_{px}(p)}{\Gamma(p)}\right),$$

where $\Gamma(p) \equiv \int_0^\infty t^{p-1}e^{-t}dt$ is the gamma function, and $\Gamma_x(p) \equiv \int_0^x t^{p-1}e^{-t}dt$ is the incomplete gamma function ($p > 0$). From (1.1), the infinitesimal drift of our suggested approximating diffusion is

$$-H_D^n(x) = \sqrt{n}\ln\left(1 - \frac{\Gamma_{px/\sqrt{n}}(p)}{\Gamma(p)}\right). \tag{1.2}$$

Its inftinitesimal variance is 1, which follows from Theorem 2, and its steady-state distribution is given in part (i) of Proposition 5. We ran each simulation to 2,000 time units so that the queue saw approximately 5,000,000 arrivals, and recorded the time average queue-length and abandonment fraction. Observe that all of our approximations in Table 1 differ from their simulated values by no more than 7%.

Our choices of gamma distributions are motivated by the aforementioned study of a bank call center in the Internet Supplement to [39] showing an estimated hazard rate function of customer abandonment times that is unstable near the origin. The hazard rate function associated with a $G(0.5)$ distribution is also unstable near the origin. We chose the $G(2)$ distribution, whose hazard rate is stable near the origin, for comparison purposes. For the reader's convenience, in Figure 1, we plot the hazard rate functions associated with both the $G(0.5)$ and $G(2)$ distributions near the origin.

The astute reader may have noticed that:

1. the approximations when abandonment times are $G(2)$ have better accuracy than when abandonment times are $G(0.5)$, and

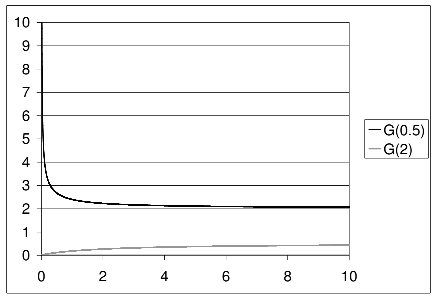

Figure 1: G(p) hazard rate functions

2. the variability of a $G(0.5)$ random variable exceeds that of a $G(2)$ random variable.

The likely reason that our approximation performs better for a less variable abandonment distribution is as follows. Our asymptotic regime is one in which abandonment times are large compared to inter-arrival and service times, and the fraction of abandoning customers becomes small. For two abandonment distributions with the same mean but different variability, the distribution with the higher variability will have more short abandonment times. Hence, we expect a higher percentage of customers to abandon under the abandonment time distribution with higher variability. Our more extensive numeric results in Section 6.3 support the qualitative statement that for less variable distributions, mean abandonment times need not be so much larger than mean inter-arrival times for a good approximation, while for highly variable distributions, mean abandonment times must be much larger.

The remainder of this paper is organized as follows. We formulate our model in Section 2. In Section 3, we describe our heavy traffic asymptotic regime and discuss our assumed hazard rate scaling. We prove existence, uniqueness, and continuity results on one- and two-sided non-linear generalized regulator mappings in Section 4, and use these results in Section 5 to establish weak convergence results for the offered waiting time process. Finally, in Section 6, we establish an asymptotic relationship between the queue-length and offered waiting time processes, provide the stationary distributions for our limiting diffusion approximations, and perform a simulation study to estimate the quality of our proposed queue-length process approximation. The proofs of all our Lemmas can be found in the Appendix.

2 Model Formulation

Our study of the GI/GI/1 queue having FIFO service and customer abandonments begins with the model introduced in Ward and Glynn [35]. The model primitives are three independent i.i.d. sequences of non-negative random variables $\{u_i, i \geq 1\}, \{v_i, i \geq 1\}$, and $\{a_i, i \geq 1\}$, which are all defined on a common probability space $(\Omega, \mathcal{F}, P)$. We assume that $E[u_1] = E[v_1] = 1$ and $\text{var}(u_1) < \infty, \text{var}(v_1) < \infty$. For a given arrival rate ρ, the ith system arrival joins the queue at time

$$t_i \equiv \sum_{j=1}^{i} \frac{u_j}{\rho},$$

has service time v_i, and will abandon if processing does not begin within a_i time units. (For the interested reader, we note that more sophisticated models of customer impatience can be found in Mandelbaum and Shimkin [23] and Zohar, Mandelbaum, and Shimkin [40].) We let F be the cumulative distribution function of a_1, and

$$h(x) = \frac{\frac{d}{dx}F(x)}{1-F(x)},\ x \geq 0$$

be the associated hazard rate function. We assume F is proper; i.e., that $\lim_{x\to\infty} F(x) = 1$. Then, Lemma 2 in Baccelli, Boyer, and Hebuterne [2] guarantees the offered waiting time process given in (2.1) below possesses a non-degenerate limiting distribution.

The length of time a customer arriving at time t has to wait for service depends upon the processing times of the customers in the queue at time t who eventually receive service (and do not abandon). In particular, for $t > 0$, the *offered waiting time* process

$$V(t) \equiv \sum_{n=1}^{A(t)} v_n \mathbf{1}\{V(t_n^-) < a_n\} - B(t) \geq 0 \tag{2.1}$$

tracks the waiting time an infinitely patient arriving customer would experience at time $t > 0$. Here, the process

$$A(t) \equiv \max\left\{ i \geq 0 : \sum_{j=1}^{i} u_j \leq \rho t \right\}$$

counts the number of customers that have arrived to the system by time $t \geq 0$ and the process

$$B(t) \equiv \int_0^t \mathbf{1}\{V(s) > 0\} ds$$

is the cumulative server busy time.

It is useful for our analysis to represent the offered waiting time process in terms of a stochastic integral and three martingales as follows. Define the σ-field

$$\mathcal{F}_i \equiv \sigma((u_1, v_1, a_1), ..., (u_i, v_i, a_i), u_{i+1}) \subset \mathcal{F}$$

such that

$$P(V(t_i^-) \geq a_i | \mathcal{F}_{i-1}) = F\left(V(t_i^-)\right),\ \ i = 1, 2, \ldots,$$

almost surely, because $V(t_i^-)$ is $\mathcal{F}_{i-1}$ measurable and a_i is independent of $\mathcal{F}_{i-1}$. The martingale

$$\left\{ \left(M_a(i) \equiv \sum_{j=1}^{i} \mathbf{1}\{V(t_j^-) \geq a_j\} - E[\mathbf{1}\{V(t_j^-) \geq a_j\} | \mathcal{F}_{j-1}], \mathcal{F}_i \right), i \geq 0 \right\} \tag{2.2}$$

is the sum of the random variables representing which customers abandon, centered by their conditional means. Also let

$$S(i) \equiv \sum_{j=1}^{i} (v_j - E[v_1])$$

be the sum of the centered service times and

$$S_a(i) \equiv \sum_{j=1}^{i} (v_j - E[v_1]) \, \mathbf{1}\{V(t_i^-) \geq a_i\}$$

be the sum of the centered service times of those customers that will eventually abandon. Define the centered process

$$X(t) \equiv E[v_1]A(t) - \rho t + S(A(t)) + t\,(\rho - 1) - S_a(A(t)) - E[v_1]M_a(A(t)), \tag{2.3}$$

and the "integral error" process

$$\epsilon(t) \equiv \int_0^t \left(\int_0^{V(s^-)} h(u)du \right) ds - E[v_1] \int_0^t F(V(s^-))dA(s). \tag{2.4}$$

(Note that even though $E[v_1] = 1$, we explicitly write $E[v_1]$ in (2.3) because its presence will be important in our heavy traffic regime defined in Section 3, where service times are scaled to become small.) Algebraic manipulations of (2.1) show that

$$V(t) = X(t) + \epsilon(t) - \int_0^t \left(\int_0^{V(s^-)} h(u)du \right) ds + I(t), \tag{2.5}$$

where

$$I(t) \equiv t - B(t) = \int_0^t \mathbf{1}\{V(s) = 0\}ds \tag{2.6}$$

is the cumulative server idle time.

We first perform our asymptotic analysis under the assumption that the abandonment distribution has support on the positive real line.

Assumption 1 *The abandonment distribution F may be expressed as*

$$F(x) = 1 - \exp\left(- \int_0^x h(u)du \right), \text{ for } x \geq 0$$

where h is a non-negative and continuous function on $[0, \infty)$.

We then extend our analysis to include abandonment distributions having compact support.

Assumption 2 *The abandonment distribution F may be expressed as*

$$F(x) = \left(1 - \exp\left(- \int_0^{x \wedge C} h(u)du \right) \right) + b\mathbf{1}\{x \geq C\}, \; x \geq 0,$$

where h is a non-negative and continuous function on $[0, C]$, and $b \equiv \exp\left(- \int_0^C h(u)du \right)$.

Assumption 2 allows distributions such as the deterministic distribution, for which $h(x) = 0,\ x < C$, and $F(x) = \mathbf{1}\{x \geq C\}$, but does not allow distributions such as the uniform distribution on $[0, 1]$, for which $h(x) = (1-x)^{-1} \to \infty$ as $x \uparrow 1$. For technical reasons, we avoid distributions whose hazard rate tends to infinity on its support.

3 Hazard Rate Scaling in Heavy Traffic

We first define our heavy traffic asymptotic regime in Subsection 3.1. Subsection 3.2 provides intuition for our assumed hazard rate scaling, and Subsection 3.3 discusses its implications in terms of customer patience.

3.1 The Heavy Traffic Asymptotic Regime

We consider a sequence of systems indexed by $n \geq 1$ in which the arrival rates become large and service times small. Our convention is to superscript any process or quantity associated with the n^{th} system by n. Specifically, the n^{th} system has arrival rate $n\rho^n$. That is, the i^{th} arrival to the n^{th} system occurs at time

$$t_i^n \equiv \sum_{j=1}^{i} \frac{u_j}{n\rho^n},$$

and the cumulative number of customer arrivals in $[0, t]$ in the n^{th} system is given by

$$A^n(t) = \max\{i \geq 0, t_i^n \leq t\}, \quad t \geq 0.$$

The service time of the i^{th} arrival is

$$v_i^n \equiv v_i/n, \tag{3.1}$$

so that the sum of centered service times becomes

$$S^n(i) = \frac{1}{n} \sum_{j=1}^{i} (v_j - E[v_1]). \tag{3.2}$$

As n increases, the mean arrival and service rates become arbitrarily close; in particular,

$$\sqrt{n}(\rho^n - 1) \to \theta, \text{ as } n \to \infty, \tag{3.3}$$

where $\theta \in \Re$.

We scale the hazard rate function by $\sqrt{n}$ so that the hazard rate function associated with customer abandonment times in the n^{th} system is

$$h^n(x) \equiv h(\sqrt{n}x). \tag{3.4}$$

To intuitively motivate the scaling in (3.4), first observe that in a conventional queueing system having $a_i = \infty$ for all $i \geq 0$, Kingman's approximation [15] shows the queue size is proportional to $(1-\rho^n)^{-1}$, which is of order $\sqrt{n}$ from assumption (3.3). Because the arrival rate in the n^{th}

system is of order n, a sample path version of Little's law known as the snapshot principle (see Reiman [26]) suggests that

$$V^n(t) \approx \frac{Q^n(t)}{n\rho^n} \propto \sqrt{n}/n = 1/\sqrt{n}, \tag{3.5}$$

meaning the offered waiting time in the n^{th} system shrinks at rate $n^{-1/2}$ as n grows large. Therefore, as in an observation made much earlier by Lehoczky [21] (and further developed in [7], [19], [20], and [18] for a $GI/GI/1{+}GI$ system operating under the earliest-deadline-first service discipline, and under the assumption that customers do not abandon the system when their deadline expires), in order that the limiting system capture the effects of customer abandonments, customer abandonment times must be shrinking (at rate $\sqrt{n}$) as n grows large. Furthermore, in order that more than only the behavior of the abandonment distribution close to the origin be used to determine whether or not a customer in the n^{th} system abandons when n is large, the hazard rate scaling must inflate its argument by $\sqrt{n}$.

Under Assumption 1, (3.4) implies the distribution of abandonment times in the n^{th} system is

$$F^n(x) = 1 - \exp\left(-\int_0^x h(\sqrt{n}u)du\right), \text{ for } x \geq 0. \tag{3.6}$$

Under Assumption 2, (3.4) implies the upper bound on abandonment times in the n^{th} system is

$$C^n \equiv \frac{C}{\sqrt{n}}, \tag{3.7}$$

and the distribution of abandonment times in the n^{th} system is

$$F^n(x) = \left(1 - \exp\left(-\frac{1}{\sqrt{n}}\int_0^{(\sqrt{n}x)\wedge C} h(w)dw\right)\right) + b^n \mathbf{1}\{\sqrt{n}x \geq C\}, \tag{3.8}$$

where

$$b^n \equiv \exp\left(-\frac{1}{\sqrt{n}}\int_0^C h(w)dw\right). \tag{3.9}$$

We assume customer abandonment times in the n^{th} system are an i.i.d. sequence of random variables $\{a_j^n, j \geq 1\}$ having distribution F^n defined in either (3.6) or (3.8). Note that the sum of centered service times of those customers that will eventually abandon becomes

$$S_a^n(i) = \frac{1}{n}\sum_{j=1}^{i} (v_j - E[v_1])\, \mathbf{1}\{V^n(t_i^{n,-}) \geq a_i^n\}. \tag{3.10}$$

It is useful for later analysis to define the fluid-scaled quantity

$$\overline{A}^n(t) \quad \equiv \quad \frac{A^n(t)}{n}, \tag{3.11}$$

and the diffusion-scaled quantities

$$\tilde{V}^n(t) \equiv \sqrt{n}V^n(t) \tag{3.12}$$

$$\tilde{A}^n(t) \equiv \sqrt{n}\left(\frac{1}{n}A^n(t) - \rho^n t\right) \tag{3.13}$$

$$\tilde{S}^n(t) \equiv \sqrt{n}S^n\left(\lfloor nt \rfloor\right) \tag{3.14}$$

$$\tilde{S}^n_a(t) \equiv \sqrt{n}S^n_a\left(\lfloor nt \rfloor\right) \tag{3.15}$$

$$\tilde{M}^n_a(t) \equiv \frac{1}{\sqrt{n}}M^n_a(\lfloor nt \rfloor) \tag{3.16}$$

$$\tilde{I}^n(t) \equiv \sqrt{n}I^n(t). \tag{3.17}$$

Recall from (3.5) that the scaling that leads to a non-degenerate limit process should inflate the offered waiting time process by $\sqrt{n}$.

We require the following technicalities. All random variables are defined on a common probability space $(\Omega, \mathcal{F}, P)$. For each positive integer d, let $D([0,\infty), \Re^d)$ be the space of right continuous functions with left limits (RCLL) in $\Re^d$ having time domain $[0,\infty)$. We endow $D([0,\infty), \Re^d)$ with the usual Skorokhod J_1 topology, and let M^d denote the Borel σ-algebra associated with the J_1 topology. All stochastic processes are measurable functions from $(\Omega, \mathcal{F}, P)$ into $(D([0,\infty), \Re^d), M^d)$ for some appropriate dimension d. We will often use the notation $\xi^n = \{\xi^n(t), t \geq 0\}$ to denote the stochastic process associated with a collection of random variables $\{\xi^n(t), t \geq 0\}$. Suppose $\{\xi^n\}_{n=1}^{\infty}$ is a sequence of stochastic processes. The notation $\xi^n \Rightarrow \xi$ means that the probability measures induced by the ξ^n's on $(D([0,\infty), \Re^d), M^d)$ converge weakly to the probability measure on $(D([0,\infty), \Re^d), M^d)$ induced by the stochastic process ξ. The notation $\stackrel{D}{=}$ means equal in distribution.

The functional strong law of large numbers (see, for example, Theorem 5.10 in Chen and Yao [5]) establishes

$$\overline{A}^n \to e, \tag{3.18}$$

P-almost surely, uniformly on compact sets, as $n \to \infty$, where $e(t) = t$ for all $t \geq 0$ is the identity function. Let $W_{S,1}$ and $W_{S,2}$ be independent, standard Brownian motions. The functional central limit theorem for renewal processes (see, for example Theorem 5.11 in Chen and Yao [5]) establishes

$$\tilde{A}^n \Rightarrow \text{var}(u_1)W_{S,1},$$

as $n \to \infty$, and Donsker's theorem (see, for example, Theorem 14.1 in Billingsley [3]) establishes

$$\tilde{S}^n \Rightarrow \text{var}(v_1)W_{S,2},$$

as $n \to \infty$. The assumed independence of the inter-arrival and service time sequences implies the joint convergence

$$\left(\tilde{A}^n, \tilde{S}^n\right) \Rightarrow \left(\text{var}(u_1)W_{S,1}, \text{var}(v_1)W_{S,2}\right), \tag{3.19}$$

as $n \to \infty$. We often use the random time change theorem in our proofs, and a convenient statement of this result can be found in Chapter 3, Section 14 of Billingsley [3]. In general, addition is not a continuous map from $D\left([0,\infty), \Re\right) \times D\left([0,\infty), \Re\right) \to D\left([0,\infty), \Re\right)$; however, addition is a continuous map on the subspace of continuous functions. All limit processes in this paper are continuous, and so we often use the continuous mapping theorem (see, for example, Theorem 3.4.1 of Whitt [36]) in association with the addition operator and obtained limit processes without further explanation. Finally, the space $D([0,\infty),\Re)$ is separable and complete by Theorem 16.3 in Billingsley [3], and so relative compactness and tightness in $D([0,\infty),\Re)$ are equivalent by Prohorov's theorem (see, for example, Theorem 5.1 in Billingsley [3] for the direct half and Theorem 5.2 in [3] for the converse half). We use the two words interchangeably.

3.2 Intuition for the Hazard Rate Scaling

In order to produce an interesting limiting diffusion approximation, we would like the state-dependent rate at which customers are abandoning the system, appropriately scaled, to converge to a non-degenerate limit. To calculate this state-dependent rate, we must first determine the probability that each customer in the queue will abandon the system in the next small amount of time. We can calculate this abandonment probability for each customer using the hazard rate function associated with the abandonment distribution as follows.

First, as in Whitt [38], we assume that few customers have abandoned, and so the amount of time that the ith customer from the back of the queue has been waiting is close to i/n, because i customers have arrived after this customer, the average inter-arrival time in the nth system is $1/(\rho^n n)$, and ρ^n is close to 1. This then implies that the probability that this customer will abandon the system in the next δ time units is close to $h^n(i/n)\delta$, and so, summing over all the customers in line at time t, the abandonment rate at time t is approximately

$$\sum_{i=1}^{Q^n(t)} h^n\left(\frac{i}{n}\right).$$

From (3.5), $Q^n \approx n\rho^n V^n$, and so since ρ^n is close to 1 for large n, the above sum is approximately

$$\sum_{i=1}^{nV^n(t)} h^n\left(\frac{i}{n}\right) = \sum_{i=1}^{\sqrt{n}\tilde{V}^n(t)} h\left(\frac{i}{\sqrt{n}}\right),$$

where the equality follows from the definition of h^n in (3.4) and $\tilde{V}^n$ in (3.12).

Since the relationship (3.5) also suggests that the queue-length is growing at rate $\sqrt{n}$, to have any hope of obtaining a finite, state-dependent total system abandonment rate, we must scale by $n^{-1/2}$. Then, assuming that $\tilde{V}^n \Rightarrow V$ as $n \to \infty$, we find that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\sqrt{n}\tilde{V}^n(t)} h\left(\frac{i}{\sqrt{n}}\right) \Rightarrow \int_0^{V(t)} h(u)du, \tag{3.20}$$

as $n \to \infty$, also using the definition of the Riemann-Stieltjes integral. Hence we have a limiting regime in which the total system abandonment rate converges to a non-degenerate limit. Moreover, the entire abandonment distribution influences this limiting system abandonment rate.

We further conjecture that for large n, the instantaneous rate of abandonment approximately equals the arrival rate times the probability that an arriving customer abandons. Consider a customer who arrives to the queue in the n^{th} system at time t and finds the offered waiting time to be $V^n(t)$. When abandonment times are unbounded, the probability that this customer abandons is

$$F^n(V^n(t)) \quad = \quad 1 - \exp\left(- \int_0^{V^n(t)} h(\sqrt{n}u)du \right),$$

and a change of variables shows that the right hand side of the above expression is equivalently rewritten as

$$1 - \exp\left(- \frac{1}{\sqrt{n}} \int_0^{\tilde{V}^n(t)} h(u)du \right).$$

Again assuming $\tilde{V}^n \Rightarrow V$ as $n \to \infty$, a Taylor expansion of e^x about zero shows that the probability an arriving customer abandons decreases at rate $\sqrt{n}$ as n grows large. In particular, applying L'Hopital's rule shows

$$\sqrt{n}F^n(V^n(t)) \Rightarrow \int_0^{V(t)} h(u)du \tag{3.21}$$

as $n \to \infty$. Comparing the limits (3.20) and (3.21) side-by-side shows

$$\sum_{i=1}^{\sqrt{n}\tilde{V}^n(t)} h\left(\frac{i}{\sqrt{n}}\right) \approx nF^n(V^n(t)),$$

suggesting that for large n the rate of abandonment approximately equals the arrival rate times the probability of abandonment.

3.3 Implications of the Scaling

In this section, we discuss the implications of our assumed hazard rate scaling (3.4) on customer impatience. We begin with the following definition. Recall that F^n represents the abandonment time distribution of customers in the n^{th} system.

Definition 1 *We say that customers are becoming more impatient on the diffusive time scale if*

$$F^{n+1}((n+1)^{-1/2}x) \geq F^n(n^{-1/2}x),$$

for all $x \in \mathbb{R}$ and $n \geq 1$.

The diffusive time scale is of interest because from (3.5), customer waiting times in the nth system are of order $n^{-1/2}$.

It is easy to see that customers are always becoming more impatient on the diffusive time scale and we record this as our first proposition of this section. The proof is immediate from the representation for F^n in (3.6) under Assumption 1 and (3.8) under Assumption 2, and so is omitted.

Proposition 1 *Customers are becoming more impatient on the diffusive time scale for any abandonment time distribution satisfying Assumption 1 or 2.*

We next study what is occurring to customer impatience on the original time scale. The following definition characterizes customer impatience in terms of whether the abandonment time distribution function is stochastically increasing or decreasing as n increases.

Definition 2 *We say that customers are becoming more impatient (patient) on the original time scale if $F^{n+1}(x) \geq F^n(x)$ ($F^{n+1}(x) \leq F^n(x)$) for all $x \in \Re$ and $n \geq 1$.*

When we do not change the timescale as n increases, the characterization of customer patience levels is more complicated, and requires closer examination of the hazard rate function h. Recall the following definition of a distribution function G with an increasing (decreasing) average hazard rate.

Definition 3 *A distribution function G, with hazard rate h, is said to possess an increasing average hazard rate if for $0 \leq a \leq b$,*

$$\frac{1}{a}\int_0^a h(u)du \leq \frac{1}{b}\int_0^b h(u)du.$$

We say that G has a decreasing average hazard rate if the above inequality holds with $b \leq a$.

The following result characterizes impatience on the original time scale.

Proposition 2 *Under either Assumption 1 or Assumption 2, customers become more impatient (patient) on the original time scale if and only if their abandonment time distribution possesses an increasing (decreasing) average hazard rate.*

Proof of Proposition 2: Suppose first that F possesses an increasing average hazard rate. Then, for each $x \geq 0$ under Assumption 1 or $0 \leq x < C^n$ under Assumption 2, from the expression for F^n in (3.6) or (3.8),

$$1 - F^n(x) = e^{-\frac{1}{\sqrt{n}}\int_0^{\sqrt{n}x} h(u)du} \geq e^{-\frac{1}{\sqrt{n+1}}\int_0^{\sqrt{n+1}x} h(u)du} = 1 - F^{n+1}(x),$$

where the inequality follows since F possesses an increasing average hazard rate.

Suppose, on the other hand, that customers are becoming more impatient on the original time scale. Then, for each $x \geq 0$ under Assumption 1 or $0 \leq x \leq C^n$ under Assumption 2,

$$e^{-\frac{1}{\sqrt{n}}\int_0^{x\sqrt{n}} h(u)du} = 1 - F^n(x) \geq 1 - F^{n+1}(x) = e^{-\frac{1}{\sqrt{n+1}}\int_0^{x\sqrt{n+1}} h(u)du},$$

so that F possess an increasing average hazard rate. □

Because the exponential distribution is the only continuous distribution with a constant hazard rate, we also have the following immediate corollary of Proposition 2.

Corollary 1 *The level of customer patience remains constant if and only if the abandonment distribution is exponential.*

4 Non-linear Generalized Regulator Mappings

The key to our asymptotic analysis in Section 5 (that establishes the weak convergence of the offered waiting time process) is to represent the offered waiting time process in terms of a one- or two-sided non-linear generalized regulator mapping. To see what the appropriate mappings are, first observe that under assumption 1, from (2.3), (2.4), (2.5), and (3.1), the evolution equation for the offered waiting time process in the n^{th} system is

$$V^n(t) = X^n(t) + \epsilon^n(t) - \int_0^t \left(\int_0^{V^n(s^-)} h^n(u)du \right) ds + I^n(t), \tag{4.1}$$

where

$$X^n(t) \equiv \frac{1}{n}A^n(t) - \rho^n t + S^n(A^n(t)) + t(\rho^n - 1) - S_a^n(A^n(t)) - \frac{1}{n}M_a\left(A^n(t)\right) \tag{4.2}$$

and, also using the definition of F^n in (3.6)

$$\begin{aligned} \epsilon^n(t) &\equiv \int_0^t \left(\int_0^{V^n(s^-)} h^n(u)du \right) ds - \frac{1}{n}\int_0^t F^n\left(V^n(s^-)\right) dA^n(s). \qquad (4.3)\\ &= \frac{1}{\sqrt{n}}\int_0^t \left(\int_0^{\tilde{V}^n(s^-)} h(w)dw \right) ds - \int_0^t \left(1 - \exp\left(-\frac{1}{\sqrt{n}}\int_0^{\tilde{V}^n(s^-)} h(w)dw \right) \right) d\overline{A}^n(s) \end{aligned}$$

Under Assumption 2, we add and subtract the number of arrivals that find the offered waiting time process exceeding the upper bound on abandonment times, appropriately scaled,

$$U^n(t) \equiv \frac{b^n}{n}\int_0^t \mathbf{1}\{V^n(s^-) \geq C^n\} dA^n(s), \tag{4.4}$$

to the right-hand side of (4.1) to find

$$V^n(t) = X^n(t) + \epsilon_B^n(t) - \int_0^t \left(\int_0^{V^n(s^-)\wedge C^n} h^n(u)du \right) ds + I^n(t) - U^n(t), \tag{4.5}$$

where X^n is as defined in (4.2) and, also using the definition of F^n in (3.8)

$$\begin{aligned} \epsilon_B^n(t) &\equiv \int_0^t \left(\int_0^{V^n(s^-)\wedge C^n} h^n(u)du \right) ds + \frac{b^n}{n}\int_0^t \mathbf{1}\{V^n(s^-) \geq C^n\} dA^n(s) \qquad (4.6)\\ &\quad -\frac{1}{n}\int_0^t F^n\left(V^n(s^-)\right) dA^n(s)\\ &= \frac{1}{\sqrt{n}}\int_0^t \left(\int_0^{\tilde{V}^n(s^-)\wedge C} h(w)dw \right) ds - \int_0^t \left(1 - \exp\left(-\frac{1}{\sqrt{n}}\int_0^{\tilde{V}^n(s^-)\wedge C} h(w)dw \right) \right) d\overline{A}^n(s). \end{aligned}$$

Observe that the process I^n in (4.1) and (4.5) only increases when V^n is 0 and the process U^n in (4.4) only increases when V^n is equal to or exceeds C^n. Regarding $X^n + \epsilon^n$ and $X^n + \epsilon_B^n$ as the "free" processes, equations (4.1) and (4.5) immediately suggest the non-linear generalizations of the conventional one- and two-sided regulator mappings required to obtain weak convergence results on the offered waiting time process under Assumptions 1 and 2.

The remainder of this section is organized as follows. We define a one-sided non-linear generalized regulator mapping in Subsection 4.1, and prove its existence, uniqueness, and continuity. Next, in Subsection 4.2, we do the same for a two-sided non-linear generalized regulator mapping.

4.1 The One-Sided Non-Linear Generalized Regulator Mapping

The one-sided non-linear generalized regulator mapping generalizes the conventional one-sided regulator mapping first introduced in Skorokhod [29] having the explicit form

$$\phi(x)(t) \equiv x(t) + \psi(x)(t) \in [0, \infty) \text{ for all } t \geq 0, \tag{4.7}$$

for $x \in D([0, \infty), \Re)$ and

$$\psi(x)(t) \equiv \sup_{0 \leq s \leq t} \left[-x(s)\right]^+, \tag{4.8}$$

to a mapping that allows for non-linear state-space dependence.

Definition 4 *(The one-sided nonlinear generalized regulator mapping)*
Given h a non-negative, continuous function on $[0, \infty)$ and $x \in D([0, \infty), \Re)$ having $x(0) \geq 0$, the one-sided nonlinear generalized regulator mapping

$$(\phi^h, \psi^h) : D([0, \infty), \Re) \mapsto D([0, \infty), [0, \infty) \times [0, \infty))$$

is defined by

$$(\phi^h, \psi^h)(x) \equiv (z, l)$$

where

(C1) $z(t) = x(t) - \int_0^t \left(\int_0^{z(s)} h(u)du \right) ds + l(t) \in [0, \infty)$ *for all* $t \geq 0$*;*

(C2) l *is nondecreasing,* $l(0) = 0$*, and* $\int_0^\infty z(t)dl(t) = 0$*.*

When h is the zero function, ϕ and ψ in (4.7) and (4.8) uniquely satisfy Definition 4. When the function h is constant, the non-linear generalized one-sided regulator mapping becomes the linearly generalized one-sided mapping given in Section 5 of Reed and Ward [25].

For $x \in D([0, \infty), \Re)$ having $x(0) \geq 0$ and (ϕ, ψ) defined in (4.7) and (4.8), set

$$z \equiv \phi^h(x) = \phi\left(\mathcal{M}^h(x)\right) \tag{4.9}$$

$$l \equiv \psi^h(x) = \psi\left(\mathcal{M}^h(x)\right), \tag{4.10}$$

where the mapping $\mathcal{M}^h : D\left([0, \infty), \Re\right) \to D\left([0, \infty), \Re\right)$ has $\mathcal{M}^h(x) \equiv w$ for w that solves the integral equation

$$w(t) = x(t) - \int_0^t \left(\int_0^{\phi(w)(s)} h(u)du \right) ds \tag{4.11}$$

having initial condition $w(0) = x(0)$. Observe that (z, l) defined in (4.9) and (4.10) satisfy conditions (C1) and (C2) of Definition 4 because

1. From (4.7),(4.9), (4.10), and (4.11),

$$\begin{aligned} 0 &\leq \phi\left(\mathcal{M}^h(x)\right)(t) \\ &= \mathcal{M}^h(x)(t) + \psi\left(\mathcal{M}^h(x)\right)(t) \\ &= x(t) - \int_0^t \left(\int_0^{z(s)} h(u)du\right) ds + l(t); \end{aligned}$$

2. The function $\psi(\mathcal{M}^h(x))$ is non-decreasing from its definition in (4.8), $\psi\left(\mathcal{M}^h(x)\right)(0) = 0$ since $\mathcal{M}^h(x)(0) = x(0) \geq 0$ by assumption on x, and $\int_0^\infty \phi\left(\mathcal{M}^h(x)\right)(t) d\psi\left(\mathcal{M}^h(x)\right)(t) = 0$ from the definitions of ϕ and ψ in (4.7) and (4.8).

Therefore, the key to proving existence, uniqueness, and continuity of the non-linear generalized one-sided regulator mapping in Definition 4 is the following lemma that establishes the existence, uniqueness, and local Lipschitz continuity of the integral equation in (4.11). We note that if h is bounded on $[0, \infty)$ then the mapping $\mathcal{M}^h$ is globally Lipschitz; in particular, the constant κ in Lemma 1, part (ii) below depends only on T and the conclusion in (ii) holds for all $x_1, x_2 \in D([0, \infty), \Re)$.

Lemma 1 *(Properties of the Integral Equation (4.11))*
Let h be a non-negative, continuous function on $[0, \infty)$.

(i) For each $x \in D([0, \infty), \Re)$ there exists a unique w satisfying (4.11).

(ii) Let $T > 0$ and $x \in D([0, \infty), \Re)$. There exists κ, dependent on x, such that if $x_1, x_2 \in D([0, \infty), \Re)$ satisfy

$$\|x_1 - x\|_T < 1 \text{ and } \|x_2 - x\|_T < 1,$$

then

$$\|\mathcal{M}^h(x_1) - \mathcal{M}^h(x_2)\|_T < \kappa \|x_1 - x_2\|_T.$$

(iii) The function $\mathcal{M}^h$ is continuous when the space $D([0, \infty), \Re)$ is endowed with Skorohod J_1 topology.

Our next proposition establishes the existence, uniqueness, and continuity of the non-linear generalized one-sided regulator mapping. It is useful for its proof and also for later analysis to observe that Lemma 13.4.1 of Whitt [36] establishes that for any $x_1, x_2 \in D([0, \infty), \Re)$,

$$\|\psi(x_1) - \psi(x_2)\|_T \leq \|x_1 - x_2\|_T, \tag{4.12}$$

and, therefore from (4.7), as in Lemma 13.5.1 of Whitt [36],

$$\|\phi(x_1) - \phi(x_2)\|_T \leq 2\|x_1 - x_2\|_T. \tag{4.13}$$

Proposition 3 *(Properties of the Non-linear Generalized One-sided Regulator Mapping)*
Let h be a non-negative, continuous function on $[0, \infty)$.

(i) *For each $x \in D([0,\infty),\Re)$ having $x(0) \geq 0$, there exists a unique pair of functions*

$$(\phi^h, \psi^h)(x) = (z, l)$$

that satisfies (C1)-(C2) of Definition 4.

(ii) *Suppose $x \in D([0,\infty),\Re)$ and $x(0) \geq 0$. Let $h^n(x) = h(\sqrt{n}x)$ for all $x \geq 0$ be as defined in (3.4). Then,*

$$\sqrt{n}\left(\phi^{h^n}, \psi^{h^n}\right)(x) = \left(\phi^h, \psi^h\right)(\sqrt{n}x).$$

(iii) *Let $T > 0$ and $x \in D([0,\infty),\Re)$. There exists κ, dependent on x, such that if $x_1, x_2 \in D([0,\infty),\Re)$ satisfy*

$$\|x_1 - x\|_T < 1 \text{ and } \|x_2 - x\|_T < 1,$$

then

$$\|\phi^h(x_1) - \phi^h(x_2)\|_T \vee \|\psi^h(x_1) - \psi^h(x_2)\|_T \leq \kappa \|x_1 - x_2\|_T.$$

(iv) *Both the functions ϕ^h and ψ^h are continuous when the space $D([0,\infty),\Re)$ is endowed with the Skorohod J_1 topology.*

Proof of (i): Existence follows from the representations (4.9) and (4.10) and part (i) of Lemma 1. To see the representations (4.9) and (4.10) are unique, let (z, l) be a solution satisfying (C1)-(C2) of Definition 4. Because for

$$g(t) \equiv x(t) - \int_0^t \left(\int_0^{z(s)} h(u)du \right) ds, \; t \geq 0, \tag{4.14}$$

from (C1)

$$z(t) = g(t) + l(t) \geq 0,$$

and l satisfies (C2), we conclude

$$(z, l) \quad = \quad (\phi, \psi)(g). \tag{4.15}$$

If we now show that $g = \mathcal{M}^h(x)$, we will then have that

$$(z, l) = \left(\phi\left(\mathcal{M}^h(x)\right), \psi\left(\mathcal{M}^h(x)\right)\right)$$

which, by part (i) of Lemma 1 and the uniqueness of (ϕ, ψ), will uniquely define (z, l). However, by (4.15) we have $z = \phi(g)$ and so it follows upon substitution into (4.14) that

$$g(t) = x(t) - \int_0^t \left(\int_0^{\phi(g)(s)} h(u)du \right) ds$$

as desired.

Proof of (ii): Since

$$\left(\phi^{h^n}, \psi^{h^n}\right)(x) = (z, l) \tag{4.16}$$

satisfies (C1) of Definition 4,

$$z(t) = x(t) - \int_0^t \left(\int_0^{z(s)} h^n(u)du \right) ds + l(t).$$

Let $z^n \equiv \sqrt{n}z$, $x^n \equiv \sqrt{n}x$, and $l^n \equiv \sqrt{n}l$. Multiply both sides of the above equation by $\sqrt{n}$ to find

$$z^n(t) = x^n(t) - \int_0^t \left(\int_0^{z^n(s)} h(u)du \right) ds + l^n(t).$$

Since also (C2) of Definition 4 holds for (z, l), $l^n = \sqrt{n}l$ is non-decreasing, $l^n(0) = \sqrt{n}l(0) = 0$, and $\int_0^\infty z^n(t)dl^n(t) = \int_0^\infty nz(t)dl(t) = 0$, we conclude

$$\left(\phi^h, \psi^h\right)(x^n) = (z^n, l^n).$$

Therefore, from the definitions of z^n, l^n, and x^n, and the equality (4.16),

$$\sqrt{n}\left(\phi^{h^n}, \psi^{h^n}\right)(x) = \sqrt{n}(z, l) = (z^n, l^n) = \left(\phi^h, \psi^h\right)(x^n) = \left(\phi^h, \psi^h\right)(\sqrt{n}x).$$

Proof of (iii): From the representations (4.9) and (4.10), the Lipschitz property of ψ and ϕ in (4.12) and (4.13), and part (ii) of Lemma 1,

$$\begin{aligned}
&\|\phi^h(x_1) - \phi^h(x_2)\|_T \vee \|\psi^h(x_1) - \psi^h(x_2)\|_T \\
&= \|\phi\left(\mathcal{M}^h(x_1)\right) - \phi\left(\mathcal{M}^h(x_2)\right)\|_T \vee \|\psi\left(\mathcal{M}^h(x_1)\right) - \psi\left(\mathcal{M}^h(x_2)\right)\|_T \\
&\leq 2\|\mathcal{M}^h(x_1) - \mathcal{M}^h(x_2)\|_T \vee \|\mathcal{M}^h(x_1) - \mathcal{M}^h(x_2)\|_T \\
&\leq 2\kappa\|x_1 - x_2\|_T,
\end{aligned}$$

where κ is as in part (ii) of Lemma 1.

Proof of (iv): Suppose that $x^n \to x$ as $n \to \infty$ in the Skorohod J_1 topology. Then, from part (iii) of Lemma 1, we have that

$$\mathcal{M}^h(x^n) \quad \to \quad \mathcal{M}^h(x) \;\text{ as }\; n \to \infty,$$

in the Skorohod J_1 topology. Thus, since by Theorems 13.4.1 and 13.5.1 in Whitt [36], ϕ and ψ are both continuous in the Skorohod J_1 topology, and compositions of continuous functions are continuous, this then implies that as $n \to \infty$, in the Skorohod J_1 topology,

$$\phi^h(x^n) \;=\; \phi\left(\mathcal{M}^h(x^n)\right) \;\to\; \phi\left(\mathcal{M}^h(x)\right) \;=\; \phi^h(x)$$

and

$$\psi^h(x^n) \;=\; \psi\left(\mathcal{M}^h(x^n)\right) \;\to\; \psi\left(\mathcal{M}^h(x)\right) \;=\; \psi^h(x).$$

□

4.2 The Two-Sided Non-Linear Generalized Regulator Mapping

The two-sided non-linear generalized regulator mapping generalizes the conventional two-sided regulator mapping defined in Section 14.8.1 of Whitt [36], and having the explicit form given in Theorem 1.4 in Kruk et al [17].

Definition 5 *Let $C > 0$. Given h a non-negative continuous function on $[0, C]$ and $x \in D([0,\infty),\Re)$ having $0 \le x(0) \le C$, the two-sided non-linear generalized regulator mapping*

$$\left(\phi^h_C, \psi^h_{1,C}, \psi^h_{2,C}\right) : D\left([0,\infty),\Re\right) \to D\left([0,\infty),[0,C]\times[0,\infty)\times[0,\infty)\right)$$

is defined by

$$\left(\phi^h_C, \psi^h_{1,C}, \psi^h_{2,C}\right) \equiv (z, l, u)$$

where

(C1) $z(t) = x(t) - \int_0^t \left(\int_0^{z(s)} h(u)du\right) ds + l(t) - u(t) \in [0, C]$ *for all* $t \ge 0$*;*

(C2) l *and* u *are non-decreasing,* $l(0) = u(0) = 0$*, and* $\int_0^\infty z(t)dl(t) = \int_0^\infty [C - z(t)]^+ du(t) = 0$.

Similar to Section 4.1, when h is the zero function, Definition 5 defines the conventional two-sided regulator mapping, and we denote the unique mapping by $(\phi_C, \psi_{1,C}, \psi_{2,C})$. When the function h is constant, the non-linear generalized two-sided regulator mapping becomes the linearly generalized two-sided mapping given in Definition 2 of Ward and Kumar [31].

For $x \in D([0,\infty),\Re)$ having $0 \le x(0) \le C$, set

$$z \equiv \phi^h_C(x) = \phi_C\left(\mathcal{M}^h_C(x)\right) \tag{4.17}$$

$$l \equiv \psi^h_{1,C}(x) = \psi_{1,C}\left(\mathcal{M}^h_C(x)\right) \tag{4.18}$$

$$u \equiv \psi^h_{2,C}(x) = \psi_{2,C}\left(\mathcal{M}^h_C(x)\right), \tag{4.19}$$

where the mapping $\mathcal{M}^h_C : D([0,\infty),\Re) \to D([0,\infty),\Re)$ has $\mathcal{M}^h_C(x) \equiv w$ for w that solves the integral equation

$$w(t) = x(t) - \int_0^t \left(\int_0^{\phi_C(w)(s)} h(u)du\right) ds \tag{4.20}$$

having initial condition $w(0) = x(0)$. By paralleling the arguments in the beginning of Section 4.1, it is straightforward to show that (z, l, u) defined in (4.17)-(4.19) satisfy conditions (C1) and (C2) of Definition 5. Therefore, the key to understanding the properties of the non-linear generalized two-sided regulator mapping in Definition 5 is to understand the properties of the integral equation (4.20).

Lemma 2 *(Properties of the Integral Equation (4.20))*
Let h be a non-negative, continuous function on $[0, C]$.

(i) For each $x \in D([0,\infty),\Re)$, there exists a unique w satisfying (4.20).

(ii) Let $T > 0$. There exists a finite constant κ that depends only on T such that for any $x_1, x_2 \in D([0,\infty),\Re)$ having $0 \le x_1(0), x_2(0) \le C$,

$$\|\mathcal{M}^h_C(x_1) - \mathcal{M}^h_C(x_2)\|_T \le \kappa \|x_1 - x_2\|_T.$$

(iii) *The function* $\mathcal{M}_C^h$ *is continuous when the space* $D([0,\infty),\Re)$ *is endowed with the Skorohod* J_1 *topology.*

The main proposition of this subsection establishes several useful properties of the non-linear generalized two-sided regulator mapping.

Proposition 4 *(Properties of the Non-linear Generalized Two-Sided Regulator Mapping)* *Let* h *be a non-negative, continuous function on* $[0,C]$.

(i) *For each* $x \in D([0,\infty),\Re)$ *having* $0 \le x(0) \le C$, *there exists a unique pair of functions*

$$\left(\phi_C^h, \psi_{1,C}^h, \psi_{2,C}^h\right)(x) = (z, l, u)$$

that satisfies (C1)-(C2) of Definition 5.

(ii) *Suppose* $x \in D([0,\infty),\Re)$ *and* $0 \le x(0) \le C$. *Let* $h^n(x) = h(\sqrt{n}x)$ *for all* $x \ge 0$ *be as defined in (3.4). Then,*

$$\sqrt{n}\left(\phi_C^{h^n}, \psi_{1,C}^{h^n}, \psi_{2,C}^{h^n}\right)(x) = \left(\phi_{\sqrt{n}C}^h, \psi_{1,\sqrt{n}C}^h, \psi_{2,\sqrt{n}C}^h\right)(\sqrt{n}x).$$

(iii) *Let* $T > 0$. *There exists a finite constant* κ *that depends only on* T *such that for any* $x_1, x_2 \in D([0,\infty),\Re)$ *having* $0 \le x_1(0), x_2(0) \le C$

$$\|\phi_C^h(x_1) - \phi_C^h(x_2)\|_T \le \kappa \|x_1 - x_2\|_T.$$

Furthermore[2], *if* $\|x^n - x\|_T \to 0$ *as* $n \to \infty$, *then*

$$\|\psi_{j,C}^h(x^n) - \psi_{j,C}^h(x)\|_T \to 0, \ j \in \{1,2\}$$

as $n \to \infty$.

(iv) *The functions* $\phi_C^h, \psi_{1,C}^h$, *and* $\psi_{2,C}^h$ *are continuous when the space* $D([0,\infty),\Re)$ *is endowed with the Skorohod* J_1 *topology.*

Proof of (i): Existence follows from the representations (4.17)-(4.19) and part (i) of Lemma 2. The proof of uniqueness is very similar to part (i) of Proposition 3, and so is omitted.

Proof of (ii): Since

$$\left(\phi_C^{h^n}, \psi_{1,C}^{h^n}, \psi_{2,C}^{h^n}\right)(x) = (z, l, u) \tag{4.21}$$

satisfies (C1) of Definition 5,

$$z(t) = x(t) - \int_0^t \left(\int_0^{z(s)} h^n(u)du\right) ds + l(t) - u(t) \in [0, C].$$

[2] We remark that the Lipschitz property does not hold for $\psi_{1,C}^h$ and $\psi_{2,C}^h$. See Example 14.8.1 of Whitt [36] for a counterexample to the Lipschitz property for the conventional two-sided regulator mapping.

Let $z^n \equiv \sqrt{n}z$, $x^n \equiv \sqrt{n}x$, $l^n \equiv \sqrt{n}l$, $u^n \equiv \sqrt{n}u$, and $C^n \equiv \sqrt{n}C$. Multiply both sides of the above equation by $\sqrt{n}$ to find

$$z^n(t) = x^n(t) - \int_0^t \left(\int_0^{z^n(s)} h(w)dw \right) ds + l^n(t) - u^n(t) \in [0, C^n].$$

Since also (C2) of Definition 5 holds for (z, l, u), l^n and u^n are non-decreasing, $l^n(0) = u^n(0) = 0$, and

$$\int_0^\infty z^n(t)dl^n(t) = \int_0^\infty nz(t)dl(t) = 0$$
$$\int_0^\infty [C^n - z^n(t)]^+ du^n(t) = \int_0^\infty n [C - z(t)]^+ du(t) = 0,$$

we conclude

$$\left(\phi^h_{C^n}, \psi^h_{1,C^n}, \psi^h_{2,C^n}\right)(x^n) = (z^n, l^n, u^n).$$

Therefore, from the definitions of z^n, l^n, and x^n, and the equality (4.21),

$$\sqrt{n}\left(\phi^{h^n}_C, \psi^{h^n}_{1,C}, \psi^{h^n}_{2,C}\right)(x) = \sqrt{n}(z, l, u) = (z^n, l^n, u^n) = \left(\phi^h_{\sqrt{n}C}, \psi^h_{1,\sqrt{n}C}, \psi^h_{2,\sqrt{n}C}\right)(\sqrt{n}x).$$

Proof of (iii): From the representations (4.17), the Lipschitz property of $\phi_{[0,C]}$ established in Theorem 14.8.1 of [36] with Lipschitz constant 2, and part (i) of Lemma 2,

$$\begin{aligned}
\|\phi^h_C(x_1) - \phi^h_C(x_2)\|_T &= \|\phi_{[0,C]}\left(\mathcal{M}^h_C(x_1)\right) - \phi_{[0,C]}\left(\mathcal{M}^h_C(x_2)\right)\|_T \\
&\le 2\|\mathcal{M}^h_C(x_1) - \mathcal{M}^h_C(x_2)\|_T \\
&\le 2\kappa\|x_1 - x_2\|_T.
\end{aligned}$$

Next, assume $\|x^n - x\|_T \to 0$ as $n \to \infty$. Then, part (ii) of Lemma 2 guarantees

$$\|\mathcal{M}^h_C(x^n) - \mathcal{M}^h_C(x)\|_T \le \kappa\|x^n - x\|_T \to 0,$$

as $n \to \infty$. The representations of $\psi^h_{1,C}$ and $\psi^h_{2,C}$ in (4.18) and (4.19) and the continuity of the mappings $\psi_{1,C}$ and $\psi_{2,C}$ established in Theorem 14.8.1 in Whitt [36] then show that since a composition of continuous functions is continuous

$$\|\psi^h_{j,C}(x^n) - \psi^h_{j,C}(x)\|_T = \|\psi_{j,C}\left(\mathcal{M}^h_C(x^n)\right) - \psi_{j,C}\left(\mathcal{M}^h_C(x)\right)\|_T \to 0,$$

as $n \to \infty$.

Proof of (iv): Since by Theorem 14.8.2 of Whitt [36], $\phi_C, \psi_{1,C}$, and $\psi_{2,C}$ are all continuous in the Skorohod J_1 topology, by part (iv) of Lemma 2, the proof now proceeds in the same manner as the proof of part (iii) of Proposition 3. □

5 Weak Convergence of the Offered Waiting Time Process

We establish the weak convergence of the scaled offered waiting time process $\tilde{V}^n$ in (3.12) when the abandonment distribution has unbounded support (Assumption 1) in Section 5.1, and when the abandonment distribution has bounded support (Assumption 2) in Section 5.2. Specifically, we prove the following theorem.

Theorem 1 *Let W be a Brownian motion with drift θ given in (3.3), variance $\sigma^2 = var(u_1) + var(v_1)$, and initial position $W(0) = 0$.*

(i) Under assumption 1, $\left(\tilde{V}^n, \tilde{I}^n\right) \Rightarrow \left(\phi^h, \psi^h\right)(W)$, as $n \to \infty$.

(ii) Under assumption 2, $\left(\tilde{V}^n, \tilde{I}^n, \tilde{U}^n\right) \Rightarrow \left(\phi^h_C, \psi^h_{1,C}, \psi^h_{2,C}\right)(W)$, as $n \to \infty$.

The limiting virtual waiting time process of Theorem 1 may loosely be described as a diffusion process with infinitesimal drift given by

$$m(x) = \theta x - \int_0^x h(u)du, \quad \text{for } x \geq 0,$$

and infinitesimal variance σ^2. There is a lower reflecting barrier at the origin for the case of part (i). Part (ii) also requires an upper reflecting barrier at the point C which represents an upper limit on the abandonment times.

Define

$$\tilde{X}^n(t) = \sqrt{n} X^n(t) \tag{5.1}$$

and

$$\tilde{\epsilon}^n(t) = \sqrt{n} \epsilon^n(t). \tag{5.2}$$

5.1 Proof of Theorem 1 part (i):

The key to our weak convergence proof is to represent the offered waiting time process in terms of the one-sided non-linear generalized regulator mapping

$$(V^n, I^n) = \left(\phi^{h^n}, \psi^{h^n}\right)(X^n + \epsilon^n), \tag{5.3}$$

from which the representation of the scaled offered waiting time process in terms of (ϕ^h, ψ^h) follows. To see that (5.3) is valid, first observe that the evolution equation (4.1) combined with the original definition of V^n in (2.1) that guarantees $V^n(t) \geq 0$ for all $t \geq 0$ implies condition (C1) of Definition 4 holds. Next, from (2.6), for every n, I^n is non-decreasing, has $I^n(0) = 0$, and

$$\int_0^\infty V^n(t)dI^n(t) = \int_0^\infty V^n(t)\mathbf{1}\{V^n(t) = 0\}dt = 0,$$

and so (C2) is also satisfied.

The definitions of $\tilde{V}^n$ and $\tilde{I}^n$ in (3.12) and (3.17), the representation (5.3), the scaling property of the non-linear generalized one-sided regulator mapping in part (ii) of Proposition 3, and the definitions of $\tilde{X}^n$ in (5.1) and $\tilde{\epsilon}^n$ in (5.2) imply

$$\begin{aligned}\left(\tilde{V}^n, \tilde{I}^n\right) &= \sqrt{n}\left(V^n, I^n\right) \\ &= \sqrt{n}\left(\phi^{h^n}, \psi^{h^n}\right)\left(X^n + \epsilon^n\right) \\ &= \left(\phi^h, \psi^h\right)\left(\sqrt{n}\left(X^n + \epsilon^n\right)\right) \\ &= \left(\phi^h, \psi^h\right)\left(\tilde{X}^n + \tilde{\epsilon}^n\right).\end{aligned} \tag{5.4}$$

Suppose we can show

$$\tilde{X}^n \Rightarrow W \text{ and } \tilde{\epsilon}^n \Rightarrow 0, \tag{5.5}$$

as $n \to \infty$. Then, the continuous mapping theorem establishes

$$\tilde{X}^n + \tilde{\epsilon}^n \Rightarrow W,$$

as $n \to \infty$. The result in part (i) then follows from the representation of $\left(\tilde{V}^n, \tilde{I}^n\right)$ in (5.4), the continuous mapping theorem, and part (iv) of Proposition 3.

To show (5.5), we require the following three lemmas. The first establishes that the process

$$R^n(i) \equiv \sum_{j=1}^{i} \mathbf{1}\{V^n\left(t_j^{n,-}\right) \geq a_j^n\}, \tag{5.6}$$

defined so that $R^n\left(A^n(t)\right)$ is the cumulative number of customers in $[0, t]$ who have arrived by time t and will either already abandoned or will abandon after time t, is small on fluid-scale. Define

$$\overline{R}^n(t) \equiv \frac{1}{n} R^n\left(\lfloor nt \rfloor\right). \tag{5.7}$$

Lemma 3 *Under assumption 1, as* $n \to \infty$, $\overline{R}^n \Rightarrow 0$.

The second establishes the weak convergence of the diffusion-scaled martingale $\tilde{M}_a^n$ in (3.16) to the zero process.

Lemma 4 *Under assumption 1, as* $n \to \infty$, $\tilde{M}_a^n \Rightarrow 0$.

The third establishes tightness of the offered waiting time process.

Lemma 5 *The sequence* $\{\tilde{V}^n\}$ *is tight in* $D([0, \infty), \Re)$.

5.1.1 Weak convergence of $\tilde{X}^n$ in (5.5):

From the definition of $\tilde{X}^n$ in (5.1), the evolution equation for X^n in (4.2), and the diffusion-scaled processes definitions in (3.13), (3.14), (3.15), and (3.16)

$$\tilde{X}^n(t) = \tilde{A}^n(t) + \tilde{S}^n\left(\overline{A}^n(t)\right) + \sqrt{n}t\left(\rho^n - 1\right) - \tilde{S}_a^n\left(\overline{A}^n(t)\right) - \tilde{M}_a^n\left(\overline{A}^n(t)\right). \tag{5.8}$$

Because the service time sequence is i.i.d., recalling the definitions of S_a^n and $\tilde{S}_a^n$ in (3.10) and (3.15) and R^n in (5.6), for any $t \geq 0$,

$$\tilde{S}_a^n\left(\overline{A}^n(t)\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{A^n(t)}\left(v_j - E\left[v_1\right]\right)\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n\right\} \stackrel{D}{=} \frac{1}{\sqrt{n}}\sum_{j=1}^{R^n(A^n(t))}\left(w_j - E\left[w_1\right]\right),$$

where $\{w_i, i \geq 1\}$ is an i.i.d. sequence of random variables with distribution equal to that of the service time distribution and which is also independent of the model primitives introduced in Section 2. It is straightforward to show that the finite-dimensional distributions are also equivalent, and so the definitions of S^n, $\tilde{S}^n$, and $\overline{R}^n$ in (3.2), (3.14), and (5.7) imply

$$\tilde{S}^n \circ \overline{R}^n \circ \overline{A}^n \stackrel{D}{=} \tilde{S}_a^n \circ \overline{A}^n. \tag{5.9}$$

The almost sure convergence of $\overline{A}^n$ in (3.18), the weak convergence of $\overline{R}^n$ in Lemma 3, the weak convergence of $\tilde{S}^n$ in (3.19), and the random time change theorem show $\tilde{S}^n \circ \overline{R}^n \circ \overline{A}^n \Rightarrow 0$, as $n \to \infty$, and so the distributional equality in (5.9) implies

$$\tilde{S}_a^n \Rightarrow 0, \tag{5.10}$$

as $n \to \infty$. Finally, the representation of $\tilde{X}^n$ in (5.8), the almost sure convergence of $\overline{A}^n$ in (3.18), the weak convergences in (3.19) and (5.10), the heavy traffic assumption (3.3), Lemma 4, and the random time change theorem imply

$$\tilde{X}^n \Rightarrow W,$$

as $n \to \infty$.

5.1.2 Weak convergence of $\tilde{\epsilon}^n$ in (5.5):

Let $\{n_k\}$ be a subsequence along which

$$\tilde{V}^{n_k} \Rightarrow V,$$

as $n_k \to \infty$. Such a subsequence exists because $\{\tilde{V}^n\}$ is tight in $D([0,\infty),\Re)$ by Lemma 5. From (3.18), because $\overline{A}^n$ converges to a deterministic limit process, the joint convergence

$$\left(\tilde{V}^{n_k}, \overline{A}^{n_k}\right) \Rightarrow (V, e),$$

as $n_k \to \infty$, is valid. The Skorokhod representation theorem (see, for example, Theorem 3.2.2 in Whitt [36]) guarantees there exists additional random elements on $(D([0,\infty),\Re),J_1) \times D([0,\infty),\Re),J_1)$, $\left\{\left(\breve{V}^{n_k},\breve{A}^{n_k}\right)\right\}$ and $\breve{V}$, defined on a possibly additional probability space $(\breve{\Omega},\breve{\mathcal{F}},\breve{P})$ such that

$$\left(\breve{V}^{n_k},\breve{A}^{n_k}\right) \stackrel{D}{=} \left(\tilde{V}^{n_k},\overline{A}^{n_k}\right), \quad \breve{V} \stackrel{D}{=} V, \tag{5.11}$$

and

$$\breve{P}\left(\lim_{n_k\to\infty}\left(\breve{V}^{n_k},\breve{A}^{n_k}\right) = \left(\breve{V},e\right)\right) = 1. \tag{5.12}$$

Define

$$\begin{aligned}\breve{\epsilon}^n(t) \quad &\equiv \quad \int_0^t \left(\int_0^{\breve{V}^n(s^-)} h(w)dw - \int_0^{\breve{V}(s^-)} h(w)dw\right) ds \\ &\quad + \int_0^t \left(\int_0^{\breve{V}(s^-)} h(w)dw\right) ds - \int_0^t \sqrt{n}\left(1-\exp\left(-\frac{1}{\sqrt{n}}\int_0^{\breve{V}^n(s^-)} h(w)dw\right)\right) d\breve{A}^n(s).\end{aligned} \tag{5.13}$$

Observe from the definition of $\tilde{\epsilon}^n$ in (4.3), ϵ^n in (2.4), and h^n in (3.4) that

$$\begin{aligned}\tilde{\epsilon}^n(t) \quad &= \quad \sqrt{n}\epsilon^n(t) \\ &= \quad \int_0^t \left(\int_0^{\tilde{V}^n(s^-)} h(w)dw\right) ds - \int_0^t \sqrt{n}\left(1-\exp\left(-\frac{1}{\sqrt{n}}\int_0^{\tilde{V}^n(s^-)} h(w)dw\right)\right) d\overline{A}^n(s).\end{aligned} \tag{5.14}$$

From the distributional equivalence (5.11), the definition of $\breve{\epsilon}^n$ in (5.13), and the representation of $\tilde{\epsilon}^n$ in (5.14),

$$\tilde{\epsilon}^n \stackrel{D}{=} \breve{\epsilon}^n. \tag{5.15}$$

We now show that

$$\breve{\epsilon}^{n_k} \Rightarrow 0, \tag{5.16}$$

as $n_k \to \infty$. From the continuity of the integrand operator and the convergence in (5.12),

$$\int_0^{\breve{V}^{n_k}(\cdot^-)} h(w)dw \to \int_0^{\breve{V}(\cdot^-)} h(w)dw,$$

almost surely, uniformly on compact sets of $[0,\infty)$, as $n_k \to \infty$, and so

$$\int_0^{\cdot} \left(\int_0^{\breve{V}^{n_k}(s^-)} h(w)dw - \int_0^{\breve{V}(s^-)} h(w)dw\right) ds \to 0,$$

almost surely, uniformly on compact sets of $[0,\infty)$, as $n_k \to \infty$. Since

$$\sqrt{n}\left(1-\exp\left(\frac{-x}{\sqrt{n}}\right)\right) \to x,$$

as $n \to \infty$, uniformly on compact sets, we find

$$\sqrt{n_k}\left(1-\exp\left(\frac{-\int_0^{\check{V}^{n_k}(\cdot^-)} h(w)dw}{\sqrt{n_k}}\right)\right) \to \int_0^{\check{V}(\cdot^-)} h(w)dw,$$

almost surely, uniformly on compact sets of $[0,\infty)$, as $n_k \to \infty$. Lemma 8.3 in Dai and Dai [6] and (5.12) then shows

$$\int_0^{\cdot}\left(\int_0^{\check{V}_k(s^-)} h(w)dw\right)ds - \int_0^{\cdot}\sqrt{n_k}\left(1-\exp\left(\frac{\int_0^{\check{V}^{n_k}(s^-)} h(w)dw}{\sqrt{n}}\right)\right)d\breve{A}^{n_k}(s) \to 0,$$

almost surely, uniformly on compact sets of $[0,\infty)$ as $n_k \to \infty$. We conclude from (5.13) that the weak convergence in (5.16) holds.

The distributional equivalence in (5.15) then implies $\tilde{\epsilon}^{n_k} \Rightarrow 0$ as $n_k \to \infty$. Since the choice of subsequence $\{n_k\}$ was arbitrary, we conclude

$$\tilde{\epsilon}^n \Rightarrow 0,$$

as $n \to \infty$. □

5.2 Weak Convergence under Assumption 2

We desire to represent the offered waiting time process in (4.5) in terms of the non-linear generalized two-sided regulator mapping. However, because V^n may sometimes exceed C^n (which is easily seen from the evolution equation (2.1)), we cannot directly represent V^n using the two-sided non-linear generalized regulator mapping. Instead, we introduce the process

$$\mathcal{V}^n(t) \equiv V^n(t) \wedge C^n \text{ for all } t \geq 0, \tag{5.17}$$

and observe that

$$V^n(t) = \mathcal{V}^n(t) + \delta^n(t), \tag{5.18}$$

where

$$\delta^n(t) \equiv \left[V^n(t) - C^n\right]^+. \tag{5.19}$$

The following lemma shows that V^n exceeds C^n less and less often and by smaller and smaller amounts in our heavy traffic asymptotic regime. Therefore, representing the process $\mathcal{V}^n$ in terms of the two-sided non-linear generalized regulator mapping allows us to use the same continuous mapping strategy as in our proof of part (i) in Subsection 5.1 to obtain weak convergence results for the process V^n. Let

$$\tilde{\delta}^n(t) = \sqrt{n}\delta^n(t). \tag{5.20}$$

Lemma 6 *Under assumption 2, as $n \to \infty$, $\tilde{\delta}^n \Rightarrow 0$.*

The processes $\mathcal{V}^n$, I^n, and U^n can be represented as follows

$$(\mathcal{V}^n, I^n, U^n) = \left(\phi^{h^n}_{C^n}, \psi^{h^n}_{1,C^n}, \psi^{h^n}_{2,C^n}\right)(X^n + \epsilon^n_B - \delta^n). \tag{5.21}$$

To see (5.21) is valid, first observe from (4.5), (5.17), (5.18), and because the process V^n is non-negative that $0 \leq \mathcal{V}^n(t) \leq C^n$ for all $t \geq 0$ and

$$\mathcal{V}^n(t) = (X^n(t) + \epsilon^n_B(t) - \delta^n(t)) - \int_0^t \left(\int_0^{\mathcal{V}^n(s^-)} h^n(u)du\right) ds + I^n(t) - U^n(t),$$

meaning condition (C1) of Definition 5 holds. Next, from the definitions of I^n in (2.6) and U^n in (4.4), for every n, I^n and U^n are non-decreasing, have $I^n(0) = U^n(0) = 0$, and

$$\int_0^\infty \mathcal{V}^n(t)dI^n(t) = \int_0^\infty (V^n(t) \wedge C^n)\,\mathbf{1}\{V^n(t) = 0\}dt = 0$$
$$\int_0^\infty [C^n - \mathcal{V}^n(t)]^+\, dU^n(t) = \frac{b^n}{n}\int_0^\infty [C^n - (V^n(t) \wedge C^n)]^+\,\mathbf{1}\{V^n(t^-) \geq C^n\}dA^n(t) = 0,$$

and so condition (C2) of Definition 5 also holds.

Although we cannot directly parallel the representation (5.4) in the proof of part (i) in Subsection 5.1, we can use the non-linear generalized two-sided regulator mapping and the scaled processes

$$\tilde{\mathcal{V}}^n(t) = \sqrt{n}\mathcal{V}^n(t) \tag{5.22}$$
$$\tilde{\epsilon}^n_B(t) = \sqrt{n}\epsilon^n_B(t) \tag{5.23}$$
$$\tilde{U}^n(t) = \sqrt{n}U^n(t) \tag{5.24}$$

to establish a representation for $\left(\tilde{\mathcal{V}}^n, \tilde{I}^n, \tilde{U}^n\right)$ that is similar in spirit. First observe from the scalings for $\tilde{\mathcal{V}}^n$, $\tilde{I}^n$, and $\tilde{U}^n$ in (5.22), (3.17), and (5.24), the representation of $(\mathcal{V}^n, I^n, U^n)$ in terms of the non-linear generalized two-sided regulator mapping in (5.21), the definition of C^n in (3.7), and the scaling property of the non-linear generalized two-sided regulator mapping in part (ii) of Proposition 4 that

$$\begin{aligned}\left(\tilde{\mathcal{V}}^n, \tilde{I}^n, \tilde{U}^n\right) &= \sqrt{n}\left(\phi^{h^n}_{C^n}, \psi^{h^n}_{1,C^n}, \psi^{h^n}_{2,C^n}\right)(X^n + \epsilon^n_B - \delta^n) \\ &= \left(\phi^h_C, \psi^h_{1,C}, \psi^h_{2,C}\right)\left(\tilde{X}^n + \tilde{\epsilon}^n_B - \tilde{\delta}^n\right),\end{aligned} \tag{5.25}$$

also recalling the scalings for $\tilde{X}^n$, $\tilde{\epsilon}^n_B$, and $\tilde{\delta}^n$ in (5.1), (5.23), and (5.20). The representation for V^n in terms of $\mathcal{V}^n$ and δ^n in (5.18) then implies

$$\left(\tilde{V}^n, \tilde{I}^n, \tilde{U}^n\right) = \left(\tilde{\mathcal{V}}^n, \tilde{I}^n, \tilde{U}^n\right) + \left(\tilde{\delta}^n, 0, 0\right). \tag{5.26}$$

We parallel the proof of part (i) in Subsection 5.1 to show

$$\left(\tilde{\mathcal{V}}^n, \tilde{I}^n, \tilde{U}^n\right) \Rightarrow \left(\phi^h_C, \psi^h_{1,C}, \psi^h_{2,C}\right)(W), \tag{5.27}$$

as $n \to \infty$. Lemma 6, the continuous mapping theorem, and the equality (5.26) then establish the result stated in part (ii) of Theorem 1.

From the representation (5.25), Lemma 6, the continuous mapping theorem, and part (iv) of Proposition 4, establishing

$$\tilde{X}^n \Rightarrow W \text{ and } \tilde{\epsilon}^n_B \Rightarrow 0, \tag{5.28}$$

as $n \to \infty$, is sufficient to show (5.27). We require the following three Lemmas, which are the equivalents of Lemmas 3-5 when abandonment times are bounded.

Lemma 7 *Under assumption 2, as $n \to \infty$, $\overline{R}^n \Rightarrow 0$.*

Lemma 8 *Under assumption 2, as $n \to \infty$, $\tilde{M}^n_a \Rightarrow 0$.*

Lemma 9 *The sequence $\{\tilde{\mathcal{V}}^n\}$ is tight in $D([0,\infty),\Re)$.*

By Lemmas 7 and 8, the arguments showing $\tilde{X}^n \Rightarrow W$ as $n \to \infty$ in Subsection 5.1.1 remain valid. The definitions of ϵ^n_B and $\tilde{\epsilon}^n_B$ in (4.6) and (5.23), the definitions of $\mathcal{V}^n$ and $\tilde{\mathcal{V}}^n$ in (5.17) and (5.22), and the representation of F^n in (3.8) show

$$\tilde{\epsilon}^n_B(t) = \int_0^t \left(\int_0^{\tilde{\mathcal{V}}^n(s^-)} h(w)dw \right) ds - \int_0^t \sqrt{n} \left(1 - \exp\left(-\frac{1}{\sqrt{n}} \int_0^{\tilde{\mathcal{V}}^n(s^-)} h(w)dw \right) \right) d\overline{A}^n(s). \tag{5.29}$$

The representation of $\tilde{\epsilon}^n_B$ in (5.29) above has exactly the same form as that for $\tilde{\epsilon}^n$ in (5.14) in the proof of part (i) in Subsection 5.1, with $\tilde{\mathcal{V}}^n(s^-)$ replacing $\tilde{V}^n(s^-)$. Therefore, because Lemma 9 establishes the sequence $\{\tilde{\mathcal{V}}^n\}$ is tight in $D([0,\infty),\Re)$, the arguments in Subsection 5.1.2 showing $\tilde{\epsilon}^n \Rightarrow 0$ as $n \to \infty$ also show $\tilde{\epsilon}^n_B \Rightarrow 0$, as $n \to \infty$. □

6 Stationary Performance Measure Approximation

We first show in Subsection 6.1 that the asymptotic behavior of the diffusion-scaled queue-length and offered waiting time processes are identical. Next, in Subsection 6.2, we derive the stationary distributions of the limiting diffusion processes in Theorem 1, which can be used to approximate steady-state performance measures for a GI/GI/1 queue with abandonments. Finally, we perform a simulation study in Subsection 6.3 to evaluate the accuracy of our proposed steady-state performance measure approximations.

6.1 An Asymptotic Relationship Between the Queue-length and Offered Waiting Time Processes

We establish an asymptotic relationship between the queue-length and offered waiting time processes identical to that in Theorem 4 in Section 3 in Reiman [27] for a conventional GI/GI/1 queue. To handle the complications imposed by the presence of customers that may abandon the system before receiving service, we require the following Lemma. For $t \geq 0$, let $a^n(t)$ be the arrival time of the customer in service at time t in the nth system. If the server is idle, let $a^n(t) = t$.

Lemma 10 *Under either assumption 1 or 2,*

$$n^{-1/2} \sum_{i=A^n \circ a^n(\cdot)}^{A^n(\cdot)} \mathbf{1}\{V^n\left(t_i^{n,-}\right) \geq a_i^n\} \Rightarrow 0,$$

as $n \to \infty$.

Let $Q^n(t)$ be the queue-length at time $t \geq 0$ in the n^{th} system, and $\tilde{Q}^n(t) = n^{-1/2}Q^n(t)$ be the diffusion-scaled queue-length.

Theorem 2 *Under either assumption 1 or 2,*

$$\tilde{Q}^n - \tilde{V}^n \Rightarrow 0,$$

as $n \to \infty$.

For the proofs of both Lemma 10 and Theorem 2, it is useful to notice that the convergence in (16) of Theorem 4 in Section 3 in [27] continues to hold in our setting. In particular, because the server works at rate 1 and the system is FIFO,

$$V^n(a^n(t)^-) \leq t - a^n(t) \leq V^n(a^n(t)^-) + v^n_{A^n(a^n(t))},$$

and so, recalling the scaling of the service times in (3.1) and the definition of $\tilde{V}^n$ in (3.12),

$$\tilde{V}^n\left(a^n(t)^-\right) \leq \sqrt{n}\left(t - a^n(t)\right) \leq \tilde{V}^n\left(a^n(t)^-\right) + \frac{v_{A^n(a^n(t))}}{\sqrt{n}}.$$

Because $\sup_{k=1,\ldots,nt} n^{-1/2}v_k \Rightarrow 0$ as $n \to \infty$ from Lemma 3 in Iglehart and Whitt [12], for each $T \geq 0$,

$$\sup_{0\leq t\leq T}\left|\sqrt{n}\left(t - a^n(t)\right) - \tilde{V}^n\left(a^n(t)^-\right)\right| \Rightarrow 0, \tag{6.1}$$

as $n \to \infty$, which dividing by $\sqrt{n}$, implies

$$a^n \Rightarrow e, \tag{6.2}$$

as $n \to \infty$.

Proof of Theorem 2: Since the service discipline is FIFO, the number of customers currently in queue is less than the number that have arrived after the customer currently in service plus one, $A^n(t) - A^n(a^n(t)) + 1$. Additionally, the current queue-length exceeds $A^n(t) - A^n(a^n(t))$ minus the number of customers that have arrived after the one currently in service that will eventually abandon, and so

$$A^n(t) - A^n(a^n(t)) - \sum_{i=A^n(a^n(t))}^{A^n(t)} \mathbf{1}\{V^n(t_i^{n,-}) \geq a_i\} \leq Q^n(t) \leq A^n(t) - A^n(a^n(t)) + 1,$$

or, also using the definition of $\tilde{A}^n$ in (3.13),

$$\tilde{A}^n(t) - \tilde{A}^n(a^n(t)) + \sqrt{n}\rho^n(t - a^n(t)) - n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} \mathbf{1}\{V^n(t_i^{n,-}) \geq a_i^n\}$$
$$\leq \tilde{Q}^n(t) \leq \tilde{A}^n(t) - \tilde{A}^n(a^n(t)) + \sqrt{n}\rho^n(t - a^n(t)) + n^{-1/2}.$$

Subtracting $\tilde{V}^n$ from all sides and adding and subtracting several terms shows

$$\begin{aligned} \left|\tilde{Q}^n(t) - \tilde{V}^n(t)\right| \quad \leq \quad & \left|\tilde{A}^n(t) - \tilde{A}^n(a^n(t))\right| + n^{-1/2} + \left|\tilde{V}^n(t)(\rho^n - 1)\right| \\ & + \left|\rho^n \left(\sqrt{n}(t - a^n(t)) - \tilde{V}^n(a^n(t)^-)\right)\right| + \left|\rho^n \left(\tilde{V}^n(a^n(t)^-) - \tilde{V}^n(t)\right)\right| \\ & + n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} \mathbf{1}\left\{V^n(t_i^{n,-}) \geq a_i^n\right\}. \end{aligned} \tag{6.3}$$

The weak convergence of a^n to the identity process in (6.2), the functional central limit theorem, and Theorem 1 imply

$$\tilde{A}^n - \tilde{A}^n \circ a^n \Rightarrow 0 \text{ and } \tilde{V}^n - \tilde{V}^n \circ a^n \Rightarrow 0, \tag{6.4}$$

as $n \to \infty$. Finally, the inequality (6.3), (6.4), the convergence $\rho^n \to 1$ as $n \to \infty$ in (3.3), the weak convergence in (6.1), and Lemma 10 imply the stated result. □

6.2 Approximating the Stationary Distribution of the Offered Waiting Time Process

We establish the stationary distributions of the diffusions $\phi^h(W)$ and $\phi^h_C(W)$, and also the average pushing at the upper boundary for the diffusion $\phi^h_C(W)$. We write the stationary distributions in terms of the cumulative hazard rate function $H(x) = \int_0^x h(y)dy$ in order to provide intuition on the condition that the diffusion $\phi^h(W)$ has a unique stationary distribution. Specifically, observe in condition (i) in Proposition 5 below that if $\theta \leq 0$, then a unique stationary distribution exists because $\phi^h(W)$ is a negative drift diffusion with reflection at the origin. Also, if $\theta > 0$ and there exists z_0 such that $H(z) > \theta$ for all $z > z_0$, then again a unique stationary distribution exists and $\phi^h(W)$ will drift towards $z^\star \equiv \{z : H(z) = \theta\}$ similar to the conventional Ornstein-Uhlenbeck process. Otherwise, if neither of the aforementioned conditions is satisfied, $\phi^h(W)$ has a positive drift and so a stationary distribution does not exist.

Proposition 5 *Let W be a Brownian motion with drift θ, variance σ^2.*

(i) Suppose there exists z_0 such that $H(z) > \theta$ for all $z > z_0$. Then, the one-sided regulated diffusion $\phi^h(W)$ has a unique stationary distribution π with density

$$p(x) = M \exp\left(\frac{2}{\sigma^2}\left(\theta x - \int_0^x H(s)ds\right)\right), \; x \geq 0,$$

where M is such that $\int_0^\infty p(x)dx = 1$.

(ii) The two-sided regulated diffusion $\phi_C^h(W)$ has a unique stationary distribution π_C with density

$$p_C(x) = M_C \exp\left(\frac{2}{\sigma^2}\left(\theta x - \int_0^x H(s)ds\right)\right),\ 0 \le x \le C,$$

and average pushing at the upper boundary

$$\lim_{t\to\infty} t^{-1}E\left[\psi_{2,C}^h(W)(t)\right] = \frac{\sigma^2}{2}\frac{\exp\left(-\int_0^C \frac{2}{\sigma^2}(-\theta + H(x))dx\right)}{\int_0^C \exp\left(-\int_0^y \frac{2}{\sigma^2}(-\theta + H(x))dx\right)dy},$$

where M_C is such that $\int_0^C p(x)dx = 1$.

Furthermore, for any $x \in \Re$, as $t \to \infty$,

$$P\left(\phi^h(W)(t) \le x\right) \to \pi(x) \text{ and } P\left(\phi_C^h(W)(t) \le x\right) \to \pi_C(x). \tag{6.5}$$

The proof of Proposition 5 requires the following Lemma to establish (6.5).

Lemma 11 *Let W be a Brownian Motion with drift θ and variance σ^2. For each $x \in \Re$, let P_x be a probability measure such that the Brownian motion W has initial position $W(0) = x$. Suppose* $\lim_{z\to\infty} H(z) > \theta$. *Let*

$$\begin{aligned} T_0 &\equiv \inf\{t \ge 0 : \phi^h(W)(t) = 0\} \\ T_0^C &\equiv \inf\{t \ge 0 : \phi_C^h(W)(t) = 0\} \end{aligned}$$

Then,

$$\begin{aligned} P_x\left(T_0 < \infty\right) &= 1,\ \ x \ge 0 \\ P_x\left(T_0^C < \infty\right) &= 1,\ \ 0 \le x \le C. \end{aligned}$$

Proof of Proposition 5: Echeverria [8] shows that a stationary distribution π of $\phi^h(W)$ ought to satisfy

$$\int_0^\infty (Af)(y)\pi(dy) = 0 \tag{6.6}$$

for all bounded f that are twice continuously differentiable on $[0, \infty)$ and satisfy $f'(0) = 0$, where

$$Af(y) \equiv \left(\theta - \int_0^y h(u)du\right) f^{'}(y) + \frac{\sigma^2}{2} f^{''}(y).$$

Similarly, a stationary distribution π_C of $\phi_C^h(W)$ ought to satisfy

$$\int_0^C (Af)(y)\pi_C(dy) = 0 \tag{6.7}$$

for all bounded f that are twice continuously differentiable on $[0, C]$ and satisfy $f'(0) = f'(C) = 0$. It is straightforward to verify (using integration by parts) that π and π_C satisfy (6.6) and

Reneging Distribution	E[queue-length] Simulated	Approximated	% Error	P[abandon] Simulated	Approximated	% Error
Deterministic(1)	51.691	50	3.38%	0.004802	0.005	4.12%
$G(5)$	20.5980	19.8141	3.81%	0.013138	0.013324	1.41%
$G(2)$	11.2930	10.6410	5.77%	0.025864	0.026900	4.01%
$G(1)$=Exponential(1)	6.2585	5.6419	9.85%	0.052728	0.056419	7.00%
$G(0.5)$	3.2545	2.7749	14.74%	0.11273	0.13005	15.36%
$G(0.2)$	1.5029	1.14888	23.56%	0.24128	0.360384	49.36%

Table 2: A comparison of the simulated mean queue-length and abandonment probability for a GI/GI/1+GI queue with Poisson arrivals at rate 100 per unit, deterministic service with mean 1/100, and abandonment times distributed as given in Column 1.

(6.7) respectively for the desired f. Identical arguments as in the proof of Proposition 1 in [34] then establish that π and π_C are stationary distributions of $\phi^h(W)$ and $\phi^h_C(W)$ respectively.

The average pushing at the upper boundary, $\lim_{t\to\infty} t^{-1}E\left[\psi^h(W)(t)\right]$, follows by arguments mimicking those used to prove Propositions 8 and 9 in [1].

Finally, (6.5) and the uniqueness of the stationary distribution follow as in Proposition 1 of [34], because for any $x \geq 0$ $(0 \leq x \leq C)$, the probability the diffusion $\phi^h(W)$ $(\phi^h_C(W))$ hits 0 in finite time is equal to one by Lemma 11. □

6.3 Evaluation of the Proposed Diffusion Approximations via Simulation

We begin with a simulation study that explores the effect of variability in the abandonment distribution. As in the introduction, let $G(p)$ be the distribution function associated with a mean 1 gamma random variable having scale and shape parameter p. Observe that such gamma distributions are ordered in variability by the parameter p since $\text{Var}(G(p)) = \frac{1}{p}$ increases as p decreases. The variance of a deterministic distribution is 0, and so the results presented in Table 2 are ordered according to the variability of the abandonment distribution.

Each simulation run presented in Table 2 assumes Poisson arrivals having rate 100, deterministic service with mean 0.01, and is run to 50,000 time units so has approximately 5,000,000 arrivals. The abandonment distribution varies according to the first column. Recall the drift in our suggested diffusion from (1.2) for the case that abandonment times have a gamma distribution. In the case that abandonment times are deterministic, $F^n(x) = \mathbf{1}\{x \geq C/\sqrt{n}\}$, and so (noting the relationship $H(x) = -\ln(1 - G(x))$ between a distribution function G and its associated cumulative hazard function H)

$$H^n(x) = -\ln\left(1 - F^n(x)\right) = 0, \text{ for } x < \frac{C}{\sqrt{n}},$$

which from (1.1) implies our suggested approximating diffusion has $H^n_D(x) = 0$ for $x < C$. In both cases, the θ appearing in parts (i) and (ii) of Proposition 5 is 0 and the variance σ^2 is 1.

n	E[queue-length] Simulated	Approximated	% Error	P[abandon] Simulated	Approximated	% Error
1000	2.1454	1.793910	16.38%	0.17715	0.233068	31.57%
10,000	3.1025	2.73777	11.76%	0.12769	0.153616	20.30%
100,000	4.4736	4.121112	7.88%	0.090415	0.102426	13.28%
1,000,000	6.5393	6.15025	5.95%	0.063077	0.0687956	9.07%
10,000,000	9.7142	9.127725	6.04%	0.043817	0.046426	5.95%
100,000,000	13.9040	13.4975	2.92%	0.030456	0.0314284	3.19%

Table 3: A comparison of the simulated mean queue-length and abandonment probability for a GI/GI/1+GI queue with Poisson arrivals at rate n per unit, deterministic service with mean $1/n$, and abandonment times distributed $G(0.2)$.

Note that the stationary density in part (ii) of Proposition 5 reduces to the uniform distribution on $[0, C]$ for the case of constant service times, which not surprisingly coincides with the limiting result in Theorem 2.1 and Remark 2.2 in Whitt [37] for a standard GI/GI/1 queue with finite waiting room. (Intuitively, a GI/GI/1 queue with deterministic abandonment times a and arrival rate λ should resemble a GI/GI/1/λa queue.)

We calculate the approximated steady-state queue-length using Proposition 5. When $\rho^n \uparrow 1$ as $n \to \infty$, the validity of the desired limit interchange follows from the results of Kingman [14] [16] since the queue-length process in a conventional $GI/GI/1$ queue dominates that in a $GI/GI/1+GI$ queue with identical arrival and service processes. Otherwise, when $\rho^n \downarrow 1$ as $n \to \infty$, we assume the validity of the desired limit interchange.

To approximate the probability a customer abandons the system, first observe that $F^n\left(V^n(t_i^{n,-})\right)$ is the probability the ith customer abandons, given the offered waiting time at his arrival. Recalling the definitions of ϵ^n and $\tilde{\epsilon^n}$ in (4.3) and (5.2), and the weak convergence $\tilde{\epsilon}^n \Rightarrow 0$ as $n \to \infty$ proved in Subsection 5.1.2, under Assumption 1, we find

$$\begin{aligned}
\sqrt{n}\frac{\int_0^t F^n\left(V^n(s^-)\right)dA^n(s)}{A^n(t)} &= \frac{n}{A^n(t)}\left(-\tilde{\epsilon}^n(t) + \int_0^t \left(\int_0^{\tilde{V}^n(s^-)} h(w)dw\right)ds\right) \\
&\Rightarrow t^{-1}\int_0^t \left(\int_0^{\phi^h(W)(s)} h(w)dw\right)ds,
\end{aligned}$$

as $n \to \infty$, by part (i) of Theorem 1 and the continuous mapping theorem. Assuming the interchange of limit and expectation, we find that as $n \to \infty$,

$$\begin{aligned}
\sqrt{n}E_\pi\left[\frac{\int_0^t F^n\left(V^n(s^-)\right)dA^n(s)}{A^n(t)}\right] &\to t^{-1}\int_0^t E_\pi\left[\int_0^{\phi^h(W)(s)} h(w)dw\right]ds \qquad (6.8) \\
&= \int_0^\infty \left(\int_0^x h(w)dw\right)p(x)dx,
\end{aligned}$$

when the system operates in steady-state, where p is as given in part (i) of Proposition 5. Similarly, under Assumption 2, recalling the definitions of ϵ_B^n and $\tilde{\epsilon}_B^n$ in (4.6) and (5.23), and the weak convergence $\tilde{\epsilon}_B^n \Rightarrow 0$ as $n \to \infty$ argued in the proof of part (ii) of Theorem 1,

$$\begin{aligned}\sqrt{n}\frac{\int_0^t F^n\left(V^n(s^-)\right)dA^n(s)}{A^n(t)} &= \frac{n}{A^n(t)}\left(-\tilde{\epsilon}_B^n(t)+\int_0^t\left(\int_0^{\tilde{V}^n(s^-)\wedge C} h(w)dw\right)ds+\tilde{U}^n(t)\right)\\ &\Rightarrow t^{-1}\left(\int_0^t\left(\int_0^{\phi_C^h(W)(s)} h(w)dw\right)ds+\psi_{2,C}^h(W)(t)\right),\end{aligned}$$

as $n \to \infty$. Then, by part (ii) of Theorem 1 and the continuous mapping theorem, assuming the interchange of limit and expectation, we find that, as $n \to \infty$,

$$\begin{aligned}\sqrt{n}E_{\pi_C}\left[\frac{\int_0^t F^n\left(V^n(s^-)\right)dA^n(s)}{A^n(t)}\right] &\to \int_0^\infty\left(\int_0^x h(w)dw\right)p_C(x)dx \\ &\quad+\frac{\sigma^2}{2}\frac{\exp\left(-\int_0^C\frac{2}{\sigma^2}\left(-\theta+H(x)\right)dx\right)}{\int_0^C\exp\left(-\int_0^y\frac{2}{\sigma^2}\left(-\theta+H(x)\right)dx\right)dy},\end{aligned} \tag{6.9}$$

where p_C is as given in part (ii) of Proposition 5. We use the formulas in (6.8) and (6.9) to approximate the probability an arriving customer will abandon the system.

Observe the loss in approximation accuracy as the variability of the abandonment distribution increases. This is consistent with our theoretical results which, from Lemmas 3 and 7, suggest that our approximations perform better as the fraction of abandoning customers decreases. Increasing the variability of the abandonment distribution increases the recorded abandonmnet probability in Table 2.

However, even highly variable abandonment distributions, such as $G(0.2)$, lead to accurate approximations for fast enough arrival and service rates, and correspondingly small abandonment rates. Table 3 considers the worst performing case in Table 2, the $G(0.2)$ case, and shows that the accuracy in our approximations increases as the abandonment probability becomes small. Specifically, we simulate a single-server queue with abandonments having Poisson arrivals at rate n, deterministic service times $1/n$, and customer abandonment times that follow a $G(0.2)$ distribution. We run each simulation to time $n^{-1}5,000,000$ so that approximately $5,000,000$ arrivals occur. Observe that for $n \geq 100,000$, the error in our expected queue-length approximation is under 10%, and for $n \geq 1,000,000$, the error in our approximation for the probability a customer abandons is under 10%.

7 Appendix

We devote the appendix to proving Lemmas 1- 11. It is useful for the proof of Lemma 5 to define the following notation, which matches that in Billingsley [3]. For any set $S \subset [0,T], \delta > 0$, and $x \in D([0,\infty),\Re)$, let

$$w(x,S) \equiv \sup_{u,v\in S}|x(u)-x(v)| \tag{7.1}$$

and

$$w_T(x,\delta) \equiv \sup_{0\le t\le T-\delta} w(x,[t,t+\delta]). \tag{7.2}$$

Also let

$$w'_T(x,\delta) \equiv \inf \max_{1\le i\le v} w\left(x,[t_{i-1},t_i)\right),$$

where the infimum extends over all decompositions $[t_{i-1},t_i)$, $1 \le i \le v$, of $[0,T)$ such that $t_i - t_{i-1} > \delta$ for $1 \le i < v$. Similar to (12.7) in [3], since $[0,T)$ can, for each $\delta < \frac{T}{2}$ be split into subintervals $[t_{i-1},t_i)$ satisfying $\delta < t_i - t_{i-1} \le 2\delta$, for $x \in D([0,\infty),\Re)$,

$$w'_T(x,\delta) \le w_T(x,2\delta),\ \delta < \frac{T}{2}. \tag{7.3}$$

Proof of Lemma 1:
Part (i), Existence: Let $T > 0$. Because the function $\eta : D([0,\infty),\Re) \to D([0,\infty),\Re^+)$

$$\eta(w)(t) \equiv \int_0^{\phi(w)(t)} h(\zeta)d\zeta$$

is not Lipschitz cotinuous[3], Lemma 1 in Reed and Ward [25] is not directly applicable. However, define w_0 to be the zero process, and

$$w_{n+1}(t) = x(t) - \int_0^t \left(\int_0^{\phi(w_n)(s)} h(u)du\right) ds. \tag{7.4}$$

Suppose there exists $M > 0$ such that

$$\|w_n\|_T \le M \text{ for all } n \ge 0. \tag{7.5}$$

Then, since the definition of ϕ in (4.7) implies

$$\|\phi(w_n)\|_T \le 2\|w_n\|_T \le 2M,$$

we find that for any $0 \le s \le T$,

$$\int_0^{\phi(w_n)(s)} h(\zeta)d\zeta = \int_0^{\phi(w_n)(s)} \left(h(\zeta) \wedge \|h\|_{2M}\right) d\zeta,$$

and so

$$\begin{aligned} \|\eta(w_{n+1}) - \eta(w_n)\|_T &= \sup_{0\le t\le T} \int_{\phi(w_n)(t)\wedge\phi(w_{n+1})(t)}^{\phi(w_n)(t)\vee\phi(w_{n+1}(t))} \left(h(\zeta) \wedge \|h\|_{2M}\right) d\zeta \\ &\le \|h\|_{2M}\|\phi(w_{n+1}) - \phi(w_n)\|_T \\ &\le 2\|h\|_{2M}\|w_{n+1} - w_n\|_T, \end{aligned} \tag{7.6}$$

where the last inequality follows from the Lipschitz continuity of ϕ noted in (4.13). The inequality (7.6) implies the arguments used to prove existence in Lemma 1 of Reed and Ward [25] are valid when the constant κ in their proof is taken to be $2\|h\|_{2M}$.

[3] Note the function η would be Lipschitz continuous if h were bounded on $[0,\infty)$.

We now show (7.5) to complete the proof. Since the definition of ϕ in (4.7) implies $\phi(w_0)(0) = 0$, $w_1 = x$. Next, from (7.4), because h is assumed non-negative and ϕ defined in (4.7) is also positive,

$$w_n \leq w_1 \text{ for all } n \geq 1. \tag{7.7}$$

Lemma 5.1 in Kruk et al [17] establishes that for all $n \geq 2$

$$\phi\left(w_n\right) \leq \phi\left(w_1\right). \tag{7.8}$$

To see the conditions of Lemma 5.1 in [17] are satisfied, observe that

$$w_n(t) = w_1(t) - \int_0^t \left(\int_0^{\phi(w_{n-1})(s)} h(u)du \right) ds$$

is written as the difference of w_1 and a non-decreasing function. Use of (7.8) shows that for all $n \geq 3$

$$\begin{aligned} w_n(t) &= x(t) - \int_0^t \left(\int_0^{\phi(w_{n-1})(s)} h(u)du \right) ds \\ &\geq x(t) - \int_0^t \left(\int_0^{\phi(w_1)(s)} h(u)du \right) ds = w_2(t), \end{aligned} \tag{7.9}$$

for all $t \geq 0$, since h is assumed non-negative. Combining (7.7) and (7.9), we conclude

$$w_2 \leq w_n \leq w_1 \text{ for all } n \geq 2. \tag{7.10}$$

Set $M = \|w_1\|_T \vee \|w_2\|_T$. Then, (7.10) implies (7.5) is valid.

Part (i), Uniqueness: Suppose both u and v satisfy (4.11). As in (7.8) in the proof of existence, Lemma 5.1 in Kruk et al [17] establishes

$$\|\phi(u)\|_T \leq \|\phi(x)\|_T \text{ and } \|\phi(v)\|_T \leq \|\phi(x)\|_T. \tag{7.11}$$

Set $N \equiv \|\phi(x)\|_T$. Use of the integral equation definition in (4.11) and (7.11) shows

$$\begin{aligned} \triangle(t) \equiv u(t) - v(t) &= \int_0^t \left(\int_0^{\phi(v)(s)} h(x)dx - \int_0^{\phi(u)(s)} h(x)dx \right) ds \\ &= \int_0^t \left(\int_0^{\phi(v)(s)} \left(h(\zeta) \wedge ||h||_N\right) d\zeta - \int_0^{\phi(u)(s)} \left(h(\zeta) \wedge ||h||_N\right) d\zeta \right) ds, \end{aligned}$$

and so the Lipschitz continuity of ϕ in (4.13) implies

$$\begin{aligned} |\triangle(t)| &\leq \int_0^t \left| \int_0^{\phi(v)(s)} \left(h(\zeta) \wedge ||h||_N\right) d\zeta - \int_0^{\phi(u)(s)} \left(h(\zeta) \wedge ||h||_N\right) d\zeta \right| ds \\ &\leq \int_0^t ||h||_N \left|\phi(v)(s) - \phi(u)(s)\right| ds \\ &\leq 2t||h||_N \|\triangle\|_t, \end{aligned}$$

which implies $\triangle(t)=0$ for all $0\le t\le (2||h||_N)^{-1}$. For $(2||h||_N)^{-1}<t<2\,(2||h||_N)^{-1}$,

$$\begin{aligned}|\triangle(t)| &\le \|\triangle\|_{(2||h||_N)^{-1}}+\left(t-(2||h||_N)^{-1}\right)2||h||_N\|\triangle\|_{2(2||h||_N)^{-1}}\\ &\le \|\triangle\|_{2(2||h||_N)^{-1}},\end{aligned}$$

and so $\triangle(t)=0$ for all $0\le t\le 2\,(2||h||_N)^{-1}$. Continued iteration of this argument implies $\triangle=\vec{0}$.

Part (ii): From the Lipschitz continuity of ϕ noted in (4.13) and assumption,

$$\begin{aligned}\|\phi\left(x_j\right)\|_T &\le \|\phi(x)\|_T+\|\phi(x_j)-\phi(x)\|_T \\ &\le \|\phi(x)\|_T+2,\quad j\in\{1,2\}.\end{aligned}\tag{7.12}$$

As in (7.8) in the proof of existence in part (i), Lemma 5.1 in Kruk et al [17] establishes

$$\|\phi\left(\mathcal{M}^h(x_j)\right)\|_T\le\|\phi(x_j)\|_T,\ j\in\{1,2\},$$

and so from (7.12),

$$\|\phi\left(\mathcal{M}^h(x_1)\right)\|_T\vee\|\phi\left(\mathcal{M}^h(x_1)\right)\|_T\le\|\phi(x)\|_T+2.\tag{7.13}$$

Set $\bar{c}\equiv\|\phi(x)\|_T+2$, and observe from the definition of the mapping $\mathcal{M}^h$ in (4.11), the inequality (7.13), and the Lipschitz continuity of ϕ noted in (4.13), that for $0\le t\le T$,

$$\begin{aligned}\|\mathcal{M}^h(x_1)-\mathcal{M}^h(x_2)\|_t &\le \|x_1-x_2\|_t+\sup_{0\le s\le t}\left|\int_0^s\left(\int_{\phi(\mathcal{M}^h(x_1))(u)}^{\phi\left(\mathcal{M}^h(x_2)\right)(u)}h(\zeta)d\zeta\right)du\right|\\ &\le \|x_1-x_2\|_t+t\|h\|_{\bar{c}}\|\phi\left(\mathcal{M}^h\left(x_2\right)\right)-\phi\left(\mathcal{M}^h\left(x_1\right)\right)\|_t\\ &\le \|x_1-x_2\|_t+2t\|h\|_{\bar{c}}\|\mathcal{M}^h\left(x_2\right)-\mathcal{M}^h\left(x_1\right)\|_t.\end{aligned}$$

Therefore, for any $0\le t\le(4\|h\|_{\bar{c}})^{-1}$,

$$\|\mathcal{M}^h(x_2)-\mathcal{M}^h(x_1)\|_t\le\frac{\|x_1-x_2\|_t}{1-2\|h\|_{\bar{c}}t}.\tag{7.14}$$

For $(4\|h\|_{\bar{c}})^{-1}<t\le(2\|h\|_{\bar{c}})^{-1}$,

$$\begin{aligned}\|\mathcal{M}^h(x_2)-\mathcal{M}^h(x_1)\|_t &\le \|\mathcal{M}^h(x_2)-\mathcal{M}^h(x_1)\|_{\frac{1}{4\|h\|_{\bar{c}}}}+2\|x_1-x_2\|_t\\ &\quad+\sup_{0\le s\le t}\left|\int_{\frac{1}{4\|h\|_{\bar{c}}}}^s\left(\int_{\phi(\mathcal{M}^h(x_1))(u)}^{\phi\left(\mathcal{M}^h(x_2)\right)(u)}h(\zeta)d\zeta\right)du\right|,\end{aligned}$$

and so, also using (7.14), the inequality (7.13), and (4.13),

$$\|\mathcal{M}^h(x_2)-\mathcal{M}^h(x_1)\|_t\le 4\|x_1-x_2\|_t+2\left(t-\frac{1}{4\|h\|_{\bar{c}}}\right)\|h\|_{\bar{c}}\|\mathcal{M}^h(x_2)-\mathcal{M}^h(x_1)\|_t,$$

or

$$\|\mathcal{M}^h(x_2) - \mathcal{M}^h(x_1)\|_t \leq \frac{4\|x_1 - x_2\|_t}{1 - 2\|h\|_{\overline{c}}\left(t - \frac{1}{4\|h\|_{\overline{c}}}\right)}.$$

Since only a finite number of intervals of length $(4\|h\|_{\overline{c}})^{-1}$ partition the interval $[0, T]$, continued iteration of the above argument establishes

$$\|\mathcal{M}^h(x_2) - \mathcal{M}^h(x_1)\|_t \leq \kappa\|x_1 - x_2\|_t$$

for κ finite (but dependent on x through $\overline{c}$) and any $0 \leq t \leq T$.

Part (iii): Proof. Let $w \in D([0,\infty),\Re)$ be the unique solution to (4.11) for $x \in D([0,\infty),\Re)$. Note that since $x \in D([0,\infty),\Re)$, it follows that for each $T \geq 0$, there exists an $M \geq 0$ such that $\sup_{0\leq t\leq T}|x(t)| \leq (M/2 - 1)$. The following observation will be useful. Because $\int_0^t \left(\int_0^{\phi(w)(s)} h(u)du\right) ds$ is non-decreasing in t, Lemma 5.1 in Kruk et al shows $\phi(w) \leq \phi(x)$. Thus, since for any $T \geq 0$, $\phi(x)(t) \leq 2\sup_{0\leq s\leq t}|x(s)| \leq M - 2$, the following inequality is valid

$$\phi(w)(t) \leq M - 2 \text{ for all } 0 \leq t \leq T. \tag{7.15}$$

Suppose now that $x^n \to x$ as $n \to \infty$ in the Skorohod J_1 topology and let T be a continuity point of x. Then, there must exist a sequence of absolutely continuous homeomorphisms $\{\lambda^n\}$ of $[0, T]$ such that

$$||x^n \circ \lambda^n - x||_T \vee ||\lambda^n - e||_T \to 0.$$

Furthermore, it suffices to consider absolutely continuous homeomorphisms, such that

$$||x^n \circ \lambda^n - x||_T \vee ||\dot{\lambda}^n - 1||_T \to 0$$

as $n \to \infty$, see Billingsley [3] for more details. Also, for n sufficiently large we have that $\sup_{0\leq t\leq T}|x^n(t)| \leq \sup_{0\leq t\leq T}|x(t)| + 1 \leq M/2$ and so reasoning similar to the above implies that

$$\phi(w^n)(t) \leq M \text{ for all } 0 \leq t \leq T,$$

where $w^n = \mathcal{M}^h(x^n)$ is the solution to (4.11) for x^n.

Now, for all $0 \le t \le T$ and n sufficiently large,

$$\begin{aligned} ||w^n \circ \lambda^n - w||_t &= \left\| x^n \circ \lambda^n - x - \int_0^{\lambda^n} \left(\int_0^{\phi(w^n)(s)} h(u)du \right) ds - \int_0^e \left(\int_0^{\phi(w)(s)} h(u)du \right) ds \right\|_t \quad (7.16) \\ &= \left\| x^n \circ \lambda^n - x - \int_0^e \left(\int_0^{\phi(w^n)(\lambda^n(s))} h(u)du \right) \dot{\lambda}^n(s)ds - \int_0^e \left(\int_0^{\phi(w)(s)} h(u)du \right) ds \right\|_t \\ &\le ||x^n \circ \lambda^n - x||_t + ||\dot{\lambda}^n(s) - 1||_t \int_0^t \left(\int_0^{\phi(w^n)(\lambda^n(s))} h(u)du \right) ds \\ &\quad + \left\| \int_0^e \left(\int_0^{\phi(w^n)(\lambda^n(s))} h(u)du - \int_0^{\phi(w)(s)} h(u)du \right) ds \right\|_t . \\ &\le \|x^n \circ \lambda_n - x\|_t + \|\dot{\lambda}^n(s) - 1\|_t \|h\|_M MT \\ &\quad + \|h\|_M \int_0^t \sup_{0 \le s \le t} |\phi(w_n)(\lambda^n(s)) - \phi(w)(s)| \, ds. \end{aligned}$$

By Lemma 13.5.2 in [36] $\phi(w^n)(\lambda^n(s)) = \phi(w^n \circ \lambda^n)(s)$ and by Lemma 13.5.1 in [36], ϕ is Lipschitz continuous with respect to the uniform metric with Lipschitz constant 2, and so

$$\begin{aligned} &\int_0^t \sup_{0 \le s \le t} |\phi(w^n)(\lambda^n(s)) - \phi(w)(s)| \, ds \quad (7.17) \\ &= \int_0^t \sup_{0 \le s \le t} |\phi(w^n \circ \lambda^n)(s) - \phi(w)(s)| \, ds \\ &\le \int_0^t 2\|w^n \circ \lambda^n - w\|_s ds. \end{aligned}$$

We conclude from (7.16) and (7.17) that

$$\|w^n \circ \lambda^n - w\|_t \le \|x^n \circ \lambda^n - x\|_t + \|\dot{\lambda}^n(s) - 1\|_t \|h\|_M MT + 2\|h\|_M \int_0^t \|w^n \circ \lambda^n - w\|_s ds.$$

Let $\varepsilon > 0$ be arbitrarily small. Then, there exists n_0 such that

$$\|x^n \circ \lambda^n - x\|_t + \|\dot{\lambda}^n(s) - 1\|_t \|h\|_M MT \le \epsilon$$

for $n \ge n_0$. We then have that

$$\|w^n \circ \lambda^n - w\|_t \le \epsilon + 2\|h\|_M \int_0^t \|w^n \circ \lambda^n - w\|_s ds$$

for $0 \le t \le T$ and $n \ge n_0$. Therefore, by Gronwall's inequality,

$$||w^n \circ \lambda^n - w||_T \le \epsilon e^{2T\|h\|_M}.$$

Since also $\|||\dot{\lambda}^n - 1||_T \to 0$ as $n \to \infty$, it follows that

$$\|w^n \circ \lambda^n - w\|_T \vee \|\lambda^n - e\|_T \to 0,$$

as $n \to \infty$. $\square$

Proof of Lemma 2: From Lemma 1 in [25], for parts (i) and (ii) it is sufficient to verify the Lipschitz continuity of the function $\eta : D([0,\infty),\Re) \to D([0,\infty),\Re)$, defined as

$$\eta(w)(t) \equiv \int_0^{\phi_C(w)(t)} h(u)du$$

for $w \in D([0,\infty),\Re)$. Since

$$\|\eta(w_1) - \eta(w_2)\|_T \le \sup_{0\le t\le T} \int_{\phi_C(w_2)(t)}^{\phi_C(w_1)(t)} |h(u)|du \le \|h\|_C \|\phi_C(w_1) - \phi_C(w_2)\|_T,$$

and Theorem 14.8.1 in Whitt [36] establishes the mapping ϕ_C is Lipschitz continuous with Lipschitz constant 2, we conclude

$$\|\eta(w_1) - \eta(w_2)\|_T \le 2\|h\|_C \|w_1 - w_2\|_T.$$

Note that if the condition $0 \le x_1(0), x_2(0) \le C$ is not satisfied, then the above inequality is not valid, and must also accommodate the jump at time 0.

The proof of part (iii) proceeds in a similar manner to the proof of part (iii) of Lemma 1 and therefore has not been included. It is, however, necessary to note that using the explicit form of the two-sided regulator mapping in Kruk et al [17],

$$\phi_C(x)(t) = \phi(x)(t) - \sup_{0\le s\le t}\left([\phi(x)(s) - C]^+ \wedge \inf_{s\le u\le t} \phi(x)(u)\right),$$

it is straightforward to show that

$$\phi_C(x)(\lambda(t)) = \phi_C\left(x \circ \lambda\right)(t).$$

$\square$

Proof of Lemma 3: Given any $T > 0$, suppose we can show

$$\lim_{n\to\infty} E\left[\sup_{0\le t\le T} \overline{R}^n(t)\right] = 0. \tag{7.18}$$

Convergence in L_1 implies convergence in probability, and so

$$\sup_{0\le t\le T} \overline{R}^n(t) \to 0$$

in probability, as $n \to \infty$. Convergence in probability implies weak convergence, and so

$$\sup_{0\le t\le T} \overline{R}^n \Rightarrow 0,$$

as $n \to \infty$, which establishes the desired result.

To establish (7.18), we must show that for any $\delta > 0$ and all large enough n,

$$E\left[\sup_{0\le t\le T} \overline{R}^n(t)\right] < \delta. \tag{7.19}$$

First observe, using the linearity of the expectation operator and the definitions of $\tilde{V}^n$ in (3.12) and $\overline{R}^n$ in (5.7), that

$$E\left[\sup_{0\le t\le T} \overline{R}^n(t)\right] = n^{-1}\sum_{j=1}^{\lfloor nT\rfloor} P\left(\tilde{V}^n(t_j^{n,-}) \ge \sqrt{n}a_j^n\right). \tag{7.20}$$

Next, we claim there exists a K such that for all n large enough,

$$P\left(\max_{j=1,\dots,\lfloor nT\rfloor} \tilde{V}^n\left(t_j^{n,-}\right) \ge K\right) < \frac{\delta}{2T}. \tag{7.21}$$

To see (7.21), construct a second single server queue on the same probability space as the original queue with abandonments, and with the same arrival and service time sequence as the queue with abandonments, but from which no abandonments occur; i.e., $a_i = \infty$ for all $i \in \{1,2,\dots\}$. On a sample path basis, the offered waiting time process in the queue without abandonments always exceeds or is equal to the equivalent process in the queue with abandonments. Weak convergence of a process χ^n in $D([0,\infty),\Re)$ implies tightness of the sequence of random variables $\sup_{0\le t\le T}|\chi^n(t)|$. Therefore, it follows from the weak convergence of the waiting time process for a GI/GI/1 queue established in Theorem 1 in Section 3.2 of Reiman [27] that there exists n_0 such that

$$P\left(\sup_{0\le t\le T}\left|\tilde{V}^n(t)\right| \ge K\right) < \frac{\delta}{2T},\ n \ge n_0 \tag{7.22}$$

holds. Since for all $j \in \{1,\dots,\lfloor nT\rfloor\}$, by the definition of t_j^n, the strong law of large numbers, and because assumption (3.3) implies $\rho^n \to 1$ as $n \to \infty$,

$$t_j^n \le t_{\lfloor nT\rfloor}^n = \frac{\lfloor nT\rfloor}{n\rho^n}\frac{1}{\lfloor nT\rfloor}\sum_{j=1}^{\lfloor nT\rfloor} u_j \to T,$$

as $n \to \infty$, and so (7.21) holds for large enough n. Finally, recalling the definition of F^n in (3.6), because $F^n\left(n^{-1/2}K\right) \to 0$ as $n \to \infty$, we can choose n large enough so that

$$F^n\left(\frac{K}{\sqrt{n}}\right) < \frac{\delta}{2T}. \tag{7.23}$$

Therefore, for any $j \in \{1,\dots,\lfloor nT\rfloor\}$, from (7.21) and (7.23),

$$\begin{aligned} P\left(\tilde{V}^n\left(t_j^{n,-}\right) \ge \sqrt{n}a_j^n\right) &< \frac{\delta}{2T} + P\left(\tilde{V}^n\left(t_j^{n,-}\right) \ge \sqrt{n}a_j^n \cap \max_{j=1,\dots,\lfloor nT\rfloor} \tilde{V}^n\left(t_j^{n,-}\right) < K\right) \\ &\le \frac{\delta}{2T} + F^n\left(\frac{K}{\sqrt{n}}\right) < \frac{\delta}{T}, \end{aligned}$$

and so from (7.20), for large enough n,

$$E\left[\sup_{0\le t\le T}\overline{R}^n(t)\right] < n^{-1}\sum_{j=1}^{\lfloor nT\rfloor}\frac{\delta}{T}\le \delta,$$

which establishes (7.19). $\square$

Proof of Lemma 4: For any given $t, \epsilon, \delta > 0$, we must show

$$P\left(\sup_{0\le s\le t}\left|\tilde{M}_a^n(s)\right| > \epsilon\right) = P\left(\max_{i=1,\ldots,\lfloor nt\rfloor}|M_a^n(i)| > \epsilon\sqrt{n}\right) < \delta, \tag{7.24}$$

for large enough n. By a generalization of Kolmogorov's inequality (see, for example, Corollary 2.1 in Hall and Heyde [10]), and the orthogonality of martingale differences (see, for example, property (vii) on page 355 of Resnick [28]),

$$\begin{aligned} P\left(\max_{i=1,\ldots,\lfloor nt\rfloor}|M_a^n(i)| > \epsilon\sqrt{n}\right) &\le \frac{E\left|M_a^n\left(\lfloor nt\rfloor\right)\right|^2}{\epsilon^2 n} \qquad (7.25)\\ &= \frac{1}{\epsilon^2 n}E\left[\sum_{j=1}^{\lfloor nt\rfloor}\left(\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\} - E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}|\mathcal{F}_{j-1}\right]\right)^2\right]. \end{aligned}$$

Since

$$\begin{aligned} \mathbf{1}^2\{V^n\left(t_j^{n,-}\right)\ge a_j^n\} &= \mathbf{1}\{V^n\left(t_j^{n,-}\right)\ge a_j^n\}\\ \mathbf{1}\left\{V^n\left(t_j^{n,-}\right)\ge a_j^n\right\}E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right)\ge a_j^n\right\}|\mathcal{F}_{j-1}\right] &\ge 0\\ E^2\left[\mathbf{1}\{V^n\left(t_j^{n,-}\right)\ge a_j^n\}\right] &\le E\left[\mathbf{1}^2\{V^n\left(t_j^{n,-}\right)\ge a_j^n\}\right], \end{aligned}$$

it follows that

$$\begin{aligned} &\left(\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\} - E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}|\mathcal{F}_{j-1}\right]\right)^2 \qquad (7.26)\\ &\quad\le \mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\} + E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}|\mathcal{F}_{j-1}\right]. \end{aligned}$$

Furthermore,

$$E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\} + E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}|\mathcal{F}_{j-1}\right]\right] = 2E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}\right], \tag{7.27}$$

and so from (7.25), (7.26), (7.27), and the definition of $\overline{R}^n$ in (5.7),

$$P\left(\max_{i=1,\ldots,\lfloor nt\rfloor}|M_a^n(i)| > \epsilon\sqrt{n}\right) \le \frac{2}{\epsilon^2 n}\sum_{j=1}^{\lfloor nt\rfloor}E\left[\mathbf{1}\left\{V^n(t_j^{n,-})\ge a_j^n\right\}\right] = \frac{2}{\epsilon^2}E\left[\overline{R}^n(t)\right]\to 0,$$

as $n\to\infty$, by the convergence established in (7.18). $\square$

Proof of Lemma 5: We first argue that the families $\{\tilde{X}^n\}$ and $\{\tilde{\epsilon}^n\}$ are tight in $D([0,\infty),\Re)$, and then use those tightness results to establish the tightness of $\{\tilde{V}^n\}$.

Tightness of $\{\tilde{\mathbf{X}}^{\mathbf{n}}\}$:
We first argue

$$\tilde{S}_a^n \circ \overline{A}^n \Rightarrow 0, \tag{7.28}$$

as $n \to \infty$. Recall the distributional equivalence in (5.9)

$$\tilde{S}^n \circ \overline{R}^n \circ \overline{A}^n \stackrel{D}{=} \tilde{S}_a^n \circ \overline{A}^n.$$

Hence, it is sufficient to show

$$\tilde{S}^n \circ \overline{R}^n \circ \overline{A}^n \Rightarrow 0,$$

as $n \to \infty$. From the weak convergence in (3.19) and Lemma 3,

$$\left(\tilde{S}^n, \overline{R}^n\right) \Rightarrow (\mathrm{var}(v_1) W_{S,2}, 0)\,,$$

as $n \to \infty$. The joint convergence holds because $\overline{R}^n$ weakly converges to a constant. Since $\overline{R}^n$ and $\overline{A}^n$ are non-decreasing in t for each n, the almost sure convergence of $\overline{A}^n$ in (3.18), and the random time change theorem imply

$$\tilde{S}^n \circ \overline{R}^n \circ \overline{A}^n \Rightarrow 0,$$

as $n \to \infty$.

From the definition of $\tilde{X}^n$ in (5.1) and the evolution equation for X^n in (4.2),

$$\tilde{X}^n = \tilde{A}^n + \tilde{S}^n(\bar{A}^n) + \sqrt{n}t(\rho^n - 1) - \tilde{S}_a^n(\bar{A}^n(t)) - \tilde{M}_a^n(\bar{A}^n).$$

The weak convergences in (3.19), the heavy traffic assumption (3.3), the almost sure convergence of $\overline{A}^n$ in (3.18), Lemma 4, the weak convergence in (7.28), and the random time change theorem imply that

$$\tilde{X}^n \Rightarrow W,$$

as $n \to \infty$, where W is a Brownian motion with drift θ and variance $\sigma^2 = \mathrm{var}(u_1) + \mathrm{var}(u_2)$. Tightness follows because weakly convergence subsequences are relatively compact.

Tightness of $\{\tilde{\epsilon}^{\mathbf{n}}\}$:
Let $T > 0$. We verify the conditions (16.17) and (16.18) of Theorem 16.8 in Billingsley [3] to prove the tightness of $\{\tilde{\epsilon}^n\}$. In particular, we must show the following.

- **(B16.17)** For every $\eta > 0$, there exists an a and an n_0 such that

$$P\left(\sup_{0 \le t \le T} |\tilde{\epsilon}^n(t)| \ge a\right) < \eta,\ n \ge n_0;$$

- **(B16.18)** For every γ and η, there exists a δ and an n_0 such that

$$P\left(w'_T\left(\tilde{\epsilon}^n, \delta\right) \ge \gamma\right) < \eta,\ n \ge n_0.$$

To see conditions (B16.17) and (B16.18) can be satisfied, first recall from (5.14) that

$$\tilde{\epsilon}^n(t) = \int_0^t \left(\int_0^{\tilde{V}^n(s^-)} h(w)dw \right) ds - \int_0^t \sqrt{n} \left(1 - \exp\left(-\frac{1}{\sqrt{n}} \int_0^{\tilde{V}^n(s^-)} h(w)dw \right) \right) d\overline{A}^n(s). \tag{7.29}$$

Choose K, a, and n_0 large enough, and δ small enough so that

$$P\left(\sup_{0 \le t \le T} \left| \tilde{V}^n(t) \right| \ge K \right) < \frac{\eta}{2} \tag{7.30}$$

$$P\left(K\|h\|_K T + (K\|h\|_K + 1) \overline{A}^n(T) \ge a \right) < \frac{\eta}{2} \tag{7.31}$$

$$P\left(K\|h\|_K 2\delta + (1 + K\|h\|_K) \left(\overline{A}^n(t+2\delta) - \overline{A}^n(t) \right) \ge \gamma \right) < \frac{\eta}{2}, \tag{7.32}$$

for all $n \ge n_0$, where $\|h\|_K < \infty$ because h is continuous. Such a K, a, δ, and n_0 exist from the observation (7.22) in the proof of Lemma 3, and the almost sure convergence of $\overline{A}^n$ to the identity function in (3.18). Then, from (7.29), (7.30), and (7.31), also noting that

$$\sqrt{n} \left(1 - \exp\left(-x/\sqrt{n} \right) \right) \to x, \tag{7.33}$$

uniformly on compact sets of $[0, \infty)$,

$$\begin{aligned} P\left(\sup_{0 \le t \le T} |\tilde{\epsilon}^n(t)| \ge a \right) &\le \frac{\eta}{2} + P\left(\sup_{0 \le t \le T} |\tilde{\epsilon}^n(t)| \ge a \cap \sup_{0 \le t \le T} \left| \tilde{V}^n(t) \right| < K \right) \\ &\le \frac{\eta}{2} + P\left(TK\|h\|_K + (K\|h\|_K + 1) \overline{A}^n(T) \ge a \right) < \eta, \end{aligned}$$

and so (B16.17) holds. Next, from the definitions of w and w_T in (7.1) and (7.2), the inequality (7.3), and the non-negativity of h,

$$\begin{aligned} w'_T(\tilde{\epsilon}^n, \delta) &\le w_T(\tilde{\epsilon}^n, 2\delta) \\ &\le \sup_{0 \le t \le T-2\delta} \int_t^{t+2\delta} \left(\int_0^{\tilde{V}^n(s^-)} h(w)dw \right) ds + \int_t^{t+2\delta} \sqrt{n} \left(1 - \exp\left(-\frac{1}{\sqrt{n}} \int_0^{\tilde{V}^n(s^-)} h(w)dw \right) \right) d\overline{A}^n(s), \end{aligned}$$

and so, also using (7.30), (7.32), and (7.33),

$$\begin{aligned} P\left(w'_T(\tilde{\epsilon}^n, \delta) \ge \gamma \right) &\le \frac{\eta}{2} + P\left(w'_T(\tilde{\epsilon}^n, \delta) \ge \gamma \cap \sup_{0 \le t \le T} \left| \tilde{V}^n(t) \right| \le K \right) \\ &\le \frac{\eta}{2} + P\left(2\delta K\|h\|_K + (1 + K\|h\|_K) \left(\overline{A}^n(t+2\delta) - \overline{A}^n(t) \right) \ge \gamma \right) \\ &< \eta, \end{aligned}$$

which implies (B16.18) holds. We conclude $\{\tilde{\epsilon}^n\}$ is tight.

Tightness of $\{\tilde{\mathbf{V}}^{\mathbf{n}}\}$:
We show the sequence $\{\tilde{V}^n\}$ satisfies the definition of relative compactness, and so is tight in

$D([0,\infty),\Re)$. Consider any subsequence $\{\tilde{V}^{n_i}\}$. Because the families $\{\tilde{X}^n\}$ and $\{\tilde{\epsilon}^n\}$ are both tight, there exists a further subsequence $\{\tilde{X}^{n_i(m)} + \tilde{\epsilon}^{n_i(m)}\}$ such that

$$\tilde{X}^{n_i(m)} + \tilde{\epsilon}^{n_i(m)} \Rightarrow \chi,$$

as $n_i(m) \to \infty$, for some limit process χ. From the representation for $\tilde{V}^n$ in (5.4), the continuous mapping theorem, and the continuity of the mapping ϕ^h established in part (iii) of Proposition 3, on this further subsequence,

$$\tilde{V}^{n_i(m)} = \phi^h\left(\tilde{X}^{n_i(m)} + \tilde{\epsilon}^{n_i(m)}\right) \Rightarrow \phi^h(\chi),$$

as $n_i(m) \to \infty$. We conclude $\{\tilde{V}^n\}$ is relatively compact. □

Proof of Lemma 6: The offered waiting time process can only increase at arrival time points and so

$$V^n(t) \le \max_{i=1,\ldots,A^n(t)} V^n(t_i). \tag{7.34}$$

Since in the n^{th} system service times are scaled by n^{-1} and the service time of the ith arrival is only included in the offered waiting time process if $V^n(t_i) \le C^n$, recalling the definition for C^n in (3.7), we find

$$\max_{i=1,\ldots,A^n(t)} V^n(t_i) \le \frac{C}{\sqrt{n}} + \max_{i=1,\ldots,A^n(t)} \frac{v_i}{n}. \tag{7.35}$$

From (7.35), the fact that the v_i's are non-negative random variables, the definition of δ^n in (5.19), and (7.34)

$$\sup_{0\le s\le t} \sqrt{n}\delta^n(s) = \max_{i=1,\ldots,A^n(t)} \left[\tilde{V}^n(t_i) - C\right]^+ \le \max_{i=1,\ldots,A^n(t)} \frac{v_i}{\sqrt{n}}. \tag{7.36}$$

Since $\sqrt{n}\delta^n$ is a non-negative process, $n^{-1}A^n \to e$ as $n \to \infty$, almost surely, uniformly on compact sets, and Lemma 3.3 in Iglehart and Whitt [12] establishes $\max_{i=1,\ldots,nt} n^{-1/2}v_i \Rightarrow 0$ as $n \to \infty$, the random time change theorem and (7.36) imply

$$\sqrt{n}\delta^n \Rightarrow 0,$$

as $n \to \infty$. □

It is useful for the proof of Lemma 7 to observe that

$$(\mathcal{V}^n, I^n, \mathcal{U}^n) = (\phi_{C^n}, \psi_{1,C^n}, \psi_{2,C^n})\left(\chi^n - \delta^n\right), \tag{7.37}$$

where

$$\begin{aligned} \chi^n(t) &= \frac{1}{n}A^n(t) - \rho^n t + S^n\left(A^n(t)\right) + t\left(\rho^n - 1\right) \\ &\quad - S_a^n(A^n(t)) - \frac{1}{n}\sum_{j=1}^{A^n(t)} \mathbf{1}\left\{V^n(t_j^{n,-}) \ge a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right\}, \end{aligned} \tag{7.38}$$

and

$$\mathcal{U}^n(t) = \frac{1}{n}\sum_{j=1}^{A^n(t)} \mathbf{1}\left\{\tilde{V}^n\left(t_j^{n,-}\right) \geq C\right\}. \tag{7.39}$$

To see (7.37) is valid, we verify the conditions (C1) and (C2) of Definition 5 when h is the zero function.

(C1) From (5.17) and the fact that the process V^n is non-negative, $0 \leq \mathcal{V}^n \leq C^n$. We now show

$$\mathcal{V}^n = \chi^n - \delta^n + I^n - \mathcal{U}^n.$$

From (4.4)-(4.6),

$$\begin{aligned} V^n(t) &= X^n(t) + \epsilon_B^n(t) - \int_0^t \left(\int_0^{V^n(s^-)\wedge C^n} h^n(u)du\right) ds + I^n(t) - U^n(t) \\ &= X^n(t) - \frac{1}{n}\int_0^t F^n\left(V^n(s^-)\right) dA^n(s) + I^n(t). \end{aligned}$$

The definitions of X^n in (4.2) and M_a in (2.2) then imply

$$\begin{aligned} V^n(t) &= \frac{1}{n}A^n(t) - \rho^n t + S^n\left(A^n(t)\right) - S_a^n\left(A^n(t)\right) \\ &\quad + t\left(\rho^n - 1\right) - \frac{1}{n}\sum_{j=1}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n\right\} + I^n(t). \end{aligned}$$

Since the abandonment times $\{a_j^n, j = 0, 1, 2, \ldots\}$ are bounded above by $\sqrt{n}C$,

$$\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n\right\} = \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right\} + \mathbf{1}\left\{\tilde{V}^n\left(t_j^{n,-}\right) \geq C\right\},$$

and so

$$V^n(t) = \chi^n(t) + I^n(t) - \mathcal{U}^n(t).$$

From (5.18), $\mathcal{V}^n = V^n - \delta^n$, and so

$$\mathcal{V}^n(t) = \chi^n(t) - \delta^n(t) + I^n(t) - \mathcal{U}^n(t).$$

(C2) The processes I^n and $\mathcal{U}^n$ are non-decreasing functions having $I^n(0) = \mathcal{U}^n(0) = 0$, and

$$\int_0^\infty \mathcal{V}^n(t)dI^n(t) = \int_0^\infty \left(V^n(t) \wedge C^n\right)\mathbf{1}\left\{V^n(t) = 0\right\} = 0,$$

and

$$\int_0^\infty \left[C^n - \mathcal{V}^n(t)\right]^+ d\mathcal{U}^n(t) = \frac{1}{n}\int_0^\infty \left[C^n - \left(V^n(t)\wedge C^n\right)\right]^+ d\left(\sum_{j=1}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq C^n\right\}\right) = 0.$$

Proof of Lemma 7: Because $C^n = n^{-1/2}C$ upper bounds the abandonment times $\{a_j^n, j = 0, 1, 2, \ldots\}$,

$$\overline{R}^n(t) = n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\} + n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\{\tilde{V}^n\left(t_j^{n,-}\right) \geq C\} \tag{7.40}$$

Using the expression for F^n in (3.8),

$$\begin{aligned} & E\left[\sup_{0\leq t\leq T} n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\}\right] \\ = \; & n^{-1}\sum_{j=1}^{\lfloor nT \rfloor} P\left(V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right) \\ \leq \; & n^{-1}\lfloor nT \rfloor\left(1 - \exp\left(\frac{-1}{\sqrt{n}}\int_0^C h(w)dw\right)\right) \\ \rightarrow \; & 0, \end{aligned}$$

as $n \rightarrow \infty$. Convergence in L_1 implies convergence in probability, and so

$$\sup_{0\leq t\leq T} n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right\} \rightarrow 0, \tag{7.41}$$

in probability, as $n \rightarrow \infty$.

For the second term in (7.40), we first observe from (3.11), (3.13), and (3.14), and (7.38) that

$$\begin{aligned} \sqrt{n}\chi^n(t) &= \tilde{A}^n(t) + \tilde{S}^n\left(\overline{A}^n(t)\right) + \sqrt{n}t\left(\rho^n - 1\right) \\ &\quad -\tilde{S}_a^n\left(\overline{A}^n(t)\right) - \frac{1}{\sqrt{n}}\sum_{j=1}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right\}. \end{aligned}$$

Let W be a Brownian motion with drift θ and variance $\sigma^2 = \text{var}(u_1) + \text{var}(v_1)$. The almost sure convergence in (3.18), the weak convergence in (3.19), the random time change theorem, and the heavy traffic assumption in (3.3) establish the weak convergence of the first three terms in the above expression for $\sqrt{n}\chi^n$ to W. The arguments to show the weak convergence of the fourth term to 0 in (7.28) in the proof of Lemma 5 remain valid. To see the fifth term weakly converges to zero, suppose we can show

$$n^{-1/2}\sum_{j=1}^{A^n(\cdot)} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} \Rightarrow \int_0^{\cdot} h(u)du, \tag{7.42}$$

as $n \rightarrow \infty$, which implies the process on the left-hand side is tight in $D([0,\infty), \Re)$. Then, because for each $s \leq t$,

$$\sum_{j=A^n(s)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap \tilde{V}^n\left(t_j^{n,-}\right) < C\right\} \leq \sum_{j=A^n(s)}^{A^n(t)} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\},$$

it follows from Theorem 16.8 in Billingsley (which provides sufficient conditions for tightness in $D([0,\infty),\Re)$ that

$$\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{A^n(t)}\mathbf{1}\left\{V^n\left(t_j^{n,-}\right)\geq a_j^n\cap\tilde{V}^n\left(t_j^{n,-}\right)<C\right\}\right\}$$

is tight in $D([0,\infty),\Re)$. Consider any subsequence n_k on which

$$\frac{1}{\sqrt{n_k}}\sum_{j=0}^{A^{n_k}(\cdot)}\mathbf{1}\left\{V^n\left(t_j^{n_k,-}\right)\geq a_j^{n_k}\cap\tilde{V}^{n_k}\left(t_j^{n_k,-}\right)<C\right\}\Rightarrow\Psi,$$

as $n_k\to\infty$. On this subsequence, the two-sided conventional regulator mapping representation in (7.37), the scaling property in part (ii) of Proposition 4, Lemma 6, the continuous mapping theorem, and the continuity of $\psi_{2,C}$ established in Theorem 14.8.1 in Whitt [36] imply

$$\sqrt{n_k}\mathcal{U}^{n_k}=\psi_{2,C}\left(\sqrt{n_k}\chi^{n_k}-\delta^{n_k}\right)\Rightarrow\psi_{2,C}\left(W+\Psi\right),$$

as $n_k\to\infty$, and so

$$\mathcal{U}^{n_k}\Rightarrow 0,$$

as $n_k\to\infty$. Since the subsequence n_k was arbitrary,

$$\mathcal{U}^n\Rightarrow 0,$$

as $n\to\infty$. Finally, from (7.40), (7.41), the random time change theorem, and the fact that convergence in probability implies weak convergence,

$$\overline{R}^n\Rightarrow 0,$$

as $n\to\infty$.

Weak Convergence of $n^{-1/2}\sum_{j=1}^{A^n(\cdot)}\mathbf{1}\left\{a_j^n<C/\sqrt{n}\right\}$:

Consider the centered sum

$$\begin{aligned}\triangle^n(t) &\equiv \frac{1}{n}\sum_{j=1}^{\lfloor nt\rfloor}\left(\sqrt{n}\mathbf{1}\left\{a_j^n<C/\sqrt{n}\right\}-E\left[\sqrt{n}\mathbf{1}\{a_j^n<C/\sqrt{n}\}\right]\right)\\ &= \left(\frac{1}{n}\sum_{j=1}^{\lfloor nt\rfloor}\sqrt{n}\mathbf{1}\left\{a_j^n<C/\sqrt{n}\right\}\right)-\frac{\lfloor nt\rfloor}{n}\sqrt{n}P(a_1^n<C/\sqrt{n}).\end{aligned}$$

From the representation for F^n in (3.8) and L'Hopital's rule, we have that

$$\sqrt{n}P(a_1^n<C/\sqrt{n})=\sqrt{n}\left(1-\exp\left(-\frac{1}{\sqrt{n}}\int_0^C h(u)du\right)\right)\to\int_0^C h(u)du, \tag{7.43}$$

as $n \to \infty$, and so

$$\lim_{n\to\infty} \frac{\lfloor nt \rfloor}{n} \sqrt{n} P\left(a_1^n \leq C/\sqrt{n}\right) = t \int_0^C h(u)du.$$

Therefore, if we can show that $\triangle^n(t) \to 0$ in probability as $n \to \infty$, then it must be true that for each $t \geq 0$,

$$\frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \sqrt{n} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} \to t \int_0^C h(u)du, \tag{7.44}$$

in probability, as $n \to \infty$. For any $t > 0$, in order to show that $\triangle^n(t) \to 0$ in probability as $n \to \infty$, it is sufficient to show that $E\left[\triangle^n(1)^2\right] \to 0$. (This is because convergence in L^2 implies convergence in probability.) By independence,

$$\begin{aligned} E\left[\triangle^n(t)^2\right] &= \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} E\left[\left(\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} - E\left[\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\}\right]\right)^2\right] \\ &\leq \frac{2}{n} \sum_{j=1}^{\lfloor nt \rfloor} E\left[\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\}\right] \\ &= 2\frac{\lfloor nt \rfloor}{n} P\left(a_j^n < C/\sqrt{n}\right) \\ &\to 0, \end{aligned}$$

as $n \to \infty$ by (7.43), and so (7.44) is valid.

The limit point on the right-hand side of (7.44) is deterministic, and convergence in probability implies weak convergence. Repeated application of Theorem 3.9 in Billingsley [3] then implies that the finite dimensional distributions weakly converge. If we can argue the process on the left-hand side of (7.44) is tight, then we can conclude the process level convergence

$$\left\{ n^{-1/2} \sum_{j=1}^{\lfloor n\cdot \rfloor} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} \right\} \Rightarrow \int_0^{\cdot} h(u)du$$

as $n \to \infty$ is valid. The random time change theorem and (3.18) then imply that the weak convergence in (7.42) is valid, completing the proof.

We verify conditions (16.17) and (16.18) of Theorem 16.8 in Billingsley [3] to show that the process

$$\left\{ n^{-1/2} \sum_{j=1}^{A^n(\cdot)} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} \right\}$$

is tight in $D([0,\infty),\Re)$.

(B16.17) For each $a > T\int_0^C h(u)du$, from (7.44)

$$P\left(\sup_{0\leq t\leq T} \left| \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \sqrt{n} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} \right| > a\right) = P\left(\frac{1}{n} \sum_{j=1}^{\lfloor nT \rfloor} \sqrt{n} \mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} > a\right) \to 0,$$

as $n \to \infty$.

(B16.18) First note that

$$
\begin{aligned}
& w_T^{'}\left(\frac{1}{n}\sum_{j=1}^{\lfloor n\cdot\rfloor}\sqrt{n}\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\}, \delta\right) \\
\leq\ & \max_{i\in\{0,\ldots,\lfloor T/\delta\rfloor+1\}}\left(\frac{1}{n}\sum_{j=0}^{\lfloor n\delta(i+1)\rfloor}\sqrt{n}\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\} - \frac{1}{n}\sum_{j=0}^{\lfloor n\delta i\rfloor}\sqrt{n}\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\}\right) \\
\Rightarrow\ & \delta\int_0^C h(u)du \quad \text{as} \quad n \to \infty.
\end{aligned}
$$

The above weak convergence follows by the convergence of the finite dimensional distributions and the continuous mapping theorem. Thus, since convergence in distribution to a constant implies convergence in probability as well, we have that for each $\varepsilon > 0$,

$$
\lim_{\delta\to 0}\lim_{n\to\infty} P\left(w_T^{'}\left(\frac{1}{n}\sum_{j=1}^{\lfloor n\cdot\rfloor}\sqrt{n}\mathbf{1}\left\{a_j^n < C/\sqrt{n}\right\}, \delta\right) > \varepsilon\right) = 0,
$$

as $n \to \infty$.

□

Proof of Lemma 8: Since $\overline{R}^n \Rightarrow 0$ as $n \to \infty$ under Assumption 2 by Lemma 7, the arguments to prove Lemma 4 remain valid. □

Proof of Lemma 9: By the representation of $\tilde{\mathcal{V}}^n$ in (5.22) and the continuous mapping theorem, it is sufficient to show the sequences $\left\{\tilde{X}^n\right\}, \left\{\tilde{\delta}^n\right\}$, and $\{\tilde{\epsilon}_B^n\}$ are tight in $D([0,\infty),\Re)$. The sequence $\{\tilde{X}^n\}$ is tight in $D([0,\infty),\Re)$ by the same arguments as in the proof of Lemma 5. Lemma 6 establishes the sequence $\{\tilde{\delta}^n\}$ is tight in $D([0,\infty),\Re)$. The following argument shows the sequence $\{\tilde{\epsilon}_B^n\}$ is tight in $D([0,\infty),\Re)$. The evolution equation for $\tilde{\epsilon}_B^n$ in (5.29) is exactly that for $\tilde{\epsilon}^n$ in (5.14), with $\tilde{\mathcal{V}}^n$ replacing $\tilde{V}^n$. Therefore, almost the same arguments as in Lemma 5 can be used to show $\tilde{\epsilon}_B^n$ satisfies conditions (16.17) and (16.18) in Theorem 16.8 in Billingsley [3] (given in (B16.17) and (B16.18) in the proof of Lemma 5 in terms of $\tilde{\epsilon}^n$), which establishes $\{\tilde{\epsilon}_B^n\}$ is tight in $D([0,\infty),\Re)$. The difference is a simplification: the choice of large K in (7.30) is not necessary because from the definitions of C^n in (3.7), $\mathcal{V}^n$ in (5.17), and $\tilde{\mathcal{V}}^n$ in (5.22),

$$
\tilde{\mathcal{V}}^n(t) \leq C \text{ for all } t \geq 0.
$$

□

Proof of Lemma 10: We first prove Lemma 10 under Assumption 1 and then under Assumption 2.

Proof under Assumption 1: First observe from the definitions of M_a and $\tilde{M}_a^n$ in (2.2) and

(3.16) that

$$
\begin{aligned}
&n^{-1/2} \sum_{i=A^n\circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \geq a_i^n\right\} \\
&= \tilde{M}_a^n\left(\overline{A}^n(t)\right) - \tilde{M}_a^n\left(\overline{A}^n(a^n(t))\right) + n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} F^n\left(V^n(t_i^{n,-})\right).
\end{aligned}
$$

Lemma 4, the convergence of $\overline{A}^n$ to the identity process in (3.18), the fact that $a^n(t) \leq t$ for all $t \geq 0$, and the random time change theorem show

$$
\tilde{M}_a^n \circ \overline{A}^n - \tilde{M}_a^n \circ \overline{A}^n \circ a^n \Rightarrow 0, \tag{7.45}
$$

as $n \to \infty$. Therefore, to show the stated result under Assumption 1, it remains to show that

$$
n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} F^n\left(V^n(t_i^{n,-})\right) \Rightarrow 0, \tag{7.46}
$$

as $n \to \infty$.

To show (7.46), it is sufficient to show that for any $\gamma, \delta, T > 0$,

$$
P\left(\sup_{0\leq t\leq T} n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} F^n\left(V^n(t_i^{n,-})\right) > \gamma\right) < \delta. \tag{7.47}
$$

For any $\delta > 0$, from (7.21) and the convergence of $\overline{A}^n$ in (3.18), we can choose K large enough so that

$$
P\left(\max_{j=1,\ldots,\lfloor n\overline{A}^n(t)\rfloor} \tilde{V}^n\left(t_j^{n,-}\right) \geq K\right) < \frac{\delta}{2},
$$

and so

$$
\begin{aligned}
&P\left(\sup_{0\leq t\leq T} n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} F^n\left(V^n(t_i^{n,-})\right) > \gamma\right) \qquad (7.48)\\
&< \frac{\delta}{2} + P\left(\sup_{0\leq t\leq T} n^{-1/2} \sum_{i=A^n(a^n(t))}^{A^n(t)} F^n\left(V^n(t_i^{n,-})\right) > \gamma \cap \max_{j=1,\ldots,\lfloor n\overline{A}^n(t)\rfloor} \tilde{V}^n(t_j^{n,-}) < K\right) \\
&\leq \frac{\delta}{2} + P\left(\sup_{0\leq t\leq T} \left(A^n(t) - A^n(a^n(t))\right) n^{-1/2} F^n\left(\frac{K}{\sqrt{n}}\right) > \gamma\right) \\
&\leq \frac{\delta}{2} + P\left(\sup_{0\leq t\leq T} \left(\tilde{A}^n(t) - \tilde{A}^n(a^n(t))\right) F^n\left(\frac{K}{\sqrt{n}}\right) + \rho^n(t - a^n(t))\sqrt{n}F^n\left(\frac{K}{\sqrt{n}}\right) > \gamma\right).
\end{aligned}
$$

From the definition of F^n in (3.6) and L'Hopital's rule,

$$
\sqrt{n}F^n\left(\frac{K}{\sqrt{n}}\right) = \sqrt{n}\left(1 - \exp\left(-\frac{1}{\sqrt{n}}\int_0^K h(w)dw\right)\right) \to \int_0^K h(w)dw,
$$

as $n \to \infty$. The function h is continuous, and so $\sup_{0 \le w \le K} |h(w)| < \infty$. Furthermore, $\tilde{A}^n$ converges to a continuous limit process. Therefore, the weak convergence of a^n to the identity process e in (6.2) implies

$$\left(\tilde{A}^n - \tilde{A}^n \circ a^n\right) F^n \left(\frac{K}{\sqrt{n}}\right) + \rho^n \left(e - a^n\right) \sqrt{n} F^n \left(\frac{K}{\sqrt{n}}\right) \Rightarrow 0,$$

as $n \to \infty$. Since weak convergence to a constant is equivalent to convergence in probability, we can choose n large enough so that

$$P\left(\sup_{0 \le t \le T} \left(\tilde{A}^n(t) - \tilde{A}^n(a^n(t))\right) F^n \left(\frac{K}{\sqrt{n}}\right) + \rho^n (t - a^n(t)) \sqrt{n} F^n \left(\frac{K}{\sqrt{n}}\right) > \gamma\right) < \frac{\delta}{2},$$

which, from (7.48) implies (7.47) is valid, and completes the proof under Assumption 1.

Proof under Assumption 2: Define

$$\begin{aligned}
U_A^n(i) &\equiv \frac{1}{n} \sum_{j=1}^{i} \mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \ge C^n\right\} \\
\tilde{U}_A^n(t) &\equiv \sqrt{n} U_A^n(\lfloor nt \rfloor) \\
M^n(i) &\equiv \sum_{j=1}^{i} \left(\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \ge a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\} - E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \ge a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\} | \mathcal{F}_{j-1}\right]\right) \\
\tilde{M}^n(t) &\equiv \frac{1}{\sqrt{n}} M^n\left(\lfloor nt \rfloor\right),
\end{aligned}$$

and observe that

$$\begin{aligned}
& n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \ge a_i^n\right\} \qquad (7.49) \\
= \; & n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \ge a_i^n \cap V^n(t_i^{n,-}) \ge C^n\right\} \\
& + n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \ge a_i^n \cap V^n(t_i^{n,-}) < C^n\right\} \\
= \; & n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n(t_i^{n,-}) \ge C^n\right\} + n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \ge a_i^n \cap V^n(t_i^{n,-}) < C^n\right\} \\
= \; & \tilde{M}^n \circ \overline{A}^n(t) - \tilde{M}^n \circ \overline{A}^n \circ a^n(t) + \tilde{U}_A^n \circ \overline{A}^n(t) - \tilde{U}_A^n \circ \overline{A}^n \circ a^n(t) \\
& + n^{-1/2} \sum_{i=A^n \circ a^n(t)}^{A^n(t)} E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \ge a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\} | \mathcal{F}_{j-1}\right].
\end{aligned}$$

We show each of the three terms on the right-hand side of (7.49) weakly converges to 0 as $n \to \infty$.

Arguments identical to those in the proof of Lemma 4 show that for any $t > 0$

$$P\left(\max_{i=1,\ldots,\lfloor nt\rfloor} |M^n(i)| > \epsilon\sqrt{n}\right) \leq \frac{2}{\epsilon^2 n}\sum_{j=1}^{\lfloor nt\rfloor} E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\}|\mathcal{F}_{j-1}\right]. \tag{7.50}$$

Since from the expression for F^n in (3.8), recalling that $C^n = n^{-1/2}C$ from (3.7),

$$E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\}|\mathcal{F}_{j-1}\right] \leq 1 - \exp\left(-\frac{1}{\sqrt{n}}\int_0^C h(w)dw\right) \to 0,$$

as $n \to \infty$, (7.50) implies

$$P\left(\max_{i=1,\ldots,\lfloor nt\rfloor}|M^n(i)| > \epsilon\sqrt{n}\right) \to 0,$$

as $n \to \infty$, and so,

$$\tilde{M}^n \Rightarrow 0,$$

as $n \to \infty$. The almost sure convergence of $\overline{A}^n$ in (3.18) and the weak convergence of a^n in (6.2) then imply

$$\tilde{M}^n \circ \overline{A}^n - \tilde{M}^n \circ \overline{A}^n \circ a^n \Rightarrow 0, \tag{7.51}$$

as $n \to \infty$. Next, because $\tilde{U}_A^n\left(\overline{A}^n(t)\right) = (b^n)^{-1}\tilde{U}^n(t)$ (recalling the definitions of U^n and $\tilde{U}^n$ in (4.4) and (5.24)), $b^n \to 1$ as $n \to \infty$, and $\overline{A}^n \to e$ almost surely, uniformly on compact sets, as $n \to \infty$, part (ii) of Theorem 1 and the random time change theorem establish

$$\tilde{U}_A^n \Rightarrow \psi_{2,C}^h(W),$$

as $n \to \infty$, where W is a Brownian motion as defined in Theorem 1. Thus, since $\psi_{2,C}^h(W)$ is almost surely a continuos process, this then implies by (6.2) and the fact that $\overline{A}^n \to e$ as $n \to \infty$, that

$$\tilde{U}_A^n \circ \overline{A}^n - \tilde{U}_A^n \circ \overline{A}^n \circ a^n \Rightarrow 0, \tag{7.52}$$

as $n \to \infty$. Finally, for the last term on the right-hand side of (7.49), again using (3.8),

$$\begin{aligned} & n^{-1/2}\sum_{i=A^n\circ a^n(t)}^{A^n(t)} E\left[\mathbf{1}\left\{V^n\left(t_j^{n,-}\right) \geq a_j^n \cap V^n\left(t_j^{n,-}\right) < C^n\right\}|\mathcal{F}_{j-1}\right] \qquad (7.53) \\ \leq\ & \left(\overline{A}^n(t) - \overline{A}^n \circ a^n(t)\right)\sqrt{n}\left(1 - \exp\left(\frac{-1}{\sqrt{n}}\int_0^C h(w)dw\right)\right) \\ \to\ & 0, \end{aligned}$$

as $n \to \infty$, because $a^n \Rightarrow e$ as $n \to \infty$ from (6.2) and

$$\sqrt{n}\left(1 - \exp\left(\frac{-1}{\sqrt{n}}\int_0^C h(w)dw\right)\right) \to \int_0^C h(w)dw < \infty,$$

as $n \to \infty$. We conclude from (7.49), (7.51), (7.52), and (7.53) that

$$n^{-1/2} \sum_{i=A^n \circ a^n(\cdot)}^{A^n(\cdot)} \mathbf{1}\left\{V^n\left(t_i^{n,-}\right) \geq a_i^n\right\} \Rightarrow 0,$$

as $n \to \infty$, also using (6.2) and the fact that $\overline{A}^n \to e$ as $n \to \infty$, almost surely, uniformly on compact sets. □

Proof of Lemma 11: We first argue that for $0 \leq x \leq C$, $P_x\left(T_0^C < \infty\right) = 1$. Define

$$\tilde{T}_0^C \equiv \inf\{t \geq 0 : \phi_C(W)(t) = 0\},$$

and observe that

$$P_x\left(\tilde{T}_0^C < \infty\right) = 1 \tag{7.54}$$

by Problem 7 in Chapter 5 in Harrison [11]. From (4.20), for any $x \in D([0,\infty), \Re)$,

$$x(t) = \mathcal{M}_C^h(x)(t) + \int_0^t \left(\int_0^{\phi_C\left(\mathcal{M}^h(x)\right)(s)} h(u)du\right) ds$$

is written as the sum of $\mathcal{M}_C^h(x)$ and a non-decreasing function. Then, Theorem 1.6 in [17] establishes

$$\phi_C\left(\mathcal{M}_C^h(x)\right) \leq \phi_C(x)$$

for any $x \in D\left([0,\infty), \Re\right)$. We conclude $T_0^C \leq \tilde{T}_0^C$ on every sample path, which from (7.54) implies $P_x\left(T_0^C < \infty\right) = 1$.

To see $P_x\left(T_0 < \infty\right) = 1$ for all $x \geq 0$, first observe that for

$$A \equiv \frac{\sigma^2}{2}\frac{d^2}{dx^2} + (\theta - H(x))\frac{d}{dx},$$

the function

$$u_n^\epsilon(x) = 1 - \frac{\int_\epsilon^x \exp\left(\frac{2}{\sigma^2}\int_0^y (H(z)-\theta)\,dz\right)dy}{\int_\epsilon^n \exp\left(\frac{2}{\sigma^2}\int_0^y (H(z)-\theta)\,dz\right)dy} \tag{7.55}$$

solves the ordinary differential equation

$$\left(Au_n^\epsilon\right)(x) = 0,\ \epsilon \leq x \leq n, \tag{7.56}$$

with boundary conditions

$$u_n^\epsilon(\epsilon) = 1,\ u_n^\epsilon(n) = 0. \tag{7.57}$$

Next consider the diffusion

$$Z(t) = -\int_0^t H\left(Z(s)\right)ds + W(t),$$

and observe that if

$$\begin{aligned} T_\epsilon &\equiv \inf\left\{t \geq 0 : \phi^h(W)(t) \leq \epsilon\right\} \\ T_n &\equiv \inf\left\{t \geq 0 : \phi^h(W)(t) \geq n\right\} \\ T_\epsilon^z &\equiv \inf\left\{t \geq 0 : Z(t) \leq \epsilon\right\} \\ T_n^z &\equiv \inf\left\{t \geq 0 : Z(t) \geq n\right\}, \end{aligned}$$

then for $\epsilon < x < n$,

$$P_x\left(T_\epsilon < T_n\right) = P_x\left(T_\epsilon^z < T_n^z\right), \tag{7.58}$$

because for $0 \leq t < T_\epsilon^z \wedge T_n^z$, $Z(t) \in (\epsilon, n)$, and so $\phi^h(W)(t) = Z(t)$. Let $\tilde{u}_n^\epsilon$ be a bounded, twice continuously differentiable function such that

$$\tilde{u}_n^\epsilon(x) = u_n^\epsilon(x) \text{ for all } \epsilon \leq x \leq n,$$

and having bounded first derivative. Ito's formula establishes

$$d\tilde{u}_n^\epsilon\left(Z(t)\right)dt = \left(A\tilde{u}_n^\epsilon\right)\left(Z(t)\right)dt + \tilde{u}_n^{\epsilon'}\left(Z(t)\right)\sigma dW(t). \tag{7.59}$$

Our assumption that $\tilde{u}_n^{\epsilon'}$ is bounded is a sufficient condition for the stochastic integral in (7.59) to be a martingale. Applying the optional stopping theorem to the bounded stopping time $t \wedge T_\epsilon^z \wedge T_n^z$ and using the fact that $\left(A\tilde{u}_n^\epsilon\right)\left(Z(t)\right) = 0$ for $0 \leq t < T_\epsilon^z \wedge T_n^z$ from (7.56), we find

$$\tilde{u}_n^\epsilon(x) = E_x\left[\tilde{u}_n^\epsilon\left(Z\left(t \wedge T_\epsilon^z \wedge T_n^z\right)\right)\right]. \tag{7.60}$$

The exit time of a one-dimensional diffusion from a compact subinterval of $\Re$ is finite with probability 1. (See Section 5 in [13].) Therefore, taking the limit as $t \to \infty$ in (7.60) and using the bounded convergence theorem shows

$$\tilde{u}_n^\epsilon(x) = E_x\left[\tilde{u}_n^\epsilon\left(Z\left(T_\epsilon^z \wedge T_n^z\right)\right)\right].$$

The boundary conditions in (7.57) and the equality in (7.58) then imply

$$\tilde{u}_n^\epsilon(x) = P_x\left(T_\epsilon^z < T_n^z\right) = P_x\left(T_\epsilon < T_n\right). \tag{7.61}$$

Since the diffusion $\phi^h(W)$ has continuous paths almost surely,

$$\lim_{n\to\infty}\lim_{\epsilon\downarrow 0} P_x\left(T_\epsilon < T_n\right) = P_x\left(T_0 < \infty\right). \tag{7.62}$$

Furthermore, for any $x \geq 0$, from the expression for u_n^ϵ in (7.55),

$$\lim_{n\to\infty}\lim_{\epsilon\downarrow 0}\tilde{u}_n^\epsilon(x) = 1, \tag{7.63}$$

because

$$\int_0^n \exp\left(\frac{2}{\sigma^2}\int_0^y \left(H(z) - \theta\right)dz\right)dy \to \infty,$$

as $n \to \infty$, under the assumption that $\lim_{z\to\infty} H(z) > \theta$. We conclude from (7.61), (7.62), and (7.63) that

$$1 = P_x\left(T_0 < \infty\right).$$

□

References

[1] B. Ata, J. Harrison, and L. Shepp. Drift rate control of a Brownian processing system. *Annals of Applied Probability*, 15(2):1145–1160, 2005.

[2] F. Baccelli, P. Boyer, and G. Hebuterne. Single-server queues with impatient customers. *Advances in Applied Probability*, 16:887–905, 1984.

[3] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., New York, 1999. Second Edition.

[4] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, 100:36–50, 2005.

[5] H. Chen and D. D. Yao. *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization.* Springer-Verlag, New York, 2001.

[6] J. G. Dai and W. Dai. A heavy traffic limit theorem for a class of open queueing networks with finite buffers. *Queueing Systems*, 32:5–40, 1999.

[7] B. Doytchinov, J. Lehoczky, and S. Shreve. Real-time queues in heavy traffic with earliest-deadline-first queue discipline. *Annals of Applied Probability*, 11:332–378, 2001.

[8] P. Echeverria. A criterion for invariant measures of Markov processes. *Z. Wahrsch. Verw. Gebiete*, 61:1–16, 1982.

[9] O. Garnett, A. Mandelbaum, and M. I. Reiman. Designing a call center with impatient customers. *Manufacturing and Service Operations Management*, 4:208–227, 2002.

[10] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Applications.* Academic Press, New York, 1980.

[11] J. M. Harrison. *Brownian Motion and Stochastic Flow Systems.* John Wiley & Sons, New York, 1985.

[12] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic, i and ii. *Advances in Applied Probability*, 2:150–177 and 355–364, 1970.

[13] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, New York, 1991. Second Edition.

[14] J. F. C. Kingman. The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.*, 57:902–904, 1961.

[15] J. F. C. Kingman. Two similar queues in parallel. *Ann. Math. Statist.*, 32:1314–1323, 1961.

[16] J. F. C. Kingman. On queues in heavy traffic. *J. Roy. Statis. Soc. Ser. B*, 24:383–392, 1962.

[17] L. Kruk, J. Lehoczky, K. Ramanan, and S. Shreve. An explicit formula for doubly reflected processes in [0,a], 2005. Submitted.

[18] L. Kruk, J. Lehoczky, and S. Shreve. Accuracy of state space collapse for earliest-deadline-first queues. *Annals of Applied Probability*, 16(2):516–581, 2006.

[19] L. Kruk, J. Lehoczky, S. Shreve, and S. Yeung. Multiple-input heavy-traffic real-time queues. *Annals of Applied Probability*, 13(1):54–99, 2003.

[20] L. Kruk, J. Lehoczky, S. Shreve, and S. Yeung. Earliest-deadline-first service in heavy traffic acyclic networks. *Annals of Applied Probability*, 14(3):1306–1352, 2004.

[21] J. Lehoczky. Real-time queueing theory. In *Proceedings of the IEEE Real-Time Systems Symposium*, pages 186–195. IEEE, 1997.

[22] R. Lillo and M. Martin. Stability in queues with impatient customers. *Stochastic Models*, 17, 2001.

[23] A. Mandelbaum and N. Shimkin. A model for rational abandonment from invisible queues. *Queueing Systems*, 36:141–173, 2000.

[24] C. Palm. Etude des delais d'attente. *Ericsson Technics*, 5:37–56, 1937.

[25] J. E. Reed and A. R. Ward. A diffusion approximation for a generalized Jackson network with reneging. *Proceedings of the 42nd Allerton Conference on Communication, Control, and Computing*, 2004.

[26] M. I. Reiman. The heavy traffic diffusion approximation for sojourn times in Jackson networks. In R. L. Disney and T. J. Ott, editors, *Applied Probability-Computer Science, The Interface*. Birkhauser, Boston, 1982.

[27] M. I. Reiman. Some diffusion approximations with state space collapse. In F. Baccelli and G. Fayolle, editors, *Modelling and Performance Evaluation Methodology*, pages 209–240. Springer-Verlag, 1984.

[28] S. I. Resnick. *A Probability Path*. Birkhauser, Boston, 1999.

[29] A. V. Skorokhod. Stochastic equations for diffusions in a bounded region 1,2. *Theor. of Prob. and Its Appl.*, 6:264–274, 1961.

[30] R. E. Stanford. Reneging phenomena in single channel queues. *Mathematics of Operations Research*, 4:162–178, 1979.

[31] A. Ward and S. Kumar. Asymptotically optimal admission control of a queue with impatient customers. *Mathematics of Operations Research, submitted.*, 2006.

[32] A. R. Ward and N. Bambos. On stability of queueing networks with job deadlines. *Journal of Applied Probability*, 40:293–304, 2003.

[33] A. R. Ward and P. W. Glynn. A diffusion approximation for a Markovian queue with reneging. *Queueing Systems*, 43:103–128, 2003.

[34] A. R. Ward and P. W. Glynn. Properties of the reflected Ornstein-Uhlenbeck process. *Queueing Systems*, 44:109–123, 2003.

[35] A. R. Ward and P. W. Glynn. A diffusion approximation for a GI/GI/1 queue with balking or reneging. *Queueing Systems*, 50:371–400, 2005.

[36] W. Whitt. *Stochastic-Process Limits*. Springer, New York, 2002.

[37] W. Whitt. Heavy-traffic limits for loss proportions in single-server queues. *Queueing Systems*, 46:507–536, 2004.

[38] W. Whitt. Engineering solution of a basic call center model. *Management Science*, 51(2):221–235, 2005.

[39] S. Zeltyn and A. Mandelbaum. Call centers with impatient customers: Many-server asymptotics of the M/M/n+G queue. *Queueing Systems*, 51(3/4):361–402, 2005.

[40] E. Zohar, A. Mandelbaum, and N. Shimkin. Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science*, 48:566–583, 2002.