

Convergence of a Queueing System in Heavy Traffic with General Patience-time Distributions

Chihoon Lee
Department of Statistics
Colorado State University
Fort Collins, CO 80523, USA

Ananda Weerasinghe *
Department of Mathematics
Iowa State University
Ames, IA 50011, USA

March 29, 2011

Abstract

We analyze a sequence of single-server queueing systems with impatient customers in heavy traffic. Our state process is the offered waiting time and the customer arrival process has a state dependent intensity. Service times and customer patient-times are independent, i.i.d. with general distributions subject to mild constraints. We establish the heavy traffic approximation for the scaled offered waiting time process and obtain a diffusion process as the heavy traffic limit. The drift coefficient of this limiting diffusion is influenced by the sequence of patience-time distributions in a non-linear fashion. We also establish an asymptotic relationship between the scaled version of offered waiting time and queue-length. As a consequence, we obtain the heavy traffic limit of the scaled queue-length. We introduce an infinite-horizon discounted cost functional whose running cost depends on the offered waiting time and server idle time processes. Under mild assumptions, we show that the expected value of this cost functional for the n -th system converges to that of the limiting diffusion process as n tends to infinity.

Keywords: Stochastic control, Controlled queueing systems, Heavy traffic theory, Diffusion approximations, Customer abandonment, Customer impatience, Reneging.

AMS Subject Classifications: primary 60K25, 68M20, 90B22; secondary 90B18.

1 Introduction

In this article, we study a heavy traffic approximation result for a sequence of single-server queueing systems with impatient customers. Customers are served under First-Come-First-Serve (FCFS) service discipline. In the n -th system, where $n = 1, 2, 3, \dots$, the arrival process has a dynamic intensity which depends on the offered waiting time and this intensity is of order $\mathcal{O}(n)$ for large n , and the service-times are i.i.d. with a general distribution where the mean service-time is of order $\mathcal{O}(1/n)$ for large n . The customers abandon the system if the service is not initiated within their patience-time. In the n -th system, customers act independently, and their patience-times are i.i.d. distributed and this distribution may depend on n .

In many real world examples, such as telephone call centers or internet traffic, customers may not observe the actual queue-length but often approximate waiting time is available to them. In our model, offered waiting time (or the workload process) is the basic state process and the arrival

*Research supported by U. S. Army Research Office grant W911NF0710424.

intensity of the customers is dependent on it. To motivate this work, consider a processing facility where each customer or job arrives with a deadline. Upon the arrival of each customer, a system manager learns about the customer deadline as well as the required service time. Hence, the information on offered waiting time is available to the manager and accordingly, the manager can influence the arrival intensity by means of admission control.

In practice, customer abandonment is a well documented significant feature of the queueing systems. In the queueing models, Palm [23] initiated the importance of incorporating this feature. In the telephone call center setting with many-server systems, such models are considered in [13, 19, 10, 11, 36, 29, 22, 25]. For single server setting, Ward and co-authors addressed several performance evaluation issues of such systems in [30, 31, 26]. For general queueing systems in heavy traffic (with or without customer abandonment), there are numerous articles that address the issue of system optimization and [3, 5, 14, 15] is a partial list of such articles.

The results established in this article are closely related to the works of [26] and [30, 31], but they differ in three main aspects: First, in the n -th system, the intensity of our arrival process is non-constant and may depend on the current value of the offered waiting time. Loosely speaking, system manager may exercise adjustments of order $\mathcal{O}(\sqrt{n})$ to the admission rate of the n -th system without disturbing the delicate balance in heavy traffic conditions. But such adjustments have an influence on the drift coefficient of the limiting diffusion process as described in our Theorem 4.10. In controlled queueing systems, such adjustments are known as “thin control” and we refer to [1, 15] for such problems.

Second, our assumptions on patience-time distributions are quite general. In Markovian abandonment regimes [30] and also in [31] (for many-server queues in Halfin-Whitt heavy traffic regime see [4, 11, 10, 21, 13, 22, 25]) where the same patience-time distribution is used in the modeling, only the behavior of patience-time distribution in a neighborhood of origin effects the dynamics of the limiting diffusion. But, in an interesting article [26], Reed and Ward consider the patient time distribution of the n -th system to have a hazard rate intensity dependent on n (see [25] for a many-server Halfin-Whitt heavy traffic case). They provide statistical data in support of their choice. The dynamics of their limiting diffusion process depends on the entire patience-time distribution function. Our results incorporate both of these scenarios in the same general framework as illustrated in the examples of Section 3, and our assumptions can be satisfied by many other classes of patience-time distributions. The heavy traffic limit for the diffusion-scaled waiting time process is established in Theorem 4.10 and it describes the effect of patience-time distributions on the limiting diffusion. One key ingredient in our proof of Theorem 4.10 is the martingale functional central limit theorem, and this approach helps us to accommodate these general assumptions. This is in contrast with the proofs in [26]. The diffusion-scaled offered waiting time process turned out to be the reflected process under a generalized Skorokhod map introduced in the subsection 4.3. The martingale central limit theorem helps us to establish the weak convergence of the input process related to this generalized Skorokhod map, where the output is the above described reflected process. Then, the continuity properties of the generalized Skorokhod map yield the weak convergence of the diffusion-scaled offered waiting times and also identify the diffusion limit.

Third, we use martingale moment inequalities to obtain moment bounds for the input process. Then again we employ the martingale central limit theorem and Theorem 4.10 to establish the convergence of the expected value of an infinite-horizon discounted cost functional of the n -th system to that of the limiting diffusion process as n tends to infinity. Such convergence results for the expected value of the cost functionals are important in deriving asymptotically optimal strategies for the system optimization problems in heavy traffic regimes. We refer to [32, 15] (and [5, 4, 22] in many-server Halfin-Whitt heavy traffic regime) for such results related to controlled

queueing systems. We intend to use the results obtained here to address such a controlled system optimization problem in a future article.

This article is organized as follows. In Section 2 we introduce the basic model and the key martingale relevant to the arrival process. Such a martingale formulation is used in [35] for the heavy traffic analysis of queue-length processes, when the arrival and service rates are dependent on queue-length. In Section 3, we speed up the arrival rates to be of order $\mathcal{O}(n)$ and to balance this and to obtain heavy traffic conditions, we make the average service time in the n -th system to be $\frac{1}{n}$. We carefully lay out our assumptions on arrival intensities, service times and patience-time distributions. Section 4 addresses the weak convergence of scaled offered waiting time processes in heavy traffic. We establish the fluid limit first and then use it to obtain the diffusion limit for the scaled offered waiting time process. Main result in this section is Theorem 4.10, and we use martingale functional central limit theorem to obtain this weak convergence result. In Section 5, we establish the asymptotic relationship between the scaled queue length and scaled offered waiting time processes. Here we follow the proof of a similar result in [26], but supplement it with necessary additional estimates to accommodate our general assumptions. We prove the convergence of an infinite horizon discounted cost functional of the n -th system to that of the limiting diffusion under heavy traffic in Section 6. In this cost functional, the running cost function depends on offered waiting time, and there is also a cost related to server idle time. Since the running cost function is unbounded and is of polynomial growth, we need a few additional assumptions there. To reach our conclusion, we establish necessary moment estimates and combine them with the weak convergence result in Theorem 4.10. For controlled queueing networks, such convergence results are obtained in [32, 15] and in the case of many-server systems, we refer to [5]. In the appendix in Section 7, we provide a detailed construction of the arrival process with arrival intensity dependent on the offered waiting time.

The following notation is used. The set of positive integers is denoted by \mathbb{N} , the set of real numbers by \mathbb{R} and nonnegative real numbers by \mathbb{R}_+ . Let \mathbb{R}^d be the d -dimensional Euclidean space. For $a, b \in \mathbb{R}$, let $a \wedge b \doteq \min\{a, b\}$ and $a^+ = \max\{a, 0\}$, $a^- = -\min\{a, 0\}$. We use $[a]$ to denote the integer part of $a \in \mathbb{R}$. If $(M(t))_{t \geq 0}$ is a martingale then we denote the associated quadratic variation of M on the interval $[0, T]$ by $[M](T)$. The convergence in distribution of random variables (with values in some Polish space) Φ_n to Φ will be denoted as $\Phi_n \Rightarrow \Phi$. When $\sup_{0 \leq s \leq t} |f_n(s) - f(s)| \rightarrow 0$ as $n \rightarrow \infty$, for all $t \geq 0$, we say that $f_n \rightarrow f$ uniformly on compact sets. For a real valued function f defined on some metric space X and $T \in \mathbb{R}_+$, define $\|f\|_T = \sup_{x \in [0, T]} |f(x)|$.

Finally, let $D[0, \infty)$ denote the class of right continuous functions having left limit defined from $[0, \infty)$ to \mathbb{R} , equipped with the usual Skorokhod topology.

2 Basic Model

First we describe the queueing model with FCFS service discipline and customer abandonment on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $A(t)$ be the number of customers arrived at the station by time t . The random variable t_j represents the arrival time of the j -th customer, and we assume $\mathbb{E}(t_j) < \infty$. Service time of the j -th customer is represented by the random variable v_j . We assume that the customers are impatient and the j -th customer will leave the system after waiting a random time d_j if the service does not begin by then. The sequences (v_j) and (d_j) are assumed to be i.i.d. and independent of each other, $\mathbb{E}(v_1) = 1$ and $\text{var}(v_1) = \sigma_s^2 < \infty$. We let F be the cumulative distribution function of d_1 .

The amount of time an incoming customer at time t has to wait for service depends upon the

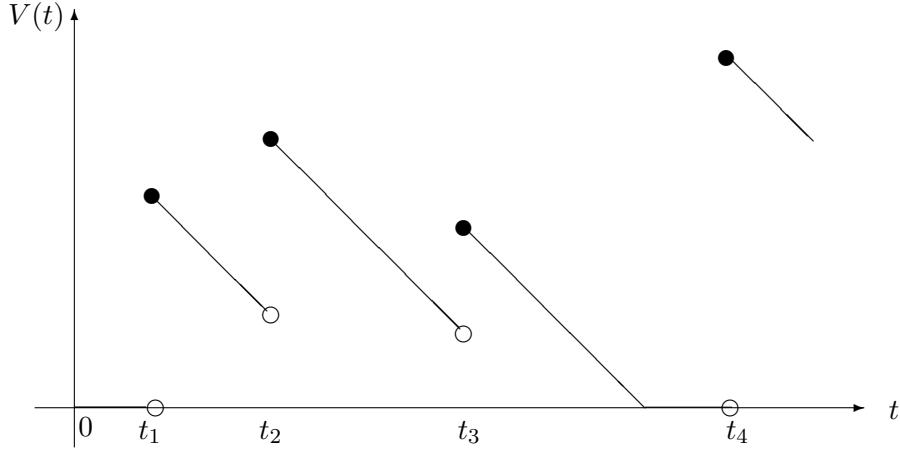


Figure 1: A typical sample path of $V(t)$.

service times of the non-abandoning customers, who are already waiting in the queue. Similar to [26], we define the offered waiting time process

$$V(t) \equiv \sum_{j=1}^{A(t)} v_j \mathbf{1}_{[V(t_{j-}) < d_j]} - \int_0^t \mathbf{1}_{[V(s) > 0]}(s) ds. \quad (2.1)$$

The process $\{V(t) : t \geq 0\}$ is non-negative, has sample paths which are right continuous with left limits (RCLL), and also at each arrival epoch t_j , it has an upward jump of size v_j . On the time interval $[t_j, t_{j+1})$, $V(t)$ is continuous, non-increasing and satisfies $V(t) = \max\{0, V(t_j) - (t - t_j)\}$. Figure 1 shows a typical sample path of the process $\{V(t)\}_{t \geq 0}$.

The quantity $V(t)$ can be interpreted as the time needed to empty the system from time t onwards if there are no arrivals after time t , and hence it is also known as the workload at time t . We note that once $V(t_n)$ is known then $V(t)$ is well defined on the next interval $[t_n, t_{n+1})$ (see below (2.11) for more details).

Next, we define the σ -fields $(\widehat{\mathcal{F}}_n)_{n \geq 0}$. Let $\widehat{\mathcal{F}}_0 \equiv \sigma(t_1)$, and for $n \geq 1$ let

$$\widehat{\mathcal{F}}_n \equiv \sigma((t_1, v_1, d_1), \dots, (t_n, v_n, d_n), t_{n+1}) \subseteq \mathcal{F}. \quad (2.2)$$

Notice that $V(t_n-)$ is $\widehat{\mathcal{F}}_{n-1}$ -measurable and the abandonment time d_n of the n -th customer is independent of $\widehat{\mathcal{F}}_{n-1}$. Hence,

$$\mathbb{P}[V(t_n-) \geq d_n | \widehat{\mathcal{F}}_{n-1}] = F(V(t_n-)) \quad (2.3)$$

holds almost surely, where F is the distribution function of d_n . We introduce two martingales $(M^v(n))$ and $(M^d(n))$ with respect to the filtration $(\widehat{\mathcal{F}}_n)_{n \geq 1}$ introduced in (2.2). We let

$$M^v(n) \equiv \sum_{j=1}^n (v_j - 1) \mathbf{1}_{[V(t_{j-}) < d_j]} \quad (2.4)$$

$$M^d(n) \equiv \sum_{j=1}^n \left(\mathbf{1}_{[V(t_{j-}) \geq d_j]} - \mathbb{E}[\mathbf{1}_{[V(t_{j-}) \geq d_j]} | \widehat{\mathcal{F}}_{j-1}] \right) \quad (2.5)$$

for all $n \in \mathcal{N}$. Clearly, $M^d(n)$ is an $\widehat{\mathcal{F}}_n$ -martingale (see also [26]). Here we show that $M^v(n)$ also is an $\widehat{\mathcal{F}}_n$ -martingale. Since $V(t_{n+1}-)$ and d_{n+1} are measurable with respect to $\sigma(\widehat{\mathcal{F}}_n, d_{n+1})$ and v_{n+1} is independent of $\sigma(\widehat{\mathcal{F}}_n, d_{n+1})$, it follows that

$$\mathbb{E} \left[(v_{n+1} - 1) \mathbf{1}_{[V(t_{n+1}-) < d_{n+1}]} | \sigma(\widehat{\mathcal{F}}_n, d_{n+1}) \right] = \mathbf{1}_{[V(t_{n+1}-) < d_{n+1}]} \mathbb{E}(v_{n+1} - 1) = 0.$$

Now conditioning both sides of (2.4) with respect to $\widehat{\mathcal{F}}_n$, we can see that $M^v(n)$ is an $\widehat{\mathcal{F}}_n$ -martingale as well. Using (2.3) in (2.5), we also see that for all $n \in \mathcal{N}$

$$M^d(n) = \sum_{j=1}^n \left[\mathbf{1}_{[V(t_j-) \geq d_j]} - F(V(t_j-)) \right]. \quad (2.6)$$

Using (2.1), (2.3)–(2.6) and after simple algebraic manipulations, we obtain the following system equation:

$$V(t) + \int_0^t F(V(s-)) dA(s) = (A(t) - t) + M^v(A(t)) - M^d(A(t)) + I(t), \quad (2.7)$$

for all $t \geq 0$, where

$$I(t) \equiv \int_0^t \mathbf{1}_{[V(s)=0]}(s) ds, \quad (2.8)$$

and $I(t)$ represents the idle time at the station during time interval $[0, t]$.

Next, we describe a filtration $(\widetilde{\mathcal{G}}_t)_{t \geq 0}$ which represents the information gathered over time by the system manager. We begin with a discrete filtration $(\overset{\circ}{\mathcal{F}}_n)_{n \geq 0}$ given by $\overset{\circ}{\mathcal{F}}_0 \equiv \{\emptyset, \Omega\}$ and

$$\overset{\circ}{\mathcal{F}}_n = \sigma((t_1, v_1, d_1), \dots, (t_n, v_n, d_n)) \quad \text{for } n \geq 1. \quad (2.9)$$

It is easy to verify that for each $t \geq 0$, $A(t)$ is a stopping time with respect to the filtration $(\overset{\circ}{\mathcal{F}}_n)_{n \geq 0}$, where $A(\cdot)$ is the arrival process with arrival times (t_j) and $\overset{\circ}{\mathcal{F}}_n \subseteq \widehat{\mathcal{F}}_n$ for all $n \geq 0$ and the filtration $(\widehat{\mathcal{F}}_n)_{n \geq 0}$ is given in (2.2). Next, we introduce the filtration $(\widetilde{\mathcal{G}}_t)_{t \geq 0}$ by

$$\widetilde{\mathcal{G}}_t \equiv \overset{\circ}{\mathcal{F}}_{A(t)} \quad \text{for all } t \geq 0. \quad (2.10)$$

Let $\lambda(\cdot)$ be a given Borel measurable function defined on $[0, \infty)$ which satisfies the condition $0 < \epsilon < \lambda(x) < C$ for all $x \geq 0$. Here ϵ and C are positive constants. In our analysis, we assume that

$$\left\{ A(t) - \int_0^t \lambda(V(s)) ds : t \geq 0 \right\} \quad (2.11)$$

is a martingale with respect to the filtration $(\widetilde{\mathcal{G}}_t)_{t \geq 0}$. We introduce yet another filtration $(\mathcal{G}_t)_{t \geq 0}$ where

$$\mathcal{G}_t \equiv \sigma(A(s), V(s) : s \leq t). \quad (2.12)$$

Notice that, once the value of $V(t_n)$ is known, the process $V(t)$ can be obtained on the interval $[t_n, t_{n+1})$ as explained earlier and hence for all $t_n \leq t < t_{n+1}$, the quantity $\int_0^t \lambda(V(s)) ds$ is also known by the time t_n . Moreover, when $t_n \leq t < t_{n+1}$, $V(t)$ is a functional of the random variables $(t_1, v_1, d_1), \dots, (t_n, v_n, d_n)$. Consequently, $\mathcal{G}_t \subseteq \widetilde{\mathcal{G}}_t$ for all $t \geq 0$ and the process $A(t) - \int_0^t \lambda(V(s)) ds$

is a martingale with respect to $(\mathcal{G}_t)_{t \geq 0}$ as well. In our proofs, we commonly use this martingale property with respect to $(\mathcal{G}_t)_{t \geq 0}$, while the martingale property with respect to $(\tilde{\mathcal{G}}_t)_{t \geq 0}$ filtration will be used only in the proof of Lemma 4.11.

We indicate the construction of such an arrival process $A(\cdot)$ and several of its properties in the Appendix. We note that since $A(\cdot)$ is a point process with (\mathcal{G}_t) -intensity $\lambda(\tilde{V}(t))$, we can use the random change of time method (see Theorem T16 and Lemma L17 in Section 6 of Chapter 2, [7]) to obtain the convenient representation

$$A(t) = Y \left(\int_0^t \lambda(V(s)) ds \right), \quad (2.13)$$

where $Y(\cdot)$ is a unit-rate Poisson process. This representation helps us in several estimates.

3 Heavy Traffic Regime

We consider a sequence of queueing systems indexed by $n \in \mathbb{N}$. In our analysis, basic state process of the n -th system will be the offered waiting time process $V_n(\cdot)$. The arrival rate $n\lambda_n(V_n(\cdot))$ of the n -th system is state dependent and the j -th customer arrival occurs at time t_j^n . The cumulative number of customer arrivals in $[0, t]$ in the system is given by $A_n(t)$. When n becomes large, arrival rate of the n -th system becomes large and thus to obtain heavy traffic conditions, we need to make the service time of the n -th system small as described below.

For the j -th arrival in the n -th system, service time is $v_j^n \equiv v_j/n$, and the abandonment time is denoted by d_j^n . As described in [26], the basic equation of the offered waiting time process $\{V_n(t) : t \geq 0\}$ is given by

$$V_n(t) = \frac{1}{n} \sum_{j=1}^{A_n(t)} v_j \mathbf{1}_{[V_n(t_j^n -) < d_j^n]} - \int_0^t \mathbf{1}_{[V_n(s) > 0]}(s) ds, \quad (3.1)$$

where $A_n(\cdot)$ is the arrival process. We introduce the filtration $\{\mathcal{G}_t^n : t \geq 0\}$ of the n -th system by $\mathcal{G}_t^n \equiv \sigma(A_n(s), V_n(s) : s \leq t)$. We also introduce the filtration $(\tilde{\mathcal{G}}_t^n)$ as similar to (2.10) and this filtration represents the information available to the system manager over time. Next, we define the discrete time filtration $(\hat{\mathcal{F}}_i^n)_{i \geq 0}$ by $\hat{\mathcal{F}}_0^n \equiv \sigma(t_1^n)$ and

$$\hat{\mathcal{F}}_i^n \equiv \sigma((t_1^n, v_1^n, d_1^n), \dots, (t_i^n, v_i^n, d_i^n), t_{i+1}^n) \quad (3.2)$$

for $i \geq 1$. Next, we define the associated continuous time filtration $(\mathcal{F}_t^n)_{t \geq 0}$ by

$$\mathcal{F}_t^n \equiv \hat{\mathcal{F}}_{[nt]}^n \equiv \sigma((t_1^n, v_1^n, d_1^n), \dots, (t_{[nt]}^n, v_{[nt]}^n, d_{[nt]}^n), t_{[nt]+1}^n). \quad (3.3)$$

Now we describe our basic assumptions:

Assumption 3.1.

- (i) The sequences $(v_j^n)_{j \geq 1}$ and $(d_j^n)_{j \geq 1}$ are independent, non-negative, i.i.d. random variables with $v_j^n \equiv \frac{v_j}{n}$ for all $j \geq 1$, $\mathbb{E}(v_j) = 1$ and $\mathbb{E}(v_j - 1)^2 = \sigma_s^2 > 0$. Furthermore, for each $j \geq 1$, the random variables v_j^n and d_j^n are independent of $\hat{\mathcal{F}}_{j-1}^n$.

(ii) The arrival process $A_n(\cdot)$ of the n -th system has an associated intensity process $n\lambda_n(V_n(\cdot))$; that is,

$$\left\{ A_n(t) - n \int_0^t \lambda_n(V_n(s)) ds : t \geq 0 \right\} \quad (3.4)$$

is a $(\tilde{\mathcal{G}}_t^n)$ -martingale. Since this process is adapted to (\mathcal{G}_t^n) and $\mathcal{G}_t^n \subseteq \tilde{\mathcal{G}}_t^n$, it is a (\mathcal{G}_t^n) -martingale as well.

Assumption 3.2.

(i) The function $\lambda_n(\cdot)$ is Borel measurable on $[0, \infty)$ and there exist two positive constants $\epsilon_0, C_0 > 0$ (independent of n and x) such that $0 < \epsilon_0 < \lambda_n(x) < C_0$ for all $x \geq 0$ and $n \geq 1$.

(ii) For each $K > 0$, $\lim_{n \rightarrow \infty} \sup_{x \in [0, K]} |\lambda_n(x) - 1| = 0$.

(iii) There exist small $\delta_0 > 0$ and $M > 0$ such that $\sup_{n \geq 1} \sup_{x \in [0, \delta_0]} \sqrt{n}(\lambda_n(x) - 1)^+ \leq M < \infty$.

(iv) There exists a non-negative, locally Lipschitz continuous function $u(\cdot)$ defined on $[0, \infty)$ such that

$$\lim_{n \rightarrow \infty} \sup_{x \in [0, K]} \left| \sqrt{n} \left(1 - \lambda_n \left(\frac{x}{\sqrt{n}} \right) \right) - u(x) \right| = 0,$$

for each $K > 0$.

Assumption 3.3. Let $F_n(\cdot)$ be the right continuous abandonment distribution function of the i.i.d. sequence $(d_j^n)_{j \geq 1}$. Assume that $F_n(0) = 0$ and there exists a non-negative, locally Lipschitz continuous function $H(\cdot)$ such that

$$\lim_{n \rightarrow \infty} \sup_{x \in [0, K]} \left| \sqrt{n} F_n \left(\frac{x}{\sqrt{n}} \right) - H(x) \right| = 0,$$

for each $K > 0$. As a consequence, we have $H(0) = 0$ and $\lim_{n \rightarrow \infty} F_n(x/\sqrt{n}) = 0$ for each $x \geq 0$.

Remark 3.4. We provide concrete examples that satisfy the above set of assumptions.

1. An example of arrival rate function $\lambda_n(\cdot)$: Let $u(\cdot)$ be non-negative, locally Lipschitz continuous and

$$\lambda_n(x) = 1 - \frac{u(\sqrt{n}x)}{\sqrt{n}} + \frac{\theta_n(x)}{\sqrt{n}},$$

where $\theta_n(\cdot)$ is a bounded function such that $\lim_{n \rightarrow \infty} \|\theta_n\|_K = 0$ for each $K > 0$.

2. Examples of abandonment distribution functions (F_n) :

(a) Let $F_n \equiv F$ for all n , and F be differentiable with a bounded derivative on $[0, \delta]$ for some $\delta > 0$. Hence, let $H(x) = F'(0)x$ in Assumption 3.3.

(b) We may take $F_n(x) = 1 - \exp(-\int_0^x h(\sqrt{n}u)du)$ for $x \geq 0$, where h is a non-negative continuous function as in (14) of [26]. In this case, $H(x) = \int_0^x h(u)du$ and it satisfies Assumption 3.3 since h is continuous. Indeed, for any general sequence (F_n) , if $F_n'(\frac{x}{\sqrt{n}})$ converges to a non-negative function $h(x)$ uniformly on compact sets, then (F_n) satisfies Assumption 3.3 with the limiting function $H(x) = \int_0^x h(u)du$.

(c) Here we provide a simple example to illustrate that there can be many limiting functions $H(\cdot)$ other than the ones described in (a) and (b) above. Let $H(\cdot)$ be any non-negative, non-decreasing, locally Lipschitz continuous function which satisfies $H(0) = 0$ and $H(+\infty) = +\infty$. We let $F_n(x) = \frac{1}{\sqrt{n}} \min\{H(\sqrt{n}x), \sqrt{n}\}$ for all $x \geq 0$. Then, for each $n \geq 1$, $F_n(0) = 0, F_n(+\infty) = 1$ and F_n is a continuous, non-decreasing probability distribution function. It is evident that the sequence of distribution functions F_n satisfies Assumption 3.3 with limiting function $H(\cdot)$.

Remark 3.5. To describe a specific example of a heavy traffic regime using the same arrival process, we can consider the system $(A(\cdot), V(\cdot))$ satisfying (2.1), (2.7)–(2.11). Then we can scale these processes as described next. First, we introduce the filtration (\mathcal{G}_t^n) by $\mathcal{G}_t^n \equiv \mathcal{G}_{nt}$ for each $n \geq 1$, where $\mathcal{G}_t = \sigma(A(s), V(s) : 0 \leq s \leq t)$. Now let $A_n(t) \equiv A(nt)$ and $V_n(t) \equiv V(nt)$ for all $t \geq 0$. Then using (2.11) and by a change of variable in integration, it easily follows that $\{A_n(t) - n \int_0^t \lambda_n(V_n(s)) ds : t \geq 0\}$ is a (\mathcal{G}_t^n) -martingale.

Throughout, one can consider the arrival intensity $\lambda_n(\cdot)$ as a “control process” related to the n -th system. In a future article, we intend to address an optimal control problem associated with this heavy traffic regime, which minimizes a prescribed cost functional. We refer to [2, 3, 15] for related “thin control” problems and also refer to Chapter VII of [7].

It will be helpful to define fluid-scaled and diffusion-scaled quantities to carry out our analysis. We let

$$\bar{A}_n(t) \equiv \frac{A_n(t)}{n} \quad \text{and} \quad \hat{A}_n(t) \equiv \frac{1}{\sqrt{n}} \left(A_n(t) - n \int_0^t \lambda_n(V_n(s)) ds \right) \quad (3.5)$$

for all $t \geq 0$. We also introduce the diffusion-scaled offered waiting time process

$$\hat{V}_n(t) \equiv \sqrt{n} V_n(t) \quad \text{for all } t \geq 0. \quad (3.6)$$

Since $V_n(\cdot)$ and $\hat{V}_n(\cdot)$ are RCLL processes (and hence with countably many discontinuities), when integrated with respect to Lebesgue measure, it follows that

$$\int_0^t f(V_n(s-)) ds = \int_0^t f\left(\frac{\hat{V}_n(s-)}{\sqrt{n}}\right) ds = \int_0^t f\left(\frac{\hat{V}_n(s)}{\sqrt{n}}\right) ds = \int_0^t f(V_n(s)) ds$$

for all $t \geq 0$, where f is any bounded Borel-measurable function. Hence, throughout this article, we use $\int_0^t \lambda_n(V_n(s)) ds$ or $\int_0^t \lambda_n\left(\frac{\hat{V}_n(s)}{\sqrt{n}}\right) ds$ appropriately, when integrated with respect to Lebesgue measure.

We also define the diffusion-scaled martingales with respect to the filtration (\mathcal{F}_t^n) (see (3.3)), given by

$$\widehat{M}_n^v(t) \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} (v_j - 1) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]}, \quad \widehat{M}_n^d(t) \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \left(\mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} - \mathbb{E}(\mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} | \mathcal{F}_{j-1}^n) \right). \quad (3.7)$$

Using (3.1), (3.4) and the state equation described in (2.7), and after simple algebraic manipulations, we obtain

$$\begin{aligned} V_n(t) + \frac{1}{n} \int_0^t F_n(V_n(s-)) dA_n(s) &= \frac{1}{n} \left(A_n(t) - n \int_0^t \lambda_n(V_n(s)) ds \right) \\ &+ \frac{1}{\sqrt{n}} \left(\widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) \right) \\ &+ \int_0^t [\lambda_n(V_n(s)) - 1] ds + I_n(t), \end{aligned} \quad (3.8)$$

where $I_n(t) = \int_0^t \mathbf{1}_{[V_n(s)=0]} ds$ for all $t \geq 0$.

4 Weak Convergence

4.1 Fluid limits

Throughout we use $\|\cdot\|_T$ defined by $\|f\|_T = \sup_{t \in [0, T]} |f(s)|$ for any f in $D[0, \infty)$. Our aim here is first to establish the fluid limit $\lim_{n \rightarrow \infty} \|V_n\|_T = 0$ in probability for each $T > 0$. We intend to employ several properties of the Skorokhod map Γ (see, for example, [20, 8, 33, 16]) in the discussion below. The Skorokhod map $\Gamma : D[0, \infty) \rightarrow D[0, \infty)$ is explicitly defined by

$$\Gamma(f)(t) = f(t) + \sup_{s \in [0, t]} (-f(s))^+ \quad \text{for all } t \geq 0. \quad (4.1)$$

Given a function f in $D[0, \infty)$, the pair $(\Gamma(f), \sup_{s \in [0, \cdot]} (-f(s))^+)$ is called the ‘‘Skorokhod decomposition’’ of f and this decomposition is unique. In (3.8), we let

$$X_n(t) \equiv \frac{1}{n}(A_n(t) - nt) + \frac{1}{\sqrt{n}} \left(\widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) \right) - \frac{1}{n} \int_0^t F_n(V_n(s-)) dA_n(s). \quad (4.2)$$

Thus, by (3.8)–(4.2), we observe that (V_n, I_n) is the Skorokhod decomposition of the process X_n and thus

$$V_n(t) = \Gamma(X_n)(t), \quad \text{for all } t \geq 0. \quad (4.3)$$

Theorem 4.1. (*Fluid limit*) For each $T > 0$,

$$\|V_n\|_T \Rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (4.4)$$

Proof. First we show that

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\widehat{A}_n\|_T = 0 \quad \text{a.s.}, \quad (4.5)$$

for each $T > 0$. For the n -th system, we consider the martingale $\widehat{A}_n(\cdot)$ described in (3.5). Using a random time change theorem for point processes (use Theorem T16 in page 41 of [7] with $\mathcal{F}_t \equiv \sigma(A_n(s), V_n(s) : 0 \leq s \leq t)$ and Lemma L17 therein and the fact that $\lambda_n(x) > \epsilon_0 > 0$ to guarantee $\int_0^\infty n \lambda_n(V_n(s)) ds = +\infty$ a.s.), there is a unit-rate Poisson process $Y_n(\cdot)$ such that $\widehat{A}_n(t) = \widehat{Y}_n(\int_0^t \lambda_n(V_n(s)) ds)$ for all $t \geq 0$. Here $\widehat{Y}_n(t) \equiv (Y_n(nt) - nt)/\sqrt{n}$ for all $n \geq 1$. Thus, using part (i) of Assumption 3.2, we have $\|\widehat{A}_n\|_T \leq \|\widehat{Y}_n\|_{C_0 T}$ and we can estimate $\mathbb{P}[\frac{1}{\sqrt{n}} \|\widehat{A}_n\|_T > \epsilon]$ for $\epsilon > 0$ arbitrary. Since \widehat{Y}_n also is a martingale, using Doob’s inequality we have

$$\mathbb{P} \left[\frac{1}{\sqrt{n}} \|\widehat{A}_n\|_T > \epsilon \right] \leq \mathbb{P} \left[\|\widehat{Y}_n\|_{C_0 T} > \epsilon \sqrt{n} \right] \leq \frac{\mathbb{E}[\phi(|\widehat{Y}_n(C_0 T)|)]}{\phi(\epsilon \sqrt{n})},$$

where $\phi(\cdot)$ is a non-negative, convex, strictly increasing function on \mathbb{R}_+ . Let $\theta > 1/2$ be fixed. Then there is a real number $x_\theta > 0$ so that $e^x < (1+x) + \theta x^2$ for $0 < x < x_\theta$. We pick any $\alpha > 0$ so that $0 < \alpha < x_\theta$ and let $\phi(x) \equiv e^{\alpha x}$ for all $x > 0$. Then by an elementary computation, we obtain

$$\frac{\mathbb{E}[\phi(|\widehat{Y}_n(C_0 T)|)]}{\phi(\epsilon \sqrt{n})} \leq e^{\theta \alpha^2 C_0 T} e^{-\epsilon \sqrt{n}}.$$

(See also Theorem 5.18, page 114 of Chen and Yao [9].) Consequently,

$$\mathbb{P} \left[\frac{1}{\sqrt{n}} \|\widehat{A}_n\|_T > \epsilon \right] \leq e^{\theta \alpha^2 C_0 T} e^{-\epsilon \sqrt{n}},$$

where $\theta > 1/2$, $\alpha > 0$ and $C_0 > 0$ are constants independent of n . Now we can apply Borel-Cantelli lemma to conclude the a.s. limit in (4.5). Hence, there is $n_0(\omega) \in \mathbb{N}$ such that $\widehat{A}_n(T) \leq \sqrt{n}$ for all $n \geq n_0(\omega)$. This together with Assumption 3.2(i) implies that

$$A_n(T) \leq \sqrt{n} \widehat{A}_n(T) + C_0 n T \leq n + C_0 n T \leq K_1 n \quad \text{for all } n \geq n_0(\omega),$$

for some constant $K_1 > 0$ which is independent of n . Next, using (3.1)

$$\|V_n\|_T \leq \frac{1}{n} \sum_{j=1}^{A_n(T)} v_j \leq \frac{1}{n} \sum_{j=1}^{K_1 n} v_j \quad \text{for all } n \geq n_0(\omega).$$

But $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{K_1 n} v_j$ exists a.s. by SLLN and hence $\|V_n\|_T \leq K_2 T$ for all $n \geq n_1(\omega)$ and for some constant $K_2 > 0$ which is independent of n . This, together with Assumption 3.2(ii), implies that

$$\int_0^T |\lambda_n(V_n(s)) - 1| ds \leq \sup_{x \in [0, K_2 T]} |\lambda_n(x) - 1| T \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Hence $\lim_{n \rightarrow \infty} \int_0^T |\lambda_n(V_n(s)) - 1| ds = 0$ a.s. Since

$$\sup_{t \in [0, T]} |\bar{A}_n(t) - t| \leq \frac{1}{\sqrt{n}} \sup_{t \in [0, T]} |\widehat{A}_n(t)| + \int_0^T |\lambda_n(V_n(s)) - 1| ds,$$

using the above fact with (4.5), we obtain

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} |\bar{A}_n(t) - t| = 0 \quad \text{a.s.} \tag{4.6}$$

Next, we consider the martingale term $\frac{1}{\sqrt{n}}(\widehat{M}_n^v(t) - \widehat{M}_n^d(t))$. Notice that

$$\mathbb{E} \left([\widehat{M}_n^v](T) \right) \leq \frac{1}{n^2} \sum_{j=1}^{[nT]} \mathbb{E}(v_j - 1)^2 \leq \frac{\sigma_s^2 T}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Similarly, it follows that $\mathbb{E} \left([\widehat{M}_n^d](T) \right) \leq \frac{4T}{n} \rightarrow 0$ as $n \rightarrow \infty$. We consider the vector valued martingale $\mathbb{M}_n(t) = (\widehat{M}_n^v(t)/\sqrt{n}, \widehat{M}_n^d(t)/\sqrt{n})$ and define $\mathbb{M}_n^*(t) \equiv \sup_{s \in [0, t]} |\mathbb{M}_n(s)|$ for all $t \geq 0$. Using

Doob's inequality for the submartingale $|\mathbb{M}_n(t)|^2$ once more, we obtain $\mathbb{E} \left[\sup_{t \in [0, T]} |\mathbb{M}_n(t)|^2 \right] \leq CT/n$

where $C > 0$ is a generic constant independent of n . We conclude that $\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} |\mathbb{M}_n(t)|^2 \right] = 0$.

Consequently, $(\mathbb{M}_n^*(T))^2 \Rightarrow 0$ as $n \rightarrow \infty$. This, together with (4.6) and the random change of time theorem (cf. Section 14, [6]), implies that

$$\mathbb{M}_n^*(\bar{A}_n(T)) \Rightarrow 0 \tag{4.7}$$

as $n \rightarrow \infty$. Hence, using (4.6) and (4.7), we have

$$\sup_{t \in [0, T]} |\bar{A}_n(t) - t| + \mathbb{M}_n^*(\bar{A}_n(T)) \rightarrow 0 \text{ in probability,} \quad (4.8)$$

as $n \rightarrow \infty$. Let

$$\begin{aligned} \mathcal{T}_n(t) &\equiv X_n(t) + \frac{1}{n} \int_0^t F_n(V_n(s-)) dA_n(s) \\ &= \frac{1}{n} (A_n(t) - nt) + \frac{1}{\sqrt{n}} \left(\widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) \right), \end{aligned} \quad (4.9)$$

where X_n is described in (4.2). With (4.8) in hand and using (4.2), we observe that

$$\lim_{n \rightarrow \infty} \|\mathcal{T}_n\|_T = 0 \text{ in probability,} \quad (4.10)$$

for each $T > 0$. By (4.2), we have $\mathcal{T}_n(t) \geq X_n(t)$ for all $t \geq 0$ and $\mathcal{T}_n(t) - X_n(t)$ is a non-negative, non-decreasing process in $D[0, \infty)$. Therefore, we can use the comparison theorem for the Skorokhod map Γ (Propositions 3.4 and 3.5 of [8]) to conclude that

$$0 \leq V_n(t) = \Gamma(X_n)(t) \leq \Gamma(\mathcal{T}_n)(t) \text{ for all } t \geq 0. \quad (4.11)$$

Since $\|\Gamma(\mathcal{T}_n)\|_T \leq 2\|\mathcal{T}_n\|_T$ by the Lipschitz continuity of Γ , using (4.10)–(4.11) we can conclude

$$\lim_{n \rightarrow \infty} \|V_n\|_T = 0 \text{ in probability.} \quad (4.12)$$

This completes the proof. ■

Remark 4.2. *In Theorem 6.5 of Section 6, we are able to show that $\lim_{n \rightarrow \infty} \mathbb{E} [\|V_n\|_T^m] = 0$ for some $m > 2$, under an additional hypothesis given in (6.6).*

4.2 Diffusion limits

Here we intend to establish the weak convergence of the process $\widehat{V}_n(\cdot)$ defined in (3.6) to a (reflected) diffusion process. We need to obtain several technical results to achieve this objective. Our first proposition is an improvement of (4.4). Using (3.5)–(3.8), we can describe the state equation for $\widehat{V}_n(\cdot)$ by

$$\begin{aligned} \widehat{V}_n(t) &+ \frac{1}{\sqrt{n}} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) \\ &= \widehat{A}_n(t) + \widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) + \sqrt{n} \int_0^t \left(\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right) ds + \sqrt{n} I_n(t), \end{aligned} \quad (4.13)$$

where $I_n(t) = \int_0^t \mathbf{1}_{[\widehat{V}_n(s)=0]}(s) ds$. Notice that $(\widehat{V}_n, \sqrt{n} I_n)$ is indeed the Skorokhod decomposition (see the line below (4.1)) of the process $\sqrt{n} X_n(\cdot)$ where X_n is described in (4.2). Then, $\Gamma(\sqrt{n} X_n)(t) = \widehat{V}_n(t)$ for all $t \geq 0$ where Γ is given in (4.1). We use this fact in the following proposition.

Proposition 4.3. *We have for each $T > 0$,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\|\widehat{V}_n\|_T > K \right] = 0. \quad (4.14)$$

Proof. We introduce $\widehat{X}_n(t) \equiv \sqrt{n}X_n(t)$ and

$$\widehat{Z}_n(t) \equiv \widehat{X}_n(t) + \frac{1}{\sqrt{n}} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) + \sqrt{n} \int_0^t \left(\lambda \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right)^- ds$$

for all $t \geq 0$, where X_n is defined in (4.2) and $x^- = -\min\{x, 0\}$. Notice that $\{\widehat{Z}_n(t) - \widehat{X}_n(t) : t \geq 0\}$ is a non-negative, non-decreasing process and thus by a comparison argument as in (4.11), we obtain $0 \leq \widehat{V}_n(t) \leq \Gamma(\widehat{Z}_n)(t)$ for all $t \geq 0$. Consequently, using the Lipschitz continuity of Γ , we get

$$\|\widehat{V}_n\|_T \leq 2\|\widehat{Z}_n\|_T \quad \text{for all } T \geq 0.$$

But

$$\widehat{Z}_n(t) = \widehat{A}_n(t) + \widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) + \sqrt{n} \int_0^t (\lambda_n(V_n(s)) - 1)^+ ds \quad \text{for all } t \geq 0,$$

and hence we have

$$\|\widehat{V}_n\|_T \leq C_1 \left[\|\widehat{A}_n\|_T + \sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))| + \sup_{t \in [0, T]} |\widehat{M}_n^d(\bar{A}_n(t))| + \sqrt{n} \int_0^T (\lambda_n(V_n(s)) - 1)^+ ds \right], \quad (4.15)$$

where $C_1 > 0$ is a generic constant independent of T . To estimate $\mathbb{P}[\|\widehat{V}_n\|_T > K]$ for $K > 0$, we estimate the probability corresponding to each term in the right hand side of (4.15). Throughout, we consider $K > 0$ to be a generic constant. First, we estimate $\mathbb{P}[\|\widehat{A}_n\|_T > K]$. Using the same technique used in the proof of (4.5), we obtain

$$\mathbb{P} \left[\|\widehat{A}_n\|_T > K \right] \leq \frac{\mathbb{E}|\widehat{Y}_n(C_0T)|^2}{K^2} \leq \frac{CT}{K^2},$$

where $C_0 > 0$ is the constant as in Assumption 3.2 (i) and $C > 0$ is a generic constant independent of K . Here $\widehat{Y}_n(t) \equiv (Y_n(nt) - nt)/\sqrt{n}$ for all $t \geq 0$ and $Y_n(\cdot)$ is a unit-rate Poisson process. Hence

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\|\widehat{A}_n\|_T > K \right] = 0. \quad (4.16)$$

Next we consider $\mathbb{P}[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))| > K]$, and here we intend to use (4.6). We have

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))| > K \right] &\leq \mathbb{P} \left[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))| > K, \bar{A}_n(T) \leq 2T \right] + \mathbb{P}[\bar{A}_n(T) > 2T] \\ &\leq \mathbb{P} \left[\sup_{t \in [0, 2T]} |\widehat{M}_n^v(t)| > K \right] + \mathbb{P}[\bar{A}_n(T) > 2T]. \end{aligned}$$

Notice that the quadratic variation process $[\widehat{M}_n^v]$ satisfies

$$[\widehat{M}_n^v](t) \leq \frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1)^2 \quad \text{and hence} \quad \mathbb{E}([\widehat{M}_n^v](2T)) \leq 2T\sigma_s^2,$$

where $\sigma_s^2 \equiv \mathbb{E}(v_j - 1)^2 > 0$ is a finite constant. Thus, from Doob's maximal inequality for submartingales (cf. [18]) we obtain $\mathbb{P} \left[\sup_{t \in [0, 2T]} |\widehat{M}_n^v(t)| > K \right] \leq CT/K^2$, where $C > 0$ is a constant independent of T and n . Hence $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, 2T]} |\widehat{M}_n^v(t)| > K \right] = 0$ and by (4.6), $\lim_{n \rightarrow \infty} \mathbb{P}[\bar{A}_n(T) > 2T] = 0$. Thus we have

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))| > K \right] = 0. \quad (4.17)$$

The proof of $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} |\widehat{M}_n^d(\bar{A}_n(t))| > K \right] = 0$ is very similar to that of (4.17).

For the last term in the right hand side of (4.15), we intend to use (4.4). Recall $\delta_0 > 0$ and $M > 0$ are as in Assumption 3.2(iii). Then we have

$$\begin{aligned} \mathbb{P} \left[\sqrt{n} \int_0^T (\lambda_n(V_n(s)) - 1)^+ ds > K \right] &\leq \mathbb{P} \left[\sqrt{n} \int_0^T (\lambda_n(V_n(s)) - 1)^+ ds > K, \|V_n\|_T < \delta_0 \right] \\ &\quad + \mathbb{P} [\|V_n\|_T \geq \delta_0] \\ &\leq \mathbb{P} [MT > K, \|V_n\|_T \leq \delta_0] + \mathbb{P} [\|V_n\|_T \geq \delta_0]. \end{aligned}$$

Notice that $\lim_{K \rightarrow \infty} \mathbb{P} [MT > K, \|V_n\|_T \leq \delta_0] = 0$ and by (4.12) we obtain $\lim_{n \rightarrow \infty} \mathbb{P} [\|V_n\|_T \geq \delta_0] = 0$. Consequently,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sqrt{n} \int_0^T (\lambda_n(V_n(s)) - 1)^+ ds > K \right] = 0 \quad (4.18)$$

Now, (4.15)–(4.18) imply (4.14) and this completes the proof. \blacksquare

Next, we introduce

$$R_n(i) \equiv \sum_{j=1}^i \mathbf{1}_{[V_n(t_j^n -) \geq d_j^n]}, \quad (4.19)$$

which represents the number of customers who abandoned the system among the first i customers. We also define its fluid-scaled term

$$\bar{R}_n(t) \equiv \frac{1}{n} R_n(\lceil nt \rceil) = \frac{1}{n} \sum_{j=1}^{\lceil nt \rceil} \mathbf{1}_{[V_n(t_j^n -) \geq d_j^n]} \quad (4.20)$$

for all $t \geq 0$. We intend to show $\bar{R}_n(\cdot) \Rightarrow 0$. In the case of constant intensity, this is indeed proved in the Lemma 5.5 of [26]. But, our proof mainly uses the previous proposition and martingale property of \widehat{A}_n .

Lemma 4.4. *For each $T > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{E}[\bar{R}_n(T)] = 0. \quad (4.21)$$

Proof. Consider the martingale $\{\widehat{A}_n(t) : t \geq 0\}$ and the stopping times $\{t_{[nT]}^n : n \geq 1\}$. Let $\bar{M} > 0$ be a constant and $\tau_n \equiv t_{[nT]}^n \wedge \bar{M}$. Then $A_n(\tau_n) \leq A_n(t_{[nT]}^n) = [nT]$. Since τ_n is a bounded stopping

time, $\mathbb{E}[\widehat{A}_n(\tau_n)] = 0$. Thus $0 \leq \mathbb{E}[A_n(\tau_n) - n \int_0^{\tau_n} \lambda_n(V_n(s))ds]$ and using Assumption 3.2, we have

$$n\epsilon_0 \mathbb{E}[\tau_n] \leq \mathbb{E}[A_n(\tau_n)] \leq [nT],$$

which implies $\mathbb{E}[\tau_n] \leq T/\epsilon_0$. By letting $\bar{M} \uparrow +\infty$, we have

$$\mathbb{E}[t_{[nT]}^n] \leq C_1 T, \quad (4.22)$$

where $C_1 > 0$ is a generic constant. Next, we estimate $\mathbb{P}[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \geq K]$. Let $\epsilon > 0$ be arbitrary. We pick a large constant C_2 such that $0 < \frac{C_1 T}{C_2 T} < \frac{\epsilon}{4}$. Then we have

$$\begin{aligned} \mathbb{P}\left[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \geq K\right] &\leq \mathbb{P}\left[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \geq K, t_{[nT]}^n < C_2 T\right] + \mathbb{P}[t_{[nT]}^n \geq C_2 T] \\ &\leq \mathbb{P}\left[\|\widehat{V}_n\|_{C_2 T} > K\right] + \frac{\epsilon}{4}, \end{aligned}$$

where the second inequality follows from Chebyshev's inequality and (4.22). Also,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left[\|\widehat{V}_n\|_{C_2 T} \geq K\right] = 0$$

by (4.14). Hence, there exists a $K_0 > 0$ such that for all $K > K_0$, $\limsup_{n \rightarrow \infty} \mathbb{P}[\|\widehat{V}_n\|_{C_2 T} \geq K] < \epsilon/4$ and as a consequence we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \geq K\right] < \frac{\epsilon}{2} \quad \text{for all } K > K_0. \quad (4.23)$$

To estimate $\mathbb{E}[\bar{R}_n(T)]$, we pick $K > K_0$ and consider $\mathbb{P}[V_n(t_j^n -) > d_j^n]$, where $j = 1, 2, \dots, [nT]$. Then it follows that

$$\begin{aligned} \mathbb{P}[V_n(t_j^n -) > d_j^n] &\leq \mathbb{P}\left[V_n(t_j^n -) > d_j^n > \frac{K}{\sqrt{n}}\right] + \mathbb{P}\left[d_j^n \leq \frac{K}{\sqrt{n}}\right] \\ &\leq \mathbb{P}\left[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) > K\right] + F_n\left(\frac{K}{\sqrt{n}}\right). \end{aligned}$$

By Assumption 3.3, $\lim_{n \rightarrow \infty} F_n(\frac{K}{\sqrt{n}}) = 0$ and consequently, there is a $n_0 \geq 1$ such that

$$\sup_{1 \leq j \leq [nT]} \mathbb{P}[V_n(t_j^n -) > d_j^n] < \epsilon$$

for all $n \geq n_0$. Hence by (4.20), $\mathbb{E}[\bar{R}_n(T)] \leq \frac{1}{n}[nT]\epsilon \leq T\epsilon$ and we conclude that $\lim_{n \rightarrow \infty} \mathbb{E}[\bar{R}_n(T)] = 0$. This completes the proof. \blacksquare

Our next step is to show that the term $\frac{1}{\sqrt{n}} \int_0^t F_n\left(\frac{\widehat{V}_n(s-)}{\sqrt{n}}\right) dA_n(s)$ in the state equation (4.13) can be well approximated by $\int_0^t H(\widehat{V}_n(s))ds$, where $H(\cdot)$ is given in Assumption 3.3.

Lemma 4.5. *We have for each $T > 0$,*

$$\sup_{t \in [0, T]} \left| \frac{1}{\sqrt{n}} \int_0^t F_n\left(\frac{\widehat{V}_n(s-)}{\sqrt{n}}\right) dA_n(s) - \int_0^t H(\widehat{V}_n(s))ds \right| \rightarrow 0 \text{ in probability as } n \rightarrow \infty. \quad (4.24)$$

Proof. We recall $\bar{A}_n(t) = \frac{1}{n}A_n(t)$ and it satisfies (4.6). Hence we can write

$$\begin{aligned} & \frac{1}{\sqrt{n}} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) - \int_0^t H(\widehat{V}_n(s)) ds \\ &= \int_0^t \sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) (d\bar{A}_n(s) - ds) + \int_0^t \left(\sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) - H(\widehat{V}_n(s)) \right) ds. \end{aligned} \quad (4.25)$$

To obtain (4.24), we estimate the right hand side of (4.25) using (4.6) and Assumption 3.3. First we note that $\{\bar{M}_n^A(t) \equiv \bar{A}_n(t) - \int_0^t \lambda_n(V_n(s)) ds : t \geq 0\}$ is a martingale and its quadratic variation is given by $[\bar{M}_n^A](T) = \frac{1}{n} \bar{A}_n(T)$. By random time change theorem of point processes (see (2.13) and the proof of (4.5)),

$$\bar{A}_n(T) = \frac{1}{n} Y_n \left(n \int_0^T \lambda_n(V_n(s)) ds \right) \leq \frac{1}{n} Y_n(nC_0T),$$

where Y_n is a unit-rate Poisson process and $C_0 > 0$ is as in Assumption 3.2(i). Thus, $[\bar{M}_n^A](T) \leq \frac{1}{n^2} Y_n(nC_0T)$ and

$$d\bar{A}_n(t) - dt = d\bar{M}_n^A(t) + (\lambda_n(V_n(t)) - 1)dt$$

and the first term on the right side of (4.25) is equal to

$$\sqrt{n} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d\bar{M}_n^A(s) + \sqrt{n} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) (\lambda_n(V_n(s)) - 1) ds. \quad (4.26)$$

We consider an arbitrary $\delta > 0$. Since $0 \leq F_n(x) \leq 1$ for all x and $F_n \left(\frac{\widehat{V}_n(t-)}{\sqrt{n}} \right)$ is a predictable process, the integral $\int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d\bar{M}_n^A(s)$ defines a martingale and its quadratic variation process is given by $\int_0^t F_n^2 \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d[\bar{M}_n^A](s)$ (see [24]). Hence using Doob's maximal inequality, we have

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} \sqrt{n} \left| \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d\bar{M}_n^A(s) \right| > \delta \right] &\leq \frac{n}{\delta^2} \mathbb{E} \left[\int_0^T F_n^2 \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d[\bar{M}_n^A](s) \right] \\ &\leq \frac{n}{\delta^2} \mathbb{E} [F_n(\|V_n\|_T) [\bar{M}_n^A](T)] \\ &\leq \frac{1}{n\delta^2} \mathbb{E} [F_n(\|V_n\|_T) Y_n(nC_0T)] \\ &\leq \frac{1}{n\delta^2} (\mathbb{E}[F_n^2(\|V_n\|_T)] \mathbb{E}[Y_n^2(nC_0T)])^{1/2} \\ &\leq \frac{C_1 T}{\delta^2} (\mathbb{E}[F_n^2(\|V_n\|_T)])^{1/2}. \end{aligned} \quad (4.27)$$

In the above estimation, for the last two inequalities, we have used Cauchy-Schwarz inequality and the fact that $\mathbb{E}[Y_n^2(nC_0T)] \leq C_1^2 n^2 T^2$ for some generic constant $C_1 > 0$ independent of n and T .

Next, we will show $\mathbb{E}[F_n^2(\|V_n\|_T)]$ approaches 0 as $n \rightarrow \infty$. By Assumption 3.3, there exist n_0 and $M_1 > 0$ such that

$$\sup_{x \in [0, K]} F_n \left(\frac{x}{\sqrt{n}} \right) < \frac{M_1}{\sqrt{n}} \quad \text{for all } n \geq n_0.$$

We consider $n > n_0$ and then

$$\begin{aligned} \mathbb{E} [F_n^2(\|V_n\|_T)] &= \mathbb{E} \left[F_n^2(\|V_n\|_T) \mathbf{1}_{\|\|V_n\|_T \leq \frac{K}{\sqrt{n}}}\right] + \mathbb{E} \left[F_n^2(\|V_n\|_T) \mathbf{1}_{\|\|V_n\|_T > \frac{K}{\sqrt{n}}}\right] \\ &\leq \frac{M_1^2}{n} + \mathbb{P}[\sqrt{n}\|V_n\|_T > K]. \end{aligned}$$

Now, letting $n \rightarrow \infty$ and then $K \rightarrow \infty$ and using (4.14), we obtain $\lim_{n \rightarrow \infty} \mathbb{E}[F_n^2(\|V_n\|_T)] = 0$. Consequently, by (4.27), we conclude that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \sqrt{n} \left| \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) d\bar{M}_n^A(t) \right| > \delta \right] = 0. \quad (4.28)$$

Similar to the previous estimation, we obtain

$$\begin{aligned} &\mathbb{P} \left[\sqrt{n} \int_0^T F_n(V_n(s-)) |\lambda_n(V_n(s)) - 1| ds > \delta \right] \\ &\leq \mathbb{P} \left[\sqrt{n} \int_0^T F_n(V_n(s-)) |\lambda_n(V_n(s)) - 1| ds > \delta, \sqrt{n}\|V_n\|_T \leq K \right] + \mathbb{P}[\sqrt{n}\|V_n\|_T > K] \\ &\leq \mathbb{P} \left[M_1 \int_0^T |\lambda_n(V_n(s)) - 1| ds > \delta \right] + \mathbb{P}[\sqrt{n}\|V_n\|_T > K]. \end{aligned} \quad (4.29)$$

In the derivation of (4.6), we have obtained $\lim_{n \rightarrow \infty} \int_0^T |\lambda_n(V_n(s)) - 1| ds = 0$ a.s. This together with (4.14) implies that the right hand side of (4.29) tends to zero as $n \rightarrow \infty$. This yields

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sqrt{n} \int_0^T F_n(V_n(s-)) |\lambda_n(V_n(s)) - 1| ds > \delta \right] = 0. \quad (4.30)$$

Consequently, using (4.26), (4.28) and (4.30), we obtain for each $T > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \int_0^t \sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) (d\bar{A}_n(s) - ds) \right| > \delta \right] = 0. \quad (4.31)$$

Finally, we intend to establish

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \int_0^t \left(\sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) - H(\widehat{V}_n(s)) \right) ds \right| > \delta \right] = 0. \quad (4.32)$$

Pick $\epsilon > 0$ so that $0 < \epsilon < \frac{\delta}{T}$. By Assumption 3.3, we take any $K > 0$ and then there is a $n_1 \in \mathbb{N}$ such that $\sup_{x \in [0, K]} |\sqrt{n} F_n(\frac{x}{\sqrt{n}}) - H(x)| < \epsilon$ for all $n \geq n_1$. We consider $n > n_1$ and estimate

$$\begin{aligned} &\mathbb{P} \left[\int_0^T \left| \left(\sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) - H(\widehat{V}_n(s)) \right) \right| ds > \delta \right] \\ &\leq \mathbb{P} \left[\int_0^T \left| \left(\sqrt{n} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) - H(\widehat{V}_n(s)) \right) \right| ds > \delta, \sqrt{n}\|V_n\|_T \leq K \right] + \mathbb{P}[\sqrt{n}\|V_n\|_T > K] \\ &\leq \mathbb{P}[\epsilon T > \delta, \sqrt{n}\|V_n\|_T \leq K] + \mathbb{P}[\sqrt{n}\|V_n\|_T > K]. \end{aligned}$$

Since $\epsilon T < \delta$, the first term of the above is 0 for all $n > n_1$. Also, $\lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}[\sqrt{n}\|V_n\|_T > K] = 0$. Hence (4.32) follows. Therefore, (4.31) and (4.32) yield (4.24). This completes the proof. \blacksquare

Our next lemma shows that the term $\sqrt{n} \int_0^t (1 - \lambda_n(\widehat{V}_n(s)/\sqrt{n})) ds$ can be well approximated by $\int_0^t u(\widehat{V}_n(s)) ds$, where the function $u(\cdot)$ is as given in Assumption 3.2.

Lemma 4.6. *We have for each $T > 0$,*

$$\int_0^T \left| \sqrt{n} \left(1 - \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) \right) - u(\widehat{V}_n(s)) \right| ds \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty, \quad (4.33)$$

and consequently,

$$\sup_{t \in [0, T]} \left| \int_0^t \left[\sqrt{n} \left(1 - \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) \right) - u(\widehat{V}_n(s)) \right] ds \right| \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty. \quad (4.34)$$

Proof. Fix $T > 0$. Let $\delta > 0$ and pick $\epsilon > 0$ small so that $\epsilon T < \delta$. Let $K > 0$ be arbitrary. By Assumption 3.2(iv), there is $n_0 \equiv n_0(K)$ so that $\sup_{x \in [0, K]} |\sqrt{n}(1 - \lambda_n(x/\sqrt{n})) - u(x)| < \epsilon$ whenever

$n \geq n_0$. Thus, on the set $[\|\widehat{V}_n\|_T < K]$, we have

$$\int_0^T \left| \sqrt{n} \left(1 - \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) \right) - u(\widehat{V}_n(s)) \right| ds < \epsilon T < \delta, \quad \text{for all } n \geq n_0.$$

Following an estimation similar to that of Lemma 4.5, we can have

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left[\int_0^T \left| \sqrt{n} \left(1 - \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) \right) - u(\widehat{V}_n(s)) \right| ds > \delta \right] \leq \limsup_{n \rightarrow \infty} \mathbb{P}[\|\widehat{V}_n\|_T > K].$$

Hence, using (4.14), desired conclusion (4.33) follows. ■

The following result is an immediate consequence of Lemmas 4.5 and 4.6. Therefore, we omit the proof.

Lemma 4.7. *For all $t \geq 0$, let*

$$\epsilon_n(t) \equiv \int_0^t \frac{1}{\sqrt{n}} F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) - \int_0^t H(\widehat{V}_n(s)) ds + \int_0^t \left[\sqrt{n} \left(1 - \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) \right) - u(\widehat{V}_n(s)) \right] ds. \quad (4.35)$$

Then for each $T > 0$, $\|\epsilon_n\|_T \rightarrow 0$ in probability as $n \rightarrow \infty$.

To discuss the weak convergence of the process $\{\widehat{V}_n(t) : t \geq 0\}$, we rewrite the state equation (4.13) in the following form:

$$\widehat{V}_n(t) = \xi_n(t) - \epsilon_n(t) - \int_0^t u(\widehat{V}_n(s)) ds - \int_0^t H(\widehat{V}_n(s)) ds + \sqrt{n} I_n(t), \quad (4.36)$$

where

$$\xi_n(t) \equiv \widehat{A}_n(t) + \widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)), \quad (4.37)$$

and $I_n(t)$, $\epsilon_n(t)$ are given in (3.8), (4.35), respectively.

4.3 Generalized Skorokhod map and weak convergence

Following Section 4 of [26], we introduce the generalized Skorokhod map.

Definition 4.8. *Let $p : [0, \infty) \rightarrow [0, \infty)$ be a locally Lipschitz continuous function. Then for a given x in $D[0, \infty)$ with $x(0) \geq 0$, there exists a unique pair of functions (z, ℓ) such that z, ℓ are also in $D[0, \infty)$ and*

- (i) $z(t) = x(t) - \int_0^t p(z(u))du + \ell(t)$, $z(t) \geq 0$, for all $t \geq 0$,
- (ii) $\ell(0) = 0$, $\ell(\cdot)$ is non-decreasing, and $\int_0^\infty z(t)d\ell(t) = 0$.

We use the notation in [26] and introduce two functions $\phi^p : D[0, \infty) \rightarrow D[0, \infty)$ and $\psi^p : D[0, \infty) \rightarrow D[0, \infty)$ given by

$$(\phi^p, \psi^p)(x) = (z, \ell). \quad (4.38)$$

This describes a generalized Skorokhod decomposition of the function x in $D[0, \infty)$. Since (4.36) describes precisely this decomposition, it is easy to observe that

$$(\phi^p, \psi^p)(\xi_n - \epsilon_n) = (\widehat{V}_n, \sqrt{n}I_n), \quad (4.39)$$

where $p(x) = u(x) + H(x)$ for all $x \geq 0$, and in this case $p(\cdot)$ is a locally Lipschitz continuous function.

In [26], the function $p(\cdot)$ is of the form $p(x) = \int_0^x h(u)du$ where $h(\cdot)$ is a non-negative continuous function. But their discussion on existence and uniqueness of the pair (z, ℓ) for a given x in $D[0, \infty)$ as well as on the continuity properties of (ϕ^p, ψ^p) in $D[0, \infty)$ endowed with the Skorokhod J_1 -topology holds for a non-negative, locally Lipschitz continuous function $p(\cdot)$ with a few minor changes in their proofs. We state these results in the following proposition and indicate the necessary changes required in the proofs given in [26].

Proposition 4.9. *(Lemma 4.1 and Proposition 4.1 of Reed and Ward [26]) Let $p : [0, \infty) \rightarrow [0, \infty)$ be a non-negative, locally Lipschitz continuous function. Then the following results hold.*

- (i) *For each x in $D[0, \infty)$ with $x(0) \geq 0$, there exists a unique pair of functions (z, ℓ) satisfying the Definition 4.8.*
- (ii) *The functions ϕ^p and ψ^p defined in (4.38) are continuous on $D[0, \infty)$, when it is endowed with the Skorokhod's J_1 -topology.*

Proof. Proofs of the above statements essentially follow from those of Lemma 4.1 and Proposition 4.1 of [26] with the changes described below. Given x in $D[0, \infty)$, Piccard's iteration scheme was used in Lemma 4.1 of [26] to obtain a unique solution to

$$w(t) = x(t) - \int_0^t p(\Gamma(w)(s))ds, \quad \text{for } t \geq 0, \quad (4.40)$$

where Γ is the Skorokhod map defined in (4.1). Given x in $D[0, T]$, they introduce the iterative scheme by $w_0(t) \equiv 0$ on $[0, T]$ and

$$w_n(t) = x(t) - \int_0^t p(\Gamma(w_{n-1})(s))ds, \quad \text{for all } 0 \leq t \leq T \quad \text{and } n \geq 1. \quad (4.41)$$

In this situation, we need to establish the bound $\sup_{n \geq 1} \sup_{t \in [0, T]} |w_n(t)| \leq M < \infty$, where $M > 0$ is a constant depending on x .

Since $x(t) - w_n(t) = \int_0^t p(\Gamma(w_{n-1})(s)) ds$ is a non-negative, non-decreasing function, by a comparison result for the Skorokhod map (cf. [20]), we have $\Gamma(w_n)(t) \leq \Gamma(x)(t)$ for all $0 \leq t \leq T$. Next, introduce $p^*(y) \equiv \max_{z \in [0, y]} p(z)$, then we have

$$p(\Gamma(w_n)(t)) \leq p^*(\Gamma(w_n)(t)) \leq p^*(\Gamma(x)(t)) \quad \text{for all } 0 \leq t \leq T,$$

and hence by (4.41), this implies $x(t) - w_n(t) \leq \int_0^t p^*(\Gamma(x)(s)) ds$. By (4.41), $x(t) - w_n(t) \geq 0$ for all $0 \leq t \leq T$ since $p(\cdot)$ is non-negative. Therefore, $x(t) - \int_0^t p^*(\Gamma(x)(s)) ds \leq w_n(t) \leq x(t)$ for all $0 \leq t \leq T$ and $n \geq 1$. Hence, the required bound $\sup_{n \geq 1} \sup_{t \in [0, T]} |w_n(t)| \leq M < \infty$ holds for some

$M > 0$. Now one can follow the proof of convergence of the sequence (w_n) to a function w in $D[0, T]$ as in [26]. Moreover, $\sup_{t \in [0, T]} |w(t)| \leq M < \infty$ holds.

The uniqueness of the solution w to (4.40) for a given x in $D[0, \infty)$ with $x(0) \geq 0$ is quite straightforward. Assume w_1 and w_2 are two solutions to (4.40). Then following the same proof above, we have $\sup_{t \in [0, T]} |w_i(t)| \leq M$ for $i = 1, 2$, and consequently $\sup_{t \in [0, T]} |\Gamma(w_i)(t)| \leq 2M$ where

$M > 0$ is a constant. Let K_M be the Lipschitz constant of $p(\cdot)$ on the interval $[0, 2M]$. Then using (4.40) for w_1 and w_2 , we obtain $\|w_1 - w_2\|_t \leq 2K_M \int_0^t \|w_1 - w_2\|_s ds$, where $\|\cdot\|_t$ denotes the sup norm on $[0, t]$. Hence, by Gronwall's inequality, it follows that $\|w_1 - w_2\|_t = 0$ for all $0 \leq t \leq T$ and thus the uniqueness of w in (4.40) follows.

In [26], the map \mathcal{M}^P is defined from $D[0, \infty)$ to $D[0, \infty)$ so that $\mathcal{M}^P(x) = w$, where w is the unique solution to (4.40). Then the continuity of \mathcal{M}^P on $D[0, \infty)$, when this space is endowed with the Skorokhod's J_1 -topology, essentially follows from the same proofs in parts (ii) and (iii) of Lemma 4.1 in [26].

Next, notice that given x in $D[0, \infty)$ with $x(0) \geq 0$, the pair (z, ℓ) defined in (4.38) satisfies

$$z = \Gamma(w) \equiv \Gamma(\mathcal{M}^P(x)) \quad \text{and} \quad \ell = w - \Gamma(w) = \mathcal{M}^P(x) - \Gamma(\mathcal{M}^P(x)).$$

Hence, $\phi^P(x) = \Gamma \circ \mathcal{M}^P(x)$ and $\psi^P(x) = \mathcal{M}^P(x) - \Gamma \circ \mathcal{M}^P(x)$ for each x in $D[0, \infty)$ with $x(0) \geq 0$. Since the Skorokhod map Γ is Lipschitz continuous on $D[0, \infty)$, the proof of part (ii) of the proposition is straightforward. \blacksquare

To obtain the weak convergence of $(\widehat{V}_n(\cdot))_{n \geq 1}$ and to identify the limit, we intend to show that $\xi_n(\cdot) \Rightarrow \sigma W(\cdot)$ as $n \rightarrow \infty$ in $D[0, \infty)$, where $W(\cdot)$ is a standard Brownian motion. Then Lemma 4.7 together with the continuous mapping theorem implies that $\xi_n(\cdot) - \epsilon_n(\cdot) \Rightarrow \sigma W(\cdot)$ as $n \rightarrow \infty$ in $D[0, \infty)$. Since both functions ϕ^P and ψ^P are continuous on $D[0, \infty)$, when this space is endowed with the Skorokhod's J_1 -topology, we can establish the following theorem for the weak convergence of the process $(\widehat{V}_n(\cdot))_{n \geq 1}$.

Theorem 4.10. (*Diffusion limit*) *The process $(\widehat{V}_n, \sqrt{n}I_n)$ converges weakly to (Z, L) as $n \rightarrow \infty$ in $D^{\otimes 2}[0, \infty)$, where (Z, L) is the unique strong solution to the reflected stochastic differential equation*

$$Z(t) = \sigma W(t) - \int_0^t u(Z(s)) ds - \int_0^t H(Z(s)) ds + L(t), \quad (4.42)$$

for all $t \geq 0$. Here, $W(\cdot)$ is a standard Brownian motion and $\sigma > 0$ is a constant which satisfies $\sigma^2 = 1 + \sigma_s^2$. The functions $u(\cdot)$ and $H(\cdot)$ are described in the Assumptions 3.2 and 3.3. The process

$Z(\cdot)$ is non-negative and has continuous sample paths. Here, $L(\cdot)$ is the local-time process of Z at the origin. The process $L(\cdot)$ is unique, continuous, non-decreasing process such that $L(0) = 0$ and

$$\int_0^t Z(s)dL(s) = 0, \quad \text{for all } t \geq 0, \quad (4.43)$$

and that $Z(t) \geq 0$ for $t \geq 0$.

Proof. Recall that the process $\epsilon_n(\cdot)$ in (4.39) converges to 0 uniformly on compact sets in probability as shown in Lemma 4.7. We intend to show $\xi_n(\cdot) \Rightarrow \sigma W(\cdot)$ in $D[0, \infty)$ in Proposition 4.12 and we assume this fact in this proof. Here W is a standard one-dimensional Brownian motion. Hence, by the continuous mapping theorem, we can conclude $\xi_n - \epsilon_n$ weakly converges to σW . Therefore, by the continuity properties of the mapping (ϕ^p, ψ^p) in (4.39) (see Proposition 4.1 in [26]), we have

$$(\phi^p, \psi^p)(\xi_n - \epsilon_n) \Rightarrow (\phi^p, \psi^p)(\sigma W) \text{ as } n \rightarrow \infty.$$

Since the reflected stochastic differential equation in (4.42)–(4.43) has a unique pathwise solution, $(\phi^p, \psi^p)(\sigma W) \equiv (Z, L)$ and the proof of the Theorem 4.10 is complete. \blacksquare

We begin with a technical lemma that will be used in the proof of Proposition 4.12.

Lemma 4.11. *Let $H_n(\cdot)$ be the process defined by*

$$\widehat{H}_n(t) = \frac{1}{\sqrt{n}} \left[([nt] + 1) - n \int_0^{t_{[nt]+1}^n} \lambda_n(V_n(s)) ds \right] \quad (4.44)$$

for all $t \geq 0$. Introduce the vector-valued process $\{\widehat{\mathbf{M}}_n(t) = (\widehat{H}_n(t), \widehat{M}_n^v(t), \widehat{M}_n^d(t)) : t \geq 0\}$, where the processes M_n^v and M_n^d are defined in (3.7). Then the following results hold:

- (i) $(\widehat{\mathbf{M}}_n(t), \mathcal{F}_t^n)$ is a mean zero martingale, where the filtration (\mathcal{F}_t^n) is defined in (3.3).
- (ii) For each $t \geq 0$, the quadratic variation processes have the following limits in probability:

- (a) $\lim_{n \rightarrow \infty} [\widehat{H}_n, \widehat{H}_n](t) = t$,
- (b) $\lim_{n \rightarrow \infty} [\widehat{M}_n^v, \widehat{M}_n^v](t) = \sigma_s^2 t$,
- (c) $\lim_{n \rightarrow \infty} [\widehat{M}_n^d, \widehat{M}_n^d](t) = 0$,
- (d) $\lim_{n \rightarrow \infty} [\widehat{H}_n, \widehat{M}_n^v](t) = \lim_{n \rightarrow \infty} [\widehat{H}_n, \widehat{M}_n^d](t) = \lim_{n \rightarrow \infty} [\widehat{M}_n^v, \widehat{M}_n^d](t) = 0$.

In part (b), σ_s^2 is given by $\sigma_s^2 = \mathbb{E}(v_1 - 1)^2$.

Proof. We already know \widehat{M}_n^v and \widehat{M}_n^d are (\mathcal{F}_t^n) -martingales from the discussion after (2.4) and (2.5). To prove part (i), it remains to show that \widehat{H}_n is also an (\mathcal{F}_t^n) -martingale. Since $\widehat{H}_n(\cdot)$ has piecewise constant paths with possible jumps at the times $\frac{k}{n}$, we consider

$$H_n(i) = \frac{1}{\sqrt{n}} \left[(i + 1) - n \int_0^{t_{i+1}^n} \lambda_n(V_n(s)) ds \right] \quad (4.45)$$

for $i = 0, 1, 2, \dots$. Notice that $\widehat{H}_n(t) = H_n([nt])$ for all $t \geq 0$ and H_n is adapted to the filtration $(\widehat{\mathcal{F}}_i^n)_{i \geq 0}$ defined in (3.2). We show that $(H_n(i), \widehat{\mathcal{F}}_i^n)$ is a martingale and from this, it follows that

$(\hat{H}_n(t), \mathcal{F}_t^n)$ also is a martingale. Following the discussion in (7.1) and (7.2), we intend to introduce two filtrations $(\mathcal{G}_t^n)_{t \geq 0}$ and $(\tilde{\mathcal{G}}_t^n)_{t \geq 0}$. Let $\mathring{\mathcal{F}}_0^n \equiv \{\emptyset, \Omega\}$ and

$$\mathring{\mathcal{F}}_j^n = \sigma((t_1^n, v_1^n, d_1^n), \dots, (t_j^n, v_j^n, d_j^n)) \quad \text{for } j \geq 1$$

as in (2.9). Then, it is easy to check that $A_n(t)$ is a $(\mathring{\mathcal{F}}_j^n)_{j \geq 0}$ -stopping time for each $t \geq 0$. Now we introduce the two filtrations $(\mathcal{G}_t^n)_{t \geq 0}$ and $(\tilde{\mathcal{G}}_t^n)_{t \geq 0}$ by

$$\mathcal{G}_t^n \equiv \sigma(A_n(s), V_n(s) : 0 \leq s \leq t), \quad \tilde{\mathcal{G}}_t^n \equiv \mathring{\mathcal{F}}_{A_n(t)}^n \quad \text{for all } t \geq 0. \quad (4.46)$$

For each i , the jump time t_i^n of the process $A_n(\cdot)$ is clearly a $(\tilde{\mathcal{G}}_t^n)$ -stopping time and $\mathbb{E}[t_i^n]$ is also finite as in (4.22). Thus the filtration $(\tilde{\mathcal{G}}_{t_i^n}^n)_{i \geq 1}$ is well defined. Since $\hat{A}_n(t)$ is a $(\tilde{\mathcal{G}}_t^n)$ -martingale as in (3.4), we have

$$\mathbb{E}[\hat{A}_n(t_{i+2}^n) | \tilde{\mathcal{G}}_{t_{i+1}^n}^n] = \hat{A}_n(t_{i+1}^n) \quad \text{for each } i = 0, 1, 2, \dots \quad (4.47)$$

Next, we observe that $\hat{A}_n(t_{i+1}^n) = H_n(i)$ and $\hat{\mathcal{F}}_i^n \subseteq \tilde{\mathcal{G}}_{t_{i+1}^n}^n$ for each $i = 0, 1, 2, \dots$. By conditioning both sides of (4.47) with respect to $\hat{\mathcal{F}}_i^n$, we obtain that $(H_n(i), \hat{\mathcal{F}}_i^n)$ is a martingale. This completes the proof of part (i).

For part (ii), first notice that \hat{H}_n can be written as

$$\hat{H}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=0}^{[nt]} \left(1 - \int_{t_j^n}^{t_{j+1}^n} n\lambda_n(V_n(s)) ds \right) \quad (4.48)$$

for all $t \geq 0$, where $t_0^n \equiv 0$. Recall that using (2.13), we can write $A_n(t) = Y_n(\int_0^t n\lambda_n(V_n(s)) ds)$ for all $t \geq 0$, where Y_n is a unit-rate Poisson process. Let $(e_j^n)_{j \geq 1}$ be the sequence of jump times of Y_n and define the sequence $(\eta_j^n)_{j \geq 1}$ by $\eta_1^n \equiv e_1^n$ and $\eta_j^n \equiv e_j^n - e_{j-1}^n$ for all $j \geq 2$. Then (η_j^n) is an i.i.d. sequence of exponential random variables with parameter 1. With the above representation, $\int_0^{t_j^n} n\lambda_n(V_n(s)) ds = e_j^n$ and hence \hat{H}_n can be written as

$$\hat{H}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=0}^{[nt]} (1 - \eta_j^n). \quad (4.49)$$

Therefore,

$$[\hat{H}_n, \hat{H}_n](t) = \frac{1}{n} \sum_{j=0}^{[nt]} (1 - \eta_j^n)^2. \quad (4.50)$$

Let $(\tilde{\eta}_j)$ be a generic i.i.d. sequence of exponential random variables with parameter 1. Then for each $\epsilon > 0$,

$$\mathbb{P} \left[\left| [\hat{H}_n, \hat{H}_n](t) - t \right| < \epsilon \right] = \mathbb{P} \left[\left| \frac{1}{n} \sum_{j=0}^{[nt]} (1 - \tilde{\eta}_j)^2 - t \right| < \epsilon \right]$$

and by strong law of large numbers, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{[nt]} (1 - \tilde{\eta}_j)^2 = t$ a.s. Consequently, for each $t \geq 0$, $\lim_{n \rightarrow \infty} [\hat{H}_n, \hat{H}_n](t) = t$ in probability.

Next, we consider the quadratic variation process $[\widehat{M}_n^v, \widehat{M}_n^v](t)$. Using (3.7), we obtain

$$[\widehat{M}_n^v, \widehat{M}_n^v](t) = \frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1)^2 \mathbf{1}_{[V_n(t_j^-) < d_j^n]}.$$

Let $S_n(t) = \frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1)^2$. Then

$$|S_n(t) - [\widehat{M}_n^v, \widehat{M}_n^v](t)| = \frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1)^2 \mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}.$$

Since v_j is independent of $\sigma(\widehat{\mathcal{F}}_{j-1}^n \cup \{d_j^n\})$ and $\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}$ is measurable with respect to this σ -algebra, we have

$$\mathbb{E} \left[(v_j - 1)^2 \mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} \middle| \sigma(\widehat{\mathcal{F}}_{j-1}^n \cup \{d_j^n\}) \right] = \sigma_s^2 \mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}.$$

Taking the expected value in both sides, we have $\mathbb{E}[(v_j - 1)^2 \mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}] = \sigma_s^2 \mathbb{E}[\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}]$. Consequently,

$$\mathbb{E}|S_n(t) - [\widehat{M}_n^v, \widehat{M}_n^v](t)| = \frac{\sigma_s^2}{n} \mathbb{E} \left[\sum_{j=1}^{[nt]} \mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} \right] = \sigma_s^2 \mathbb{E}[\bar{R}_n(t)],$$

where $\bar{R}_n(t)$ is given in (4.20). By Lemma 4.4, we have $\lim_{n \rightarrow \infty} \mathbb{E}[\bar{R}_n(t)] = 0$ and thus

$\lim_{n \rightarrow \infty} \mathbb{E}|S_n(t) - [\widehat{M}_n^v, \widehat{M}_n^v](t)| = 0$. On the other hand, (v_j) is an i.i.d. sequence with $\mathbb{E}(v_j - 1)^2 = \sigma_s^2$. Therefore, by strong law of large numbers, $\lim_{n \rightarrow \infty} S_n(t) = \sigma_s^2 t$ a.s. Using these two facts, we can conclude $\lim_{n \rightarrow \infty} [\widehat{M}_n^v, \widehat{M}_n^v](t) = \sigma_s^2 t$ in probability for each $t > 0$.

Using (3.7), we have

$$\mathbb{E}([\widehat{M}_n^d, \widehat{M}_n^d](t)) = \frac{1}{n} \mathbb{E} \sum_{j=1}^{[nt]} \left(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} - \mathbb{E}(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} | \widehat{\mathcal{F}}_{j-1}^n) \right)^2.$$

Since $\mathbb{E}(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} | \widehat{\mathcal{F}}_{j-1}^n) \leq 1$, we obtain

$$\mathbb{E} \left(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} - \mathbb{E}(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} | \widehat{\mathcal{F}}_{j-1}^n) \right)^2 \leq 2 \mathbb{E}[\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]}].$$

Therefore, $\mathbb{E}([\widehat{M}_n^d, \widehat{M}_n^d](t)) \leq 2 \mathbb{E}[\bar{R}_n(t)]$, where $\bar{R}_n(t)$ is given in (4.20). Using (4.21), we have $\lim_{n \rightarrow \infty} \mathbb{E}([\widehat{M}_n^d, \widehat{M}_n^d](t)) = 0$ and thus $\lim_{n \rightarrow \infty} [\widehat{M}_n^d, \widehat{M}_n^d](t) = 0$ in probability for each $t > 0$.

Similar to the above computations, we have

$$[\widehat{M}_n^v, \widehat{M}_n^d](t) = -\frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1) \mathbf{1}_{[V_n(t_j^-) < d_j^n]} \mathbb{E} \left(\mathbf{1}_{[V_n(t_j^-) \geq d_j^n]} | \widehat{\mathcal{F}}_{j-1}^n \right).$$

But $V_n(t_j^-)$ and d_j^n are measurable in $\sigma(\widehat{\mathcal{F}}_{j-1}^n \cup \{d_j^n\})$ and v_j is independent of $\sigma(\widehat{\mathcal{F}}_{j-1}^n \cup \{d_j^n\})$. Also, $\mathbb{E}|v_j - 1| \leq \sqrt{\mathbb{E}(v_j - 1)^2} = \sigma_s$. Hence we can easily obtain

$$\mathbb{E}|[\widehat{M}_n^v, \widehat{M}_n^d](t)| \leq \sigma_s \mathbb{E}[\bar{R}_n(t)] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

by (4.21). Thus, $\lim_{n \rightarrow \infty} [\widehat{M}_n^v, \widehat{M}_n^d](t) = 0$ in probability for each $t > 0$.

From (4.48) and (3.7), we obtain

$$[\widehat{H}_n, \widehat{M}_n^v](t) = \frac{1}{n} \sum_{j=1}^{[nt]} (v_j - 1) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]} \left(1 - \int_{t_j^n}^{t_{j+1}^n} n \lambda_n(V_n(s)) ds \right). \quad (4.51)$$

Let $U_n(t) = [\widehat{H}_n, \widehat{M}_n^v](t)$. We claim that $(U_n(t), \mathcal{F}_t^n)$ is a martingale. Clearly, $\{U_n(t)\}$ is adapted to (\mathcal{F}_t^n) . Using the notation in (4.49), we can write

$$(v_j - 1) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]} \left(1 - \int_{t_j^n}^{t_{j+1}^n} n \lambda_n(V_n(s)) ds \right) = (v_j - 1)(1 - \eta_j^n) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]}.$$

This term is integrable since $\mathbb{E}(v_j - 1)^2 = \sigma_s^2 < \infty$ and $\mathbb{E}(1 - \eta_j^n)^2 = 1$. This term is also equal to $\sqrt{n}(v_j - 1) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]} (\widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t_j^n))$. Using the fact that $v_j, V_n(t_j^n-)$ and d_j^n are $\widetilde{\mathcal{G}}_{t_j^n}^n$ -measurable and by (4.47), we see that

$$\mathbb{E} \left[(v_j - 1) \mathbf{1}_{[V_n(t_j^n-) < d_j^n]} \left(1 - \int_{t_j^n}^{t_{j+1}^n} n \lambda_n(V_n(s)) ds \right) \middle| \widetilde{\mathcal{G}}_{t_j^n}^n \right] = 0.$$

But $\widehat{\mathcal{F}}_{j-1}^n \subseteq \widetilde{\mathcal{G}}_{t_j^n}^n$ and therefore by conditioning on $\widehat{\mathcal{F}}_{j-1}^n$, we have $\{U_n(t)\}$ is an (\mathcal{F}_t^n) -martingale. Consequently,

$$\mathbb{E}([U_n, U_n](t)) \leq \frac{1}{n} \sum_{j=1}^{[nt]} \mathbb{E} \left[(v_j - 1)^2 (\widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t_j^n))^2 \right].$$

Since $(\widehat{A}_n(t), \widetilde{\mathcal{G}}_t^n)$ is a martingale (recall (3.4)) and the quadratic variation process is given by $[\widehat{A}_n, \widehat{A}_n](t) = \frac{1}{n} A_n(t)$, we have

$$\mathbb{E} \left[(\widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t_j^n))^2 \middle| \widetilde{\mathcal{G}}_{t_j^n}^n \right] = \frac{1}{n}.$$

Also, $(v_j - 1)$ is $\widetilde{\mathcal{G}}_{t_j^n}^n$ -measurable, and hence

$$\mathbb{E} \left[(v_j - 1)^2 (\widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t_j^n))^2 \middle| \widetilde{\mathcal{G}}_{t_j^n}^n \right] = \frac{1}{n} (v_j - 1)^2.$$

Consequently, $\mathbb{E}[(v_j - 1)^2 (\widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t_j^n))^2] = \sigma_s^2/n$ and we deduce that

$$\mathbb{E}([U_n, U_n](t)) \leq \frac{\sigma_s^2 [nt]}{n^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore, $U_n(t) = [\widehat{H}_n, \widehat{M}_n^v](t) \rightarrow 0$ in probability. The proof of $\lim_{n \rightarrow \infty} [\widehat{H}_n, \widehat{M}_n^d](t) = 0$ in probability is similar to that of the previous result and therefore we omit it. This completes the proof of part (ii) of lemma. \blacksquare

Proposition 4.12. *Let ξ_n be defined by (4.37). Then the process $\xi_n(\cdot)$ converges weakly to $\sigma W(\cdot)$ in $D[0, \infty)$ as $n \rightarrow \infty$, where $W(\cdot)$ is a standard Brownian motion and $\sigma > 0$ is a constant given by $\sigma^2 = 1 + \sigma_s^2$. Here, $\sigma_s^2 = \mathbb{E}(v_1 - 1)^2$ is a constant as in Assumption 3.1.*

Proof. We consider the vector-valued process $\{(\widehat{A}_n(t), \widehat{M}_n^v(\bar{A}_n(t)), \widehat{M}_n^d(\bar{A}_n(t))) : t \geq 0\}$, where $\bar{A}_n(t) = \frac{1}{n}A_n(t)$ for all $t \geq 0$. We intend to show that this process converges weakly to $(W_1, \sigma_s W_2, 0)$ in $D^{\otimes 3}[0, \infty)$, where W_1 and W_2 are independent standard Brownian motions.

Consider process \widehat{H}_n defined in (4.44). Then

$$\widehat{H}_n(\bar{A}_n(t)) = \widehat{A}_n(t_{k+1}^n) \quad \text{if } t_k^n \leq t < t_{k+1}^n. \quad (4.52)$$

Notice that the vector-valued process $\widehat{\mathbf{M}}_n(t) = (\widehat{H}_n(t), \widehat{M}_n^v(t), \widehat{M}_n^d(t))$ for $t \geq 0$ is an (\mathcal{F}_t^n) -martingale by part (i) of Lemma 4.11. Our approach here is to use the martingale functional central limit theorem (cf. Theorem 1.4, Chapter 7 in [12] or Theorem 2.1 in [34]) to establish the weak convergence of $\widehat{\mathbf{M}}_n$ to $(W_1, \sigma_s W_2, 0)$ and then to apply random time change theorem (cf. Section 14 of [6]) to conclude $\widehat{\mathbf{M}}_n(\bar{A}_n(t))$ also converges to $(W_1, \sigma_s W_2, 0)$. Finally, we establish that for each $T > 0$, $\sup_{t \in [0, T]} |\widehat{A}_n(t) - \widehat{H}(\bar{A}_n(t))|$ converges to zero in probability. Then as a consequence

of this, $(\widehat{A}_n(\cdot), \widehat{M}_n^v(\bar{A}_n(\cdot)), \widehat{M}_n^d(\bar{A}_n(\cdot)))$ converges weakly to $(W_1, \sigma_s W_2, 0)$ in $D^{\otimes 3}[0, \infty)$.

To implement the sketch of the proof given above, we consider the vector-valued martingale $(\widehat{\mathbf{M}}_n(t), \mathcal{F}_t^n)$ and apply the martingale functional central limit theorem, Theorem 1.4 of Chapter 7 in [12]. We intend to verify the assumption in the quoted Theorem 1.4, part a). First, we show that for each $T > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sup_{t \in [0, T]} |\widehat{\mathbf{M}}_n(t) - \widehat{\mathbf{M}}_n(t-)| \right) = 0.$$

Using the representation (4.49) for \widehat{H}_n , we can write

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\widehat{H}_n(t) - \widehat{H}_n(t-)| \right] = \mathbb{E} \left(\frac{1}{\sqrt{n}} \max_{1 \leq j \leq nT} |1 - \eta_j^n| \right) \leq \left[\frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |1 - \eta_j^n|^2 \right) \right]^{1/2},$$

where (η_j^n) is an i.i.d. sequence of exponentially distributed random variables with parameter 1. If $(\tilde{\eta}_j)$ is a generic sequence of i.i.d. exponentially distributed random variables with parameter 1., then $\frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |1 - \eta_j^n|^2 \right) = \frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |1 - \tilde{\eta}_j|^2 \right)$ and since $\mathbb{E}(1 - \tilde{\eta}_j)^2 = 1$, by (7.5) (see the Appendix), we have $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |1 - \tilde{\eta}_j|^2 \right) = 0$. Hence $\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} |\widehat{H}_n(t) - \widehat{H}_n(t-)| \right] = 0$.

Similarly,

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\widehat{M}_n^v(t) - \widehat{M}_n^v(t-)| \right] \leq \mathbb{E} \left(\frac{1}{\sqrt{n}} \max_{1 \leq j \leq nT} |v_j - 1| \right) \leq \left[\frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |v_j - 1|^2 \right) \right]^{1/2}.$$

Since (v_j) is i.i.d. and $\mathbb{E}(v_j - 1)^2 = \sigma_s^2 < \infty$, again by (7.5), $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left(\max_{1 \leq j \leq nT} |v_j - 1|^2 \right) = 0$.

For $n \geq 1$, let $q_n = (q_n^{ij}(t) : t \geq 0)_{1 \leq i, j \leq 3}$ be the symmetric 3×3 matrix-valued process such that q_n^{ij} represents the (i, j) -th quadratic-covariation process of the martingale $\widehat{\mathbf{M}}_n = (\widehat{H}_n, \widehat{M}_n^v, \widehat{M}_n^d)$ (for example, $q_n^{12}(t) \equiv [\widehat{H}_n, \widehat{M}_n^v](t)$ for $t \geq 0$). In part (ii) of Lemma 4.11, we have established that $\lim_{n \rightarrow \infty} q_n^{ij}(t) = c_{ij}t$ in probability for $1 \leq i, j \leq 3$ and the constant matrix $C = (c_{ij})_{3 \times 3}$ is described by the diagonal matrix $C = \text{diag}(1, \sigma_s^2, 0)$ (see Remark 1.5 in page 340 of [12]).

Hence, the assumptions of the martingale functional central limit theorem, Theorem 1.4, part a) in pages 339–340 of [12] are satisfied. Thus, we can conclude that $\widehat{\mathbf{M}}_n$ converges weakly to $(W_1, \sigma_s W_2, 0)$ as $n \rightarrow \infty$, where W_1 and W_2 are independent standard Brownian motions. By

(4.6), $\sup_{t \in [0, T]} |\bar{A}_n(t) - t| \rightarrow 0$ as $n \rightarrow \infty$ for each $T > 0$, and hence by the random time change

theorem (Section 14, [6]), $\widehat{\mathbf{M}}_n \circ \bar{A}_n$ also converges weakly to $(W_1, \sigma_s W_2, 0)$ in $D^{\otimes 3}[0, \infty)$ as $n \rightarrow \infty$.

Now, to establish the weak convergence of the process $(\widehat{A}_n(t), \widehat{M}_n^v(\bar{A}_n(t)), \widehat{M}_n^d(\bar{A}_n(t)))$ it remains to estimate $\sup_{t \in [0, T]} |\widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t)|$ for each $T > 0$. Let $\epsilon > 0$. Notice that

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t) \right| > \epsilon \right] &\leq \mathbb{P} \left[\sup_{t \in [0, T]} \left| \widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t) \right| > \epsilon, \bar{A}_n(T) \leq 2T \right] \\ &\quad + \mathbb{P} [\bar{A}_n(T) > 2T]. \end{aligned} \quad (4.53)$$

On the set $[\bar{A}_n(T) \leq 2T]$, using (4.52), we have

$$\sup_{t \in [0, T]} \left| \widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t) \right| = \sup_{0 \leq j \leq [2nT]} \sup_{t_j^n \leq t < t_{j+1}^n} \left| \widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t) \right|.$$

But using the sequence (η_j^n) of i.i.d. exponentially distributed random variables with parameter 1 introduced in the discussion above (4.49), we have

$$\sup_{t_j^n \leq t < t_{j+1}^n} \left| \widehat{A}_n(t_{j+1}^n) - \widehat{A}_n(t) \right| \leq \frac{1}{\sqrt{n}} \left(1 + \int_{t_j^n}^{t_{j+1}^n} n \lambda_n(V_n(s)) ds \right) = \frac{1}{\sqrt{n}} (1 + \eta_{j+1}^n).$$

Therefore,

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t) \right| > \epsilon, \bar{A}_n(T) \leq 2T \right] &\leq \mathbb{P} \left[\sup_{0 \leq j \leq [2nT]} \frac{1}{\sqrt{n}} (1 + \eta_{j+1}^n) > \epsilon \right] \\ &= \mathbb{P} \left[\sup_{0 \leq j \leq [2nT]} \frac{1}{\sqrt{n}} (1 + \tilde{\eta}_{j+1}) > \epsilon \right] \end{aligned} \quad (4.54)$$

where $(\tilde{\eta}_j)$ is a generic sequence of i.i.d. exponentially distributed random variables with parameter

1. Since $\mathbb{E}(1 + \tilde{\eta}_{j+1})^2 < \infty$, by (7.5) (see the Appendix), $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\sup_{0 \leq j \leq [2nT]} (1 + \tilde{\eta}_{j+1})^2 \right] = 0$.

Hence, the right hand side of (4.54) tends to zero and consequently, the first term on the right side of (4.53) converges to zero as n tends to infinity. On the other hand, $\bar{A}_n(T)$ converges to T almost surely, and hence the second term on the right side of (4.53) also converges to zero as $n \rightarrow \infty$. Using these two limits in (4.53), we obtain

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \widehat{H}_n(\bar{A}_n(t)) - \widehat{A}_n(t) \right| > \epsilon \right] = 0 \quad \text{for each } T > 0.$$

We can combine this result with the already established weak convergence of

$(\widehat{H}_n(\bar{A}_n(t)), \widehat{M}_n^v(\bar{A}_n(t)), \widehat{M}_n^d(\bar{A}_n(t)))$ to $(W_1, \sigma_s W_2, 0)$ as $n \rightarrow \infty$ to obtain that the process $(\widehat{A}_n(t), \widehat{M}_n^v(\bar{A}_n(t)), \widehat{M}_n^d(\bar{A}_n(t)))$ converges weakly to $(W_1, \sigma_s W_2, 0)$ in $D^{\otimes 3}[0, \infty)$ as $n \rightarrow \infty$. Therefore, the process $\xi_n(\cdot)$ defined in (4.37) converges weakly to $\sigma W(\cdot)$ in $D[0, \infty)$ as $n \rightarrow \infty$, where $W(\cdot)$ is a standard one-dimensional Brownian motion and $\sigma^2 = 1 + \sigma_s^2$. This completes the proof. ■

5 Scaled Queue Length

Here we establish an asymptotic relationship between the queue-length and offered waiting time processes under heavy traffic conditions. For a conventional $GI/GI/1$ queue without abandonment, this asymptotic relationship was established in Theorem 4 of Section 3 in Reiman [27]. We essentially follow the proof of this fact in Reed and Ward [26] (Theorem 6.1) and supplement it with necessary estimates to accommodate our general assumptions. In contrast with the proof of Reed and Ward [26], we rely on the Assumption 3.3 for the sequence (F_n) of patience-time distribution functions to establish Lemma 5.2. In addition, our arrival intensity $n\lambda_n(\cdot)$ is state-dependent and hence we need Lemma 5.3 to facilitate our proof.

For $t \geq 0$, let $Q_n(t)$ be the queue length of the n -th system at time t and $\widehat{Q}_n(t) = \frac{Q_n(t)}{\sqrt{n}}$ be the diffusion-scaled queue length. Following the notation in [26], we also introduce the random variable

$$a_n(t) \equiv \text{the arrival time of the customer in service at time } t \text{ in the } n\text{-th system.}$$

If the server is idle at time t , we let $a_n(t) = t$.

Theorem 5.1. *Let \widehat{Q}_n and \widehat{V}_n be scaled queue-length and scaled offered waiting time processes, respectively. Then as $n \rightarrow \infty$,*

$$\widehat{Q}_n - \widehat{V}_n \Rightarrow 0 \quad \text{in } D[0, \infty).$$

To prove this theorem, we follow the discussion in page 21 of [26] with appropriate changes and then establish two lemmas. Recall that for the j -th arrival in the n -th system, service time is v_j/n . First, notice that $V_n(a_n(t)-) \leq t - a_n(t) \leq V_n(a_n(t)-) + \frac{1}{n}v_{A_n(a_n(t))}$ and hence

$$\widehat{V}_n(a_n(t)-) \leq \sqrt{n}(t - a_n(t)) \leq \widehat{V}_n(a_n(t)-) + \frac{1}{\sqrt{n}}v_{A_n(a_n(t))} \quad (5.1)$$

for all $t \geq 0$. For $T \geq 0$, let $u_n(T) = \max\{v_j : 1 \leq j \leq A_n(T)\}$. Then we observe that for each $T \geq 0$, $u_n(T)/\sqrt{n} \Rightarrow 0$ as $n \rightarrow \infty$. Indeed, for an arbitrary $\epsilon > 0$,

$$\mathbb{P} \left[\frac{1}{\sqrt{n}}u_n(T) > \epsilon \right] \leq \mathbb{P} \left[\frac{1}{\sqrt{n}}u_n(T) > \epsilon, \bar{A}_n(T) < 2T \right] + \mathbb{P}[\bar{A}_n(T) \geq 2T].$$

We know from (4.6) that $\lim_{n \rightarrow \infty} \mathbb{P}[\bar{A}_n(T) \geq 2T] = 0$. Also, note that

$$\mathbb{P} \left[\frac{1}{\sqrt{n}}u_n(T) > \epsilon, \bar{A}_n(T) < 2T \right] \leq \mathbb{P} \left[\frac{1}{\sqrt{n}} \max_{1 \leq j \leq 2nT} v_j > \epsilon \right]$$

and $\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{\sqrt{n}} \max_{1 \leq j \leq 2nT} v_j > \epsilon \right] = 0$ follows from (7.5) in the Appendix or Lemma 3.3 of [17].

Thus, we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{\sqrt{n}}u_n(T) > \epsilon \right] = 0. \quad (5.2)$$

Notice that for $n \geq 1$ and $0 \leq t \leq T$, we have $0 \leq v_{A_n(a_n(t))} \leq u_n(T)$ and this together with (5.2) implies that

$$\sup_{t \in [0, T]} |\sqrt{n}(t - a_n(t)) - \widehat{V}_n(a_n(t)-)| \Rightarrow 0$$

as $n \rightarrow \infty$. Dividing by \sqrt{n} , and using Theorem 4.1, we deduce that

$$\sup_{t \in [0, T]} (t - a_n(t)) \Rightarrow 0 \quad (5.3)$$

as $n \rightarrow \infty$.

The next two technical lemmas enable us to prove Theorem 5.1 and we use the above facts in their proofs. Our Lemma 5.2 corresponds to Lemma 6.1 of [26], but in our proof, we use the Assumption 3.3 for the sequence (F_n) to obtain the weak convergence result

$$\frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(\cdot))}^{A_n(\cdot)} F_n(V_n(t_j^n -)) \Rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

To accommodate our state-dependent arrival intensities, we need an additional technical result given in Lemma 5.3.

Lemma 5.2. *For each $T > 0$,*

$$\sup_{t \in [0, T]} \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n -) \geq d_j^n]} \Rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Proof. We begin with the following identity: for $t \geq 0$

$$\frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n -) \geq d_j^n]} = \widehat{M}_n^d(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(a_n(t))) + \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} F_n(V_n(t_j^n -)), \quad (5.4)$$

where $\widehat{M}_n^d(t)$ is described in (3.7) (see also (2.6)). By Proposition 4.12, $\widehat{M}_n^d \Rightarrow 0$ as $n \rightarrow \infty$ and by (4.6), $\sup_{t \in [0, T]} |\bar{A}_n(t) - t| \rightarrow 0$ as $n \rightarrow \infty$. Using these facts together with (5.3) and then applying random-time change theorem in [6], we can conclude

$$\widehat{M}_n^d(\bar{A}_n(\cdot)) - \widehat{M}_n^d \circ \bar{A}_n \circ a_n(\cdot) \Rightarrow 0 \quad (5.5)$$

in $D[0, \infty)$ as $n \rightarrow \infty$. Next, we show

$$\frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(\cdot))}^{A_n(\cdot)} F_n(V_n(t_j^n -)) \Rightarrow 0$$

as $n \rightarrow \infty$ and hence by (5.4) this will imply the stated result. For $t \geq 0$, let

$$\Upsilon_n(t) \equiv \frac{1}{\sqrt{n}} \sup_{t \in [0, T]} \sum_{j=A_n(a_n(t))}^{A_n(t)} F_n(V_n(t_j^n -))$$

and $\epsilon, \tilde{\epsilon} > 0$ be arbitrary. Choose $K > 0$ large enough then, by (4.23), there is $n_0 \in \mathbb{N}$ so that $\mathbb{P} \left[\max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \geq K \right] < \tilde{\epsilon}$ for all $n \geq n_0$. Using the fact that $F_n(\cdot)$ is non-decreasing, we obtain

$$\begin{aligned} \mathbb{P}[\Upsilon_n(T) > \epsilon] &\leq \mathbb{P} \left[\Upsilon_n(T) > \epsilon, \max_{1 \leq j \leq [nT]} \widehat{V}_n(t_j^n -) \leq K \right] + \tilde{\epsilon} \\ &\leq \mathbb{P} \left[\frac{1}{\sqrt{n}} F_n \left(\frac{K}{\sqrt{n}} \right) \sup_{t \in [0, T]} (A_n(t) - A_n(a_n(t))) > \epsilon \right] + \tilde{\epsilon}, \end{aligned} \quad (5.6)$$

for all $n \geq n_0$. We take $\delta_1 > 0$. Then by the Assumption 3.3, there is $n_1 \in \mathbb{N}$ such that $\sqrt{n}F_n(\frac{K}{\sqrt{n}}) \leq H(K) + \delta_1$ for all $n \geq n_1$. We let the constant $C_1 \equiv H(K) + \delta_1 > 0$. Then for all $n \geq \max\{n_0, n_1\}$, $\frac{1}{\sqrt{n}}F_n(K/\sqrt{n}) \leq C_1/n$ and thus

$$\mathbb{P} \left[\frac{1}{\sqrt{n}}F_n \left(\frac{K}{\sqrt{n}} \right) \sup_{t \in [0, T]} (A_n(t) - A_n(a_n(t))) > \epsilon \right] \leq \mathbb{P} \left[\sup_{t \in [0, T]} (\bar{A}_n(t) - \bar{A}_n(a_n(t))) > \frac{\epsilon}{C_1} \right]. \quad (5.7)$$

But since $a_n(t) \leq t$, we have

$$\sup_{t \in [0, T]} [\bar{A}_n(t) - \bar{A}_n(a_n(t))] \leq 2 \sup_{t \in [0, T]} |\bar{A}_n(t) - t| + \sup_{t \in [0, T]} |t - a_n(t)|.$$

Hence by (4.6) and (5.3), it follows that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{\sqrt{n}}F_n \left(\frac{K}{\sqrt{n}} \right) \sup_{t \in [0, T]} (A_n(t) - A_n(a_n(t))) > \epsilon \right] = 0$$

and using this in (5.6), we have $\lim_{n \rightarrow \infty} \mathbb{P}[\Upsilon_n(T) > \epsilon] = 0$. Using this together with (5.5) in the identity (5.4), we obtain the desired conclusion. \blacksquare

Lemma 5.3. *Let $T \geq 0$. Then the following weak convergence result holds:*

$$\sqrt{n} \sup_{t \in [0, T]} \int_{a_n(t)}^t \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right| ds \Rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (5.8)$$

Proof. Let $\epsilon > 0$ be arbitrary. We pick $K > 0$ large enough then, by (4.14), there is $n_0 \in \mathbb{N}$ so that $\mathbb{P}[\|\widehat{V}_n\|_T > K] < \epsilon/2$ for all $n \geq n_0$. Let $\delta > 0$. Using part (iv) of Assumption 3.2, there is $n_1 \in \mathbb{N}$ so that

$$\sqrt{n} \sup_{x \in [0, K]} |\lambda(x/\sqrt{n}) - 1| \leq \max_{x \in [0, K]} u(x) + \delta \quad \text{for all } n \geq n_1.$$

We let $C = \max_{x \in [0, K]} u(x) + \delta > 0$ and for $n \geq 1$, introduce

$$\mathcal{W}_n(T) = \sqrt{n} \sup_{t \in [0, T]} \int_{a_n(t)}^t \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right| ds.$$

Then, we have

$$\begin{aligned} \mathbb{P}[\mathcal{W}_n(T) > \epsilon] &\leq \mathbb{P} \left[\mathcal{W}_n(T) > \epsilon, \|\widehat{V}_n\|_T \leq K \right] + \frac{\epsilon}{2} \\ &\leq \mathbb{P} \left[C \sup_{t \in [0, T]} (t - a_n(t)) > \epsilon \right] + \frac{\epsilon}{2}, \end{aligned}$$

for all $n \geq \max\{n_0, n_1\}$. Using this together with (5.3), we obtain $\lim_{n \rightarrow \infty} \mathbb{P}[\mathcal{W}_n(T) > \epsilon] = 0$ and this yields (5.8). \blacksquare

Next, we use Lemmas 5.2 and 5.3 to prove Theorem 5.1.

Proof of Theorem 5.1. We begin with the estimate

$$A_n(t) - A_n(a_n(t)) - \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} \leq Q_n(t) \leq A_n(t) - A_n(a_n(t)) + 1$$

as explained in the proof of Theorem 6.1 in [26]. For $n \geq 1$ and $t \geq 0$, let

$$\mathcal{Y}_n(t) = \widehat{A}_n(t) - \widehat{A}_n(a_n(t)) + \sqrt{n} \int_{a_n(t)}^t \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) ds,$$

where $\widehat{A}_n(\cdot)$ is described in (3.5). Then,

$$\mathcal{Y}_n(t) - \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} \leq \widehat{Q}_n(t) \leq \mathcal{Y}_n(t) + \frac{1}{\sqrt{n}}$$

for all $t \geq 0$. Hence we can write

$$\begin{aligned} \mathcal{Y}_n(t) - \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} - \widehat{V}_n(a_n(t)) + [\widehat{V}_n(a_n(t)) - \widehat{V}_n(t)] \\ \leq \widehat{Q}_n(t) - \widehat{V}_n(t) \leq \mathcal{Y}_n(t) + \frac{1}{\sqrt{n}} - \widehat{V}_n(a_n(t)) + [\widehat{V}_n(a_n(t)) - \widehat{V}_n(t)]. \end{aligned}$$

Next, let $\mathcal{Z}_n(t) = \widehat{A}_n(t) - \widehat{A}_n(a_n(t)) + \sqrt{n} \int_{a_n(t)}^t \left[\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right] ds$, for all $t \geq 0$. Then we can employ the estimates for $\widehat{V}_n(a_n(t))$ in (5.1) and obtain,

$$\begin{aligned} \mathcal{Z}_n(t) - \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} + [\widehat{V}_n(a_n(t)) - \widehat{V}_n(t)] \\ \leq \widehat{Q}_n(t) - \widehat{V}_n(t) \leq \mathcal{Z}_n(t) + \frac{1}{\sqrt{n}} v_{A_n(a_n(t))} + [\widehat{V}_n(a_n(t)) - \widehat{V}_n(t)] + \frac{1}{\sqrt{n}}. \end{aligned}$$

Consequently,

$$\begin{aligned} |\widehat{Q}_n(t) - \widehat{V}_n(t)| &\leq |\mathcal{Z}_n(t)| + \frac{1}{\sqrt{n}} \sum_{j=A_n(a_n(t))}^{A_n(t)} \mathbf{1}_{[V_n(t_j^n-) \geq d_j^n]} \\ &\quad + |\widehat{V}_n(a_n(t)) - \widehat{V}_n(t)| + \frac{1}{\sqrt{n}} v_{A_n(a_n(t))} + \frac{1}{\sqrt{n}}. \end{aligned} \quad (5.9)$$

Since $\widehat{A}_n \Rightarrow W_1$ as $n \rightarrow \infty$, and by (5.3), we have $|\widehat{A}_n(\cdot) - \widehat{A}_n \circ a_n(\cdot)| \Rightarrow 0$ as $n \rightarrow \infty$. We use this fact together with Lemma 5.3 to conclude $|\mathcal{Z}_n(\cdot)| \Rightarrow 0$ as $n \rightarrow \infty$.

Similarly, $\widehat{V}_n(\cdot)$ converges weakly as in Theorem 4.10. This together with (5.3) yields $|\widehat{V}_n \circ a_n(\cdot) - \widehat{V}_n(\cdot)| \Rightarrow 0$ as $n \rightarrow \infty$. Finally, notice that $0 \leq \frac{1}{\sqrt{n}} v_{A_n(a_n(t))} \leq \frac{u_n(T)}{\sqrt{n}}$, where $u_n(T)$ is as in (5.2), and hence by (5.2), we deduce that $v_{A_n(a_n(\cdot))}/\sqrt{n} \Rightarrow 0$ as $n \rightarrow \infty$. Using all these facts in (5.9), we are able to conclude $\widehat{Q}_n(\cdot) - \widehat{V}_n(\cdot) \Rightarrow 0$ as $n \rightarrow \infty$. This completes the proof. \blacksquare

As a consequence of Theorem 4.10 and the convergence together theorem (Theorem 5.1), we have the following corollary.

Corollary 5.4. *The scaled queue-length process $(\widehat{Q}_n(t))_{t \geq 0}$ also converges weakly as $n \rightarrow \infty$ to the diffusion process $(Z(t))_{t \geq 0}$ of (4.42) in $D[0, \infty)$.*

6 Convergence of Cost Functionals

6.1 Introduction

Here we introduce an infinite horizon discounted cost functional associated with the n -th system described in (3.8). Our goal is to show that the expected value of this cost functional converges to the expected value of the same cost functional associated with the limiting diffusion process described in (4.42). For heavy traffic limits related to scaled queue-length processes, such convergence of cost functionals are obtained in [5, 32, 19] and they are very useful in controlled queueing systems to obtain an asymptotically optimal arrival rate $\lambda_n(\cdot)$. First we introduce the scaled idle time process $\widehat{L}_n(\cdot)$ associated with (3.8) by

$$\widehat{L}_n(t) = \sqrt{n}I_n(t) \quad \text{for all } t \geq 0. \quad (6.1)$$

Then, after scaling we can rewrite (3.8) as

$$\begin{aligned} \widehat{V}_n(t) + \frac{1}{\sqrt{n}} \int_0^t F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) &= \widehat{A}_n(t) + \widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) \\ &+ \sqrt{n} \int_0^t \left[\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right] ds + \widehat{L}_n(t), \end{aligned} \quad (6.2)$$

for all $t \geq 0$.

Let $\gamma > 0$ be a discount factor and $C(\cdot)$ be a running cost function of polynomial growth. For the n -th system described in (6.2), we introduce two types of costs: A cost of $\int_0^\infty e^{-\gamma t} C(\widehat{V}_n(t)) dt$ related to the waiting times and an idleness cost proportional to $\int_0^\infty e^{-\gamma t} d\widehat{L}_n(t)$. This idleness cost can be considered as a penalty for an empty workload in the system. Thus the infinite horizon discounted cost functional associated with the n -th system is given by

$$J(\widehat{V}_n, \widehat{L}_n) \equiv \mathbb{E} \int_0^\infty e^{-\gamma t} \left[C(\widehat{V}_n(t)) dt + p \cdot d\widehat{L}_n(t) \right], \quad (6.3)$$

where $p > 0$ and $\gamma > 0$ are fixed constants. The cost functional related to the limiting diffusion in (4.42) is given by

$$J(Z, L) \equiv \mathbb{E} \int_0^\infty e^{-\gamma t} [C(Z(t)) dt + p \cdot dL(t)]. \quad (6.4)$$

To motivate the general term $\mathbb{E}[\int_0^\infty e^{-\gamma t} C(\widehat{V}_n(t)) dt]$ in our cost structure (6.3), first we consider a cost functional of the form $\sum_{j=1}^\infty e^{-\gamma t_j^n} C_n(V_n(t_j^n -))$ where $\gamma > 0$ is a discount factor and $C_n(V_n(t_j^n -))$ represents the waiting cost of the j -th customer. Here $C_n(\cdot)$ is a non-negative cost function and $C_n(0) = 0$. We impose two conditions on $C_n(\cdot)$:

- (a) $0 \leq nC_n(\frac{x}{\sqrt{n}}) \leq K_1(1 + x^\ell)$ for all $x \geq 0$, where the $K_1 > 0$ and $\ell \geq 1$ are constants independent of n as in the assumption (6.5) below, and
- (b) there exists a non-negative function $C(\cdot)$ such that $\lim_{n \rightarrow \infty} \sup_{x \in [0, K]} |nC_n(\frac{x}{\sqrt{n}}) - C(x)| = 0$ for each $K > 0$.

For a concrete example, one may take $C_n(x) = ax^2 + a_3x^3 + \dots + a_mx^m$ and $C(x) = ax^2$, where $m \leq \ell$ and $a > 0, a_3, \dots, a_m$ are non-negative constants, $1 \leq m \leq \ell$ are integers and ℓ satisfies (6.6)

below. Next, notice that

$$\sum_{j=1}^{\infty} e^{-\gamma t_j^n} C_n(V_n(t_j^n-)) = \int_0^{\infty} e^{-\gamma t} C_n(V_n(t-)) dA_n(t)$$

and hence, by the monotone convergence theorem

$$\mathbb{E} \left[\sum_{j=1}^{\infty} e^{-\gamma t_j^n} C_n(V_n(t_j^n-)) \right] = \lim_{T \rightarrow \infty} \mathbb{E} \left[\int_0^T e^{-\gamma t} C_n(V_n(t-)) dA_n(t) \right].$$

By (3.4), we can write $A_n(t) = M_n(t) + n \int_0^t \lambda_n(V_n(s)) ds$, where $(M_n(t), \mathcal{G}_t^n)_{t \geq 0}$ is a martingale. Next, we introduce a sequence (τ_m) of $(\mathcal{G}_t^n)_{t \geq 0}$ -stopping times by $\tau_m = \inf\{t \geq 0 : V_n(t) \geq m\}$ (where the inf over an empty set is defined to be ∞). By Proposition 4.3, it is evident that $\lim_{m \rightarrow \infty} \tau_m \wedge T = T$ a.s. and hence for a fixed $T > 0$,

$$\mathbb{E} \left[\int_0^T e^{-\gamma t} C_n(V_n(t-)) dA_n(t) \right] = \lim_{m \rightarrow \infty} \mathbb{E} \left[\int_0^{\tau_m \wedge T} e^{-\gamma t} C_n(V_n(t-)) dA_n(t) \right].$$

Since $0 \leq V_n(t) \leq m$ for $0 \leq t \leq \tau_m$, the integrand is bounded and $\mathbb{E}[\int_0^{\tau_m \wedge T} e^{-\gamma t} C_n(V_n(t-)) dM_n(t)]$ equals to zero. As a result, by letting m tend to infinity and then T tend to infinity, we have

$$\mathbb{E} \left[\sum_{j=1}^{\infty} e^{-\gamma t_j^n} C_n(V_n(t_j^n-)) \right] = n \mathbb{E} \left[\int_0^{\infty} e^{-\gamma t} C_n(V_n(t)) \lambda_n(V_n(t)) dt \right].$$

Then, using Assumption 3.2 and Theorem 6.5 in this section, it is easy to verify that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^{\infty} e^{-\gamma t} \left| n \lambda_n(V_n(t)) C_n(V_n(t)) - C(\widehat{V}_n(t)) \right| dt \right] = 0.$$

Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \sum_{j=1}^{\infty} e^{-\gamma t_j^n} C_n(V_n(t_j^n-)) - \int_0^{\infty} e^{-\gamma t} C(\widehat{V}_n(t)) dt \right| \right] = 0,$$

and it makes sense to consider a cost of the form $\mathbb{E}[\int_0^{\infty} e^{-\gamma t} C(\widehat{V}_n(t)) dt]$.

Under our assumptions, we intend to show that the cost functionals in (6.3) and (6.4) are finite. Our main result here is the convergence of $J(\widehat{V}_n, \widehat{L}_n)$ to $J(Z, L)$ as n tends to infinity.

6.2 Assumptions and the convergence of the cost functionals

We need to make further assumptions in this section. We assume that the running cost function $C(\cdot)$ can have polynomial growth and the service times (v_i) have higher moments. We also need to strengthen the part (iii) of Assumption 3.2. All these assumptions will be used in the proof of main theorem (Theorem 6.3) here, but Theorem 6.5, which is of independent interest, remains valid only with the assumption (6.6) below. We will make it clear in the statements of these results. Next, we list the additional assumptions below:

- (a) There exist a constant $K_1 > 0$ and an integer $\ell \geq 1$ such that

$$0 \leq C(x) \leq K_1(1 + x^\ell), \text{ for all } x \geq 0. \quad (6.5)$$

Here $C(\cdot)$ is the running cost function in (6.3).

(b) The sequence of service times (v_i) described in Section 3 also satisfies

$$\mathbb{E}[v_i^{m(1+\epsilon)}] < \infty \text{ for some integer } m > \max\{\ell, 2\}, \text{ and small } \epsilon > 0. \quad (6.6)$$

(c) The sequence of arrival intensity functions $(\lambda_n(\cdot))$ satisfies the following two conditions:

(i) There exist two constants $\delta_0 > 0$ and $M > 0$ such that

$$\sup_{n \geq 1} \sup_{x \in [0, \delta_0]} \sqrt{n} |\lambda_n(x) - 1| < M. \quad (6.7)$$

(ii) There exist two constants $A > 0$ and $B > 0$ such that

$$\sup_{n \geq 1} \sqrt{n} (\lambda_n(x) - 1)^+ \leq A + Bx \text{ for all } x \geq 0. \quad (6.8)$$

(d) The sequence of patience-time distribution functions (F_n) also satisfies

$$0 \leq \sqrt{n} F_n \left(\frac{x}{\sqrt{n}} \right) \leq C_1 x (1 + x^r), \text{ for all } x \geq 0, \quad (6.9)$$

where $C_1 > 0$ is a generic constant independent of n and the constant $r > 0$ satisfies $2(r + 1) < m$.

Since we have already assumed that $\sqrt{n}(1 - \lambda_n(x/\sqrt{n}))$ converges to a non-negative function $u(x)$ for all $x \geq 0$ (Assumption 3.2, part (iv)), conditions (6.7) and (6.8) are not very restrictive. (See also the examples in Remark 3.4.) Assumptions (6.6) and (6.9) will be used in obtaining some uniform integrability estimates for the integrand in the cost functional $J(\widehat{V}_n, \widehat{L}_n)$.

Remark 6.1. *If the cost functional $J(\widehat{V}_n, \widehat{L}_n)$ does not deal with the idle time costs, that is if $p = 0$, then we do not need the assumptions (6.7) and (6.9). In that case, Proposition 6.10 also not necessary and the estimate (6.29) will be sufficient to obtain Theorem 6.3 below.*

Remark 6.2. *The assumption (6.9) indeed imposes some restrictions on the Assumption 3.3 of Section 3. Here we follow up on the changes required in the examples (F_n) provided in Remark 3.4.*

(a) Let $F_n \equiv F$ for all n , and assume F is differentiable with a derivative of polynomial growth satisfying

$$\sup_{y \in [0, x]} F'(y) \leq C(1 + x^r) \text{ with } 0 \leq r < m - 1,$$

where $F'(y)$ denotes a derivative of F at y . Then (F_n) satisfies Assumption 3.3 as well as (6.9).

(b) Take $F_n(x) = 1 - \exp(-\int_0^x h(\sqrt{nu}) du)$ for $x \geq 0$ and assume that h is a continuous function with polynomial growth satisfying $\sup_{y \in [0, x]} h(y) \leq C(1 + x^r)$ with $0 \leq r < m - 1$. This sequence (F_n) also satisfies Assumption 3.3 as well as (6.9).

(c) For a general sequence (F_n) , assume that $F_n(\frac{x}{\sqrt{n}})$ converges to a non-negative function $h(x)$ uniformly on compact sets and $0 \leq F_n'(\frac{x}{\sqrt{n}}) \leq C(1 + x^r)$, where $C > 0$ is a constant independent of n . Then, (F_n) satisfies Assumption 3.3 as well as (6.9).

Our main theorem in this section is the following:

Theorem 6.3. *In addition to the basic assumptions in Section 3, assume (6.5)–(6.9) to hold. Then the cost functionals $J(\widehat{V}_n, \widehat{L}_n)$ and $J(Z, L)$ are all finite and*

$$\lim_{n \rightarrow \infty} J(\widehat{V}_n, \widehat{L}_n) = J(Z, L). \quad (6.10)$$

Proof of this theorem needs several preliminary results. Using Theorem 4.10, together with Skorokhod's representation theorem, we can simply assume that $\lim_{n \rightarrow \infty} (\widehat{V}_n(t), \widehat{L}_n(t)) = (Z(t), L(t))$ for all $t \geq 0$, a.s. To obtain the convergence of cost functionals, we need to obtain a polynomial growth bound which is independent of n for the expected value of the integrand in $J(\widehat{V}_n, \widehat{L}_n)$.

Lemma 6.4. *Assume (6.6) in addition to the basic assumptions in Section 3. Let $\xi_n(\cdot)$ be the process described in (4.37). Then,*

$$\mathbb{E} [\|\xi_n\|_T^m] \leq K_2(1 + T^{m/2}), \quad (6.11)$$

where $K_2 > 0$ is a generic constant independent of n .

Proof. From (4.37), we have

$$\xi_n(t) = \widehat{A}_n(t) + \widehat{M}_n^v(\bar{A}_n(t)) - \widehat{M}_n^d(\bar{A}_n(t)) \quad \text{for all } t \geq 0.$$

First, we estimate $\mathbb{E}[\|\widehat{A}_n\|_T^m]$. By (2.13), A_n has the representation $A_n(t) = Y_n(n \int_0^t \lambda_n(V_n(s)) ds)$ for all $t \geq 0$, where Y_n is a unit-rate Poisson process. We introduce the Poisson martingale $\widehat{Y}_n(t) = \frac{1}{\sqrt{n}}(Y_n(nt) - nt)$ and then we can write $\widehat{A}_n(t) = \widehat{Y}_n(\int_0^t \lambda_n(V_n(s)) ds)$ as in (2.13). Moreover, for any integer $k \geq 1$, $\|\widehat{A}_n\|_T^{2k} \leq \|\widehat{Y}_n\|_{C_0 T}^{2k}$, where the constant $C_0 > 0$ is as in Assumption 3.2. Consequently, $\mathbb{E}[\|\widehat{A}_n\|_T^{2k}] \leq \mathbb{E}[\|\widehat{Y}_n\|_{C_0 T}^{2k}]$.

The quadratic variation process of the martingale \widehat{Y}_n is given by $[\widehat{Y}_n, \widehat{Y}_n](t) = \frac{1}{n} Y_n(nt)$ and therefore, using Burkholder's inequality (cf. [24]) we obtain $\mathbb{E}[\|\widehat{Y}_n\|_{C_0 T}^{2k}] \leq \frac{C_k}{n^k} \mathbb{E}[Y_n(nC_0 T)^k]$, where $C_k > 0$ is a constant depending only on k . Recall that if X is a Poisson random variable with parameter $\lambda > 0$, then for any integer $j \geq 1$, $\mathbb{E}[X(X-1)\cdots(X-(j-1))] = \lambda^j$. Consequently, $\mathbb{E}(X^j) = p_j(\lambda)$, where $p_j(x)$ is a degree j polynomial of the form $p_j(x) = x^j + c_{j-1}x^{j-1} + \cdots + c_1x$ and the constants c_1, c_2, \dots, c_{j-1} may depend on j . Since $Y_n(nC_0 T)$ is a Poisson random variable with parameter $nC_0 T > 0$, we can easily obtain the bound

$$\frac{1}{n^k} \mathbb{E} [Y_n(nC_0 T)^k] \leq C_1 p_k(T),$$

where $C_1 > 0$ is a constant and $p_k(x)$ is a polynomial of degree k . The constant $C_1 > 0$ and the polynomial $p_k(\cdot)$ can be chosen independent of n but it may depend on k . Using these estimates and letting $T > 1$, we have $\mathbb{E}[\|\widehat{A}_n\|_T^{2k}] \leq C_2 p_k(T) \leq \widetilde{C}_k(1 + T^k)$, where $C_2 > 0$ and $\widetilde{C}_k > 0$ are generic constants independent of n . Consequently, using Hölder's inequality,

$$\mathbb{E}[\|\widehat{A}_n\|_T^m] \leq K_m(1 + T^{m/2}), \quad (6.12)$$

where $K_m > 0$ is a constant independent of n . Since $A_n(T) \leq (\sqrt{n}\widehat{A}_n(T) + nC_0 T)$, we can easily use the above estimate to obtain

$$\mathbb{E}[(A_n(T))^k] \leq C_k n^k(1 + T^k) \quad \text{for each } k \geq 1, \quad (6.13)$$

where $C_k > 0$ is a generic constant independent of n and T .

Next, we intend to estimate $\mathbb{E}[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))|^m]$. Consider the filtration $(\widehat{\mathcal{F}}_j^n)_{j \geq 1}$ introduced in (3.2). Let $T > 0$ be fixed. Then $A_n(T)$ is a stopping time with respect to this filtration $(\widehat{\mathcal{F}}_j^n)_{j \geq 1}$, since $[A_n(T) = k] = [t_k^n \leq T < t_{k+1}^n] \in \widehat{\mathcal{F}}_k^n$. We introduce a sequence of random variables related to the n -th system by

$$S_j = \frac{1}{\sqrt{n}} \sum_{i=1}^j (v_i - 1) \mathbf{1}_{[V_n(t_i^n -) < d_i^n]} \quad \text{and} \quad S_0 = 0. \quad (6.14)$$

We suppress the dependence of S_j on n for simplicity of the presentation. Following an argument similar to the establishment of martingale property of $M^v(n)$ in (2.4), we observe that $(S_j)_{j \geq 1}$ is a martingale with respect to the filtration $(\widehat{\mathcal{F}}_j^n)_{j \geq 1}$. Next, observe that

$$\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))|^m = \sup_{j \leq A_n(T)} |S_j|^m. \quad (6.15)$$

Hence, we can use the fact that $A_n(T)$ is an $(\widehat{\mathcal{F}}_j^n)$ -stopping time to estimate $\mathbb{E}[\sup_{j \leq A_n(T)} |S_j|^m]$.

We intend to use Rosenthal's inequality for square integrable martingales (see, e.g., [28]). First notice that the predictable quadratic variation process $[S_j, S_j]$ of (S_j) satisfies $[S_j, S_j] \leq \sigma_s^2 j/n$. Using Rosenthal's inequality (Theorem 1 in Section 2 of [28] with $p = m$ and the stopping time $S \equiv A_n(T)$ therein), we obtain

$$\mathbb{E} \left[\sup_{j \leq A_n(T)} |S_j|^m \right] \leq C_m \left[\frac{\sigma^m}{n^{m/2}} \mathbb{E} \left(A_n(T)^{m/2} \right) + \mathbb{E} \left((\Delta S)_{A_n(T)}^* \right)^m \right], \quad (6.16)$$

where $C_m > 0$ is a constant which depends only on m and $(\Delta S)_t^* \equiv \sup_{s \leq t} |\Delta S_s|$. Using (6.13) and the fact that $[\mathbb{E} (A_n(T)^{m/2})]^2 \leq \mathbb{E}[A_n(T)^m]$, we have

$$\frac{\sigma^m}{n^{m/2}} \mathbb{E} \left(A_n(T)^{m/2} \right) \leq \bar{C}_1 (1 + T^{m/2}), \quad (6.17)$$

where $\bar{C}_1 > 0$ is a constant independent of n and T . It is easy to observe that

$$\mathbb{E} \left((\Delta S)_{A_n(T)}^* \right)^m \leq \frac{1}{n^{m/2}} \mathbb{E} \left(\sup_{j \leq A_n(T)} |v_j - 1|^m \right).$$

To estimate the second term in (6.16), we let $K > 2$ be a constant independent of n and T , and we pick the precise value of K later. We consider

$$\mathbb{E} \left(\sup_{j \leq A_n(T)} |v_j - 1|^m \right) \leq \mathbb{E} \left(\sup_{j \leq KnT} |v_j - 1|^m \right) + \mathbb{E} \left(\sup_{j \leq A_n(T)} |v_j - 1|^m \mathbf{1}_{[A_n(T) > KnT]} \right). \quad (6.18)$$

Using (7.6) and the estimates there in the Appendix,

$$\mathbb{E} \left(\sup_{j \leq KnT} |v_j - 1|^m \right) \leq \bar{C}_2 nT,$$

where $\bar{C}_2 > 0$ is a constant independent of n and T . Since $m > 2$, we have

$$\frac{1}{n^{m/2}} \mathbb{E} \left(\sup_{j \leq KnT} |v_i - 1|^m \right) \leq \bar{C}_2 T. \quad (6.19)$$

Next, we consider

$$J \equiv \mathbb{E} \left(\sup_{j \leq A_n(T)} |v_j - 1|^m \mathbf{1}_{[A_n(T) > KnT]} \right).$$

From (2.13), it follows that $A_n(T) \leq Y_n(nC_0T)$ where $C_0 > 0$ is the constant in Assumption 3.2 part (i) and Y_n is a unit-rate Poisson process. Hence

$$J \leq \sum_{k > KnT} \mathbb{E} \left(\sup_{j \leq k} |v_j - 1|^m \mathbf{1}_{[Y_n(nC_0T) = k]} \right).$$

Now we let $p = (1 + \epsilon)$ and $q = 1 + \frac{1}{\epsilon}$ so that $\frac{1}{p} + \frac{1}{q} = 1$, where $\epsilon > 0$ is as in (6.6). Then by Hölder's inequality,

$$J \leq \sum_{k > KnT} \left[\mathbb{E} \left(\sup_{j \leq k} |v_j - 1|^{m(1+\epsilon)} \right) \right]^{1/(1+\epsilon)} \cdot [\mathbb{P}[Y_n(nC_0T) = k]]^{\epsilon/(1+\epsilon)}.$$

Using (6.6) together with (7.5) and (7.6) in the Appendix, we have $\mathbb{E} \left(\sup_{j \leq k} |v_j - 1|^{m(1+\epsilon)} \right) \leq \bar{C}_3 k$, where $\bar{C}_3 > 0$ is a generic constant independent of n and T . Furthermore, $\mathbb{P}[Y_n(nC_0T) = k] = e^{-nC_0T} \frac{(nC_0T)^k}{k!}$. Using these two estimates and by a simple algebraic manipulation, we derive

$$J \leq \bar{C}_4 (nC_0T)^{\epsilon/(1+\epsilon)} e^{-nC_0T\epsilon/(1+\epsilon)} \sum_{k=KnT}^{\infty} k \left(\frac{(nC_0T)^k}{k!} \right)^{\epsilon/(1+\epsilon)},$$

where $\bar{C}_4 > 0$ is a generic constant independent of n and T . Next, we use the fact that $\log(k!)^{1/q} \geq \frac{1}{q}(k \log k - k)$ where $q = 1 + \frac{1}{\epsilon}$ and thus $(k!)^{\epsilon/(1+\epsilon)} \geq \left(\frac{k}{e}\right)^{k\epsilon/(1+\epsilon)}$. To simplify the notation, we also introduce the function $g(x) = x^{1/q} e^{-x/q}$ for all $x \geq 0$. Notice that g is positive, continuous, and $\lim_{x \rightarrow \infty} g(x) = 0$. Thus, $g(\cdot)$ is bounded and $0 \leq g(x) \leq \bar{M}$, where $\bar{M} = g(1)$. Then we obtain

$$J \leq \bar{C}_4 \bar{M} \sum_{k=KnT}^{\infty} k \left[\left(\frac{eC_0}{K} \right)^{\epsilon/(1+\epsilon)} \right]^k.$$

Now, we choose the constant $K > 0$ so that $\left(\frac{eC_0}{K}\right)^{\epsilon/(1+\epsilon)} < \frac{1}{2}$. Then we have

$$J \leq \bar{C}_4 \bar{M} \sum_{k=0}^{\infty} k \left(\frac{1}{2} \right)^k < \infty \quad (6.20)$$

and all the constants on the right hand side are independent of n and T . Therefore, combining (6.16)–(6.20), we obtain the estimate

$$\mathbb{E} \left[\sup_{j \leq A_n(T)} |S_j|^m \right] \leq \bar{C}_m [1 + T^{m/2}],$$

where $\bar{C}_m > 0$ is a constant independent of n and T .

Thus, we have

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\widehat{M}_n^v(\bar{A}_n(t))|^m \right] \leq \bar{K}_m (1 + T^{m/2}), \quad (6.21)$$

where $\widehat{K}_m > 0$ is a generic constant independent of n and T . A very similar computation for $\mathbb{E}[\sup_{t \in [0, T]} |\widehat{M}_n^d(\bar{A}_n(t))|^m]$ yields

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\widehat{M}_n^d(\bar{A}_n(t))|^m \right] \leq \widehat{K}_m (1 + T^{m/2}), \quad (6.22)$$

where $\widehat{K}_m > 0$ is a generic constant independent of n and T . Combining (6.12), (6.21) and (6.22), the desired conclusion (6.11) follows. \blacksquare

Next, we prove the following theorem which is of independent interest and it complements Theorem 4.1 and Proposition 4.3.

Theorem 6.5. *In addition to the basic assumptions in Section 3, assume (6.6) to hold. Then*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\|V_n\|_T^m] = 0, \quad (6.23)$$

where $m \geq 2$ is given in (6.6).

Proof. Let the processes X_n and \mathcal{T}_n be described by (4.2) and (4.9), respectively. Then using (4.11) and the Lipschitz continuity of the Skorokhod map Γ in (4.1), we have

$$0 \leq \|V_n\|_T \leq 2\|\mathcal{T}_n\|_T. \quad (6.24)$$

But using (4.2), (4.35) and a simple algebraic manipulation, we can write

$$\mathcal{T}_n(t) = \frac{1}{\sqrt{n}} \xi_n(t) + \int_0^t [\lambda(\widehat{V}_n(s)/\sqrt{n}) - 1] ds$$

for all $t \geq 0$. Since $m \geq 2$, we use the inequality $(a + b)^m \leq C_m(a^m + b^m)$, where $a \geq 0$, $b \geq 0$ and $C_m = 2^{m-1}$ to obtain

$$\mathbb{E}(\|\mathcal{T}_n\|_T^m) \leq C_m \left[\frac{1}{n^{m/2}} \mathbb{E}(\|\xi_n\|_T^m) + \mathbb{E} \left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) \right], \quad (6.25)$$

where $C_m > 0$ is a constant independent of n and T . Also, notice that

$$\begin{aligned} \mathbb{E} \left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) &\leq T^{m-1} \mathbb{E} \left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) \\ &\leq T^{m-1} \mathbb{E} \left[\left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) \mathbf{1}_{[\|V_n\|_T \leq K]} \right] \\ &\quad + T^{m-1} \mathbb{E} \left[\left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) \mathbf{1}_{[\|V_n\|_T > K]} \right], \end{aligned}$$

where $K > 0$ is a constant. Hence using part (i) of Assumption 3.2, we have

$$\mathbb{E} \left(\int_0^T \left| \lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right|^m ds \right) \leq T^m \left[\sup_{x \in [0, K]} |\lambda_n(x) - 1| + (C_0 + 1)^m \mathbb{P}[\|V_n\|_T > K] \right]. \quad (6.26)$$

The first term in the right hand side of (6.26) tends to zero as $n \rightarrow \infty$, by part (ii) of Assumption 3.2, and the second term also tends to zero as $n \rightarrow \infty$ by Theorem 4.1. Using this together with (6.11) in (6.25), yields $\lim_{n \rightarrow \infty} \mathbb{E}(\|\mathcal{T}_n\|_T^m) = 0$. Then we can use (6.24) to reach the desired conclusion (6.23). This completes the proof. \blacksquare

Remark 6.6. If $\mathbb{E}(v_i^{2+\epsilon}) < \infty$ for some $\epsilon > 0$ then $\lim_{n \rightarrow \infty} \mathbb{E}(\|\mathcal{T}_n\|_T^2) = 0$ holds.

Lemma 6.7. In addition to the basic assumptions in Section 3, assume (6.5)–(6.8) to hold. Then

$$\sup_{n \geq 1} \mathbb{E} \left(\sqrt{n} \int_0^T \left(\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right)^+ ds \right)^m \leq K_m(1 + T^{2m}). \quad (6.27)$$

Here $K_m > 0$ is a constant independent of n and T .

Proof. Assuming (6.8), we obtain

$$\begin{aligned} \mathbb{E} \left(\sqrt{n} \int_0^T \left(\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right)^+ ds \right)^m &\leq T^m \mathbb{E}(A + B\|V_n\|_T)^m \\ &\leq \widetilde{C}_m T^m [1 + \mathbb{E}(\|V_n\|_T^m)], \end{aligned} \quad (6.28)$$

where $\widetilde{C}_m > 0$ is a constant independent of n and T . The constants $A > 0$ and $B > 0$ are as in (6.8). Next, by (6.24), $\mathbb{E}\|V_n\|_T^m \leq 2^m \mathbb{E}\|\mathcal{T}_n\|_T^m$. Then, we can employ (6.25) together with (6.11) to obtain $\mathbb{E}\|\mathcal{T}_n\|_T^m \leq \widetilde{K}_m [1 + T^{m/2} + T^m]$, where $\widetilde{K}_m > 0$ is a generic constant independent of n and T . Combining these facts with (6.28), the desired result follows. ■

Proposition 6.8. In addition to the basic assumptions in Section 3, assume (6.5)–(6.8) to hold. Then we have

$$\mathbb{E}\|\widehat{V}_n\|_T^m \leq K_m [1 + T^{2m}], \quad (6.29)$$

where $K_m > 0$ is a constant independent of n and T .

Proof. Let the process \widehat{Z}_n be as in the proof of Proposition 4.3. Then $\|\widehat{V}_n\|_T \leq 2\|\widehat{Z}_n\|_T$ for all $T > 0$ as explained there. Moreover,

$$\widehat{Z}_n(t) = \xi_n(t) + \sqrt{n} \int_0^t (\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1)^+ ds$$

for all $t \geq 0$, where ξ_n is as in Lemma 6.4. Consequently,

$$\mathbb{E}\|\widehat{Z}_n\|_T^m \leq C_m \left(\mathbb{E}\|\xi_n\|_T^m + \mathbb{E} \left[\sqrt{n} \int_0^t (\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1)^+ ds \right]^m \right),$$

where $C_m > 0$ is a generic constant independent of n and T . Using this estimate, (6.11), (6.27), and the fact that $\mathbb{E}\|\widehat{V}_n\|_T^m \leq 2^m \mathbb{E}\|\widehat{Z}_n\|_T^m$, we obtain (6.29). ■

Remark 6.9. The above proposition strengthens the result in Theorem 6.5. The estimate (6.29) implies that $\mathbb{E}\|V_n\|_T^m \leq K_m(1 + T^{2m})/(\sqrt{n})^m$.

In the following proposition, we obtain uniform L^2 -estimates for $\frac{1}{\sqrt{n}} \int_0^T F_n(\widehat{V}_n(s-)/\sqrt{n}) dA_n(s)$ and for $\sqrt{n} \int_0^T |\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1| ds$.

Proposition 6.10. Under the assumptions of Theorem 6.3, the followings hold:

$$(i) \quad \mathbb{E} \left[\frac{1}{\sqrt{n}} \int_0^T F_n(\widehat{V}_n(s-)/\sqrt{n}) dA_n(s) \right]^2 \leq \bar{C}_1(1 + T^{2(m+1)}) \text{ and}$$

$$(ii) \quad \mathbb{E} \left[\sqrt{n} \int_0^T |\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1| ds \right]^2 \leq \bar{C}_2(1 + T^{2(m+1)}).$$

As a consequence,

$$\mathbb{E}(\widehat{L}_n(T))^2 \leq \bar{C}_3(1 + T^{2(m+1)}), \quad (6.30)$$

where \widehat{L}_n is as in (6.1). Here \bar{C}_1, \bar{C}_2 and \bar{C}_3 are positive constants independent of n and T .

Proof. Notice that

$$0 \leq \frac{1}{\sqrt{n}} \int_0^T F_n(\widehat{V}_n(s-)/\sqrt{n}) dA_n(s) \leq \frac{1}{\sqrt{n}} F_n(\|\widehat{V}_n\|_T/\sqrt{n}) A_n(T).$$

Using this together with (6.9), we have

$$0 \leq \frac{1}{\sqrt{n}} \int_0^T F_n(\widehat{V}_n(s-)/\sqrt{n}) dA_n(s) \leq C_1 \bar{A}_n(T) (\|\widehat{V}_n\|_T + \|\widehat{V}_n\|_T^{r+1}),$$

where $C_1 > 0$ is a constant independent of n and T . Consequently,

$$\mathbb{E} \left[\frac{1}{\sqrt{n}} \int_0^T F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) \right]^2 \leq C_2 \mathbb{E} \left[\bar{A}_n(T)^2 (\|\widehat{V}_n\|_T^2 + \|\widehat{V}_n\|_T^{2(r+1)}) \right], \quad (6.31)$$

where $C_2 > 0$ is a constant independent of n and T . Using Hölder's inequality, we obtain

$$\begin{aligned} \mathbb{E}[\bar{A}_n(T)^2 \|\widehat{V}_n\|_T^2] &\leq \left[\mathbb{E} \|\widehat{V}_n\|_T^m \right]^{2/m} \left[\mathbb{E} (\bar{A}_n(T))^{(2m)/(m-2)} \right]^{(m-2)/m} \\ &\leq K_1 (1 + T^{2m})^{2/m} (1 + T^{2m/(m-2)})^{(m-2)/m} \\ &\leq K_2 (1 + T^4) (1 + T^2) \\ &\leq K_3 (1 + T^6), \end{aligned} \quad (6.32)$$

where the second inequality follows from (6.29) and (6.13). Here $K_i > 0$ ($i = 1, 2, 3$) are constants independent of n and T . Next we estimate the term $\mathbb{E}[\bar{A}_n(T)^2 \|\widehat{V}_n\|_T^{2(r+1)}]$. By (6.9), $2(r+1) < m$ and we take $p = \frac{m}{2(r+1)} > 1$ and $q = \frac{1}{1-1/p} > 1$. Thus $\frac{1}{p} + \frac{1}{q} = 1$. Then, using Hölder's inequality again, we obtain

$$\begin{aligned} \mathbb{E} \left[\bar{A}_n(T)^2 \|\widehat{V}_n\|_T^{2(r+1)} \right] &\leq \left[\mathbb{E} \|\widehat{V}_n\|_T^m \right]^{1/p} \left[\mathbb{E} \bar{A}_n(T)^{2q} \right]^{1/q} \\ &\leq \tilde{K}_1 (1 + T^{2m})^{1/p} (1 + T^{2q})^{1/q} \\ &\leq \tilde{K}_2 (1 + T^{2m}) (1 + T^2) \\ &\leq \tilde{K}_3 (1 + T^{2(m+1)}), \end{aligned} \quad (6.33)$$

where the second inequality follows from (6.29), (6.13), and $\tilde{K}_i > 0$ ($i = 1, 2, 3$) are constants independent of n and T . Since $m > 2$, by combining (6.31)–(6.33) we obtain part (i).

For part (ii), notice that

$$\mathbb{E} \left[\sqrt{n} \int_0^T |\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1| ds \right]^2 \leq T \mathbb{E} \int_0^T \left[\sqrt{n} |\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1| \right]^2 ds. \quad (6.34)$$

By (6.7), we have

$$\mathbb{E} \left[\int_0^T \left[\sqrt{n} |\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1| \right]^2 ds \mathbf{1}_{\|\widehat{V}_n\|_T \leq \delta_0} \right] \leq M^2 T, \quad (6.35)$$

where $M > 0$ is a constant independent of n and T as given in (6.7). Also, since $|\lambda_n(x) - 1| \leq C_0 + 1$ for all $x \geq 0$, where C_0 is as in Assumption 3.2, we obtain

$$\begin{aligned} \mathbb{E} \left[\int_0^T [\sqrt{n}|\lambda_n(\widehat{V}_n(s)/\sqrt{n}) - 1|]^2 ds \mathbf{1}_{\{\|V_n\|_T > \delta_0\}} \right] &\leq (C_0 + 1)^2 T n \mathbb{P}[\|V_n\|_T > \delta_0] \\ &\leq (C_0 + 1)^2 T \frac{n}{n^{m/2}} \frac{\mathbb{E}(\|\widehat{V}_n\|_T^m)}{\delta_0^m}, \end{aligned} \quad (6.36)$$

where (6.36) is from Chebychev's inequality. Since $m > 2$, $\frac{n}{n^{m/2}} < 1$ and by (6.29), the left side of (6.36) is bounded above by $\widetilde{C}_0(1+T^{2m})$ for some constant $\widetilde{C}_0 > 0$. Thus by combining (6.34)–(6.36), we establish part (ii).

For (6.30), using (6.1) and (6.2), we notice that

$$\widehat{L}_n(T) = \widehat{V}_n(T) + \frac{1}{\sqrt{n}} \int_0^T F_n \left(\frac{\widehat{V}_n(s-)}{\sqrt{n}} \right) dA_n(s) - \xi_n(T) - \sqrt{n} \int_0^T \left[\lambda_n \left(\frac{\widehat{V}_n(s)}{\sqrt{n}} \right) - 1 \right] ds, \quad (6.37)$$

where $\xi_n(\cdot)$ is described in (4.37). From (6.11), (6.29) and Jensen's inequality, we have

$$\begin{aligned} \mathbb{E}[|\xi_n(T)|^2] &\leq (\mathbb{E}[\|\xi_n\|_T^m])^{2/m} \leq K_1(1+T^{m/2})^{2/m} \leq \widetilde{K}_1(1+T) \text{ and,} \\ \mathbb{E}[\|\widehat{V}_n(T)\|^2] &\leq (\mathbb{E}[\|\widehat{V}_n\|_T^m])^{2/m} \leq K_2(1+T^{2m})^{2/m} \leq \widetilde{K}_2(1+T^2). \end{aligned}$$

Notice that $m > 2$, $\widetilde{K}_1 = 2K_1$, $\widetilde{K}_2 = 2K_2$ and these constants are independent of n and T . Now using these two estimates together with parts (i) and (ii) of this proposition in (6.37), we obtain (6.30). \blacksquare

With all these preliminary results in hand, now we are able to prove Theorem 6.3.

Proof of Theorem 6.3. First we consider the cost functional $J(\widehat{V}_n, \widehat{L}_n)$ in (6.3). With the polynomial bound (6.30) in hand, using integration by parts, it can be easily verified that

$$\mathbb{E} \left[\int_0^\infty e^{-\gamma t} d\widehat{L}_n(t) \right] = \gamma \mathbb{E} \left[\int_0^\infty e^{-\gamma t} \widehat{L}_n(t) dt \right].$$

Therefore, we have the representation

$$J(\widehat{V}_n, \widehat{L}_n) = \mathbb{E} \int_0^\infty e^{-\gamma t} [C(\widehat{V}_n(t)) + \gamma p \widehat{L}_n(t)] dt. \quad (6.38)$$

Since $(\widehat{V}_n, \widehat{L}_n) \rightarrow (Z, L)$ a.s. as $n \rightarrow \infty$, using (6.30) together with Fatou's lemma, we obtain

$$\mathbb{E}[L(T)^2] \leq \bar{C}_3(1+T^{2(m+1)}), \quad (6.39)$$

where $\bar{C}_3 > 0$ is a constant as in (6.30). Hence, using integration by parts again, we can also write $J(Z, L)$ described in (6.4) as

$$J(Z, L) = \mathbb{E} \int_0^\infty e^{-\gamma t} [C(Z(t)) + \gamma p L(t)] dt. \quad (6.40)$$

Let us consider the term $\mathbb{E}[\int_0^\infty e^{-\gamma t} C(\widehat{V}_n(t)) dt]$ in (6.38). Let μ be the probability measure on the Borel σ -algebra \mathcal{B} of $[0, \infty)$, defined by $\mu(B) = \gamma \int_B e^{-\gamma t} dt$ for each Borel set B . Consider the

probability measure $\mu \otimes \mathbb{P}$ on the space $[0, \infty) \times \Omega$ equipped with the product σ -algebra $\mathcal{B} \otimes \mathcal{F}$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is our probability space. Then using Fubini's theorem, we have

$$\mathbb{E}_{\mu \otimes \mathbb{P}}[C(\widehat{V}_n)] = \gamma \mathbb{E} \left[\int_0^\infty e^{-\gamma t} C(\widehat{V}_n(t)) dt \right]. \quad (6.41)$$

Since $\widehat{V}_n(t) \rightarrow Z(t)$ for all $t \geq 0$ a.s., we have $C(\widehat{V}_n)$ converges to $C(Z)$ almost surely in $\mu \otimes \mathbb{P}$ as $n \rightarrow \infty$. Next, we show the uniform integrability of $(C(\widehat{V}_n))$. Let $m > 2$ be as in (6.6). Using the assumptions (6.5), (6.6) and the simple inequality $0 \leq (1 + x^r) \leq 2(1 + x)^r \leq 2^r(1 + x^r)$ for $r = \frac{m}{\ell} > 1$ and $x \geq 0$, we obtain

$$\begin{aligned} \sup_{n \geq 1} \mathbb{E}_{\mu \otimes \mathbb{P}}[(C(\widehat{V}_n))^r] &\leq \gamma K_1 \sup_{n \geq 1} \mathbb{E} \left[\int_0^\infty e^{-\gamma t} (1 + (\widehat{V}_n(t))^m) dt \right] \\ &\leq \gamma K_2 \int_0^\infty e^{-\gamma t} (1 + t^{2m}) dt \\ &< \infty, \end{aligned}$$

where the second inequality follows from (6.29). The constants $K_1, K_2 > 0$ are independent of n . Hence $\mathbb{E}_{\mu \otimes \mathbb{P}}[C(\widehat{V}_n)] < \infty$ for all $n \geq 1$ and $\lim_{n \rightarrow \infty} \mathbb{E}_{\mu \otimes \mathbb{P}}[C(\widehat{V}_n)] = \mathbb{E}_{\mu \otimes \mathbb{P}}[C(Z)]$. Indeed, $\mathbb{E}_{\mu \otimes \mathbb{P}}[C(Z)]$ is finite and bounded above by $\gamma K_2 \int_0^\infty e^{-\gamma t} (1 + t^{2m}) dt$. Hence

$$\lim_{n \rightarrow \infty} \gamma \int_0^\infty e^{-\gamma t} C(\widehat{V}_n(t)) dt = \gamma \mathbb{E} \int_0^\infty e^{-\gamma t} C(Z(t)) dt. \quad (6.42)$$

In a similar manner, we can establish uniform integrability of (\widehat{L}_n) by using (6.30),

$$\begin{aligned} \sup_{n \geq 1} \mathbb{E}_{\mu \otimes \mathbb{P}}[(\widehat{L}_n)^2] &= \gamma \sup_{n \geq 1} \mathbb{E} \left[\int_0^\infty e^{-\gamma t} (\widehat{L}_n(t))^2 dt \right] \\ &\leq \bar{C}_3 \gamma \int_0^\infty e^{-\gamma t} (1 + t^{2(m+1)}) dt < \infty, \end{aligned}$$

where $\bar{C}_3 > 0$ is a constant independent of n as in (6.30). Hence, using Theorem 4.10, we can conclude

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu \otimes \mathbb{P}}[\widehat{L}_n] = \mathbb{E}_{\mu \otimes \mathbb{P}}[L] \quad \text{and} \quad \mathbb{E}_{\mu \otimes \mathbb{P}}[L] \leq \bar{C}_3 \gamma \int_0^\infty e^{-\gamma t} (1 + t^{2(m+1)}) dt < \infty.$$

This yields

$$\lim_{n \rightarrow \infty} \mathbb{E} \int_0^\infty e^{-\gamma t} \widehat{L}_n(t) dt = \mathbb{E} \int_0^\infty e^{-\gamma t} L(t) dt. \quad (6.43)$$

Since $\gamma > 0$ and $p > 0$ are constants in (6.38) and (6.40), it immediately follows that $J(\widehat{V}_n, \widehat{L}_n)$ converges to $J(Z, L)$ as $n \rightarrow \infty$. This completes the proof of Theorem 6.3. \blacksquare

7 Appendix

In this appendix, we use a result on change of intensities of point processes to construct an arrival process $A(\cdot)$ as described in Section 2.

Proposition 7.1. *There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an arrival process $A(\cdot)$ and two filtrations $(\mathcal{G}_t)_{t \geq 0}$ and $(\widetilde{\mathcal{G}}_t)_{t \geq 0}$ defined on this space satisfying the conditions (2.9)–(2.12) of Section 2. Moreover, the following results hold:*

(i) The service time and patience-time sequences (v_i) and (d_i) are i.i.d. and independent of each other. Each v_i has mean 1 and variance $\sigma^2 > 0$. Each d_i has the given distribution function F .

(ii) For each $n \geq 0$, the random variables $\{(d_{n+j}, v_{n+j}) : j = 1, 2, \dots\}$ are independent of $\widehat{\mathcal{F}}_n$, where the filtration $(\widehat{\mathcal{F}}_n)_{n \geq 0}$ is as in (2.2). In particular, for each $n \geq 0$, the random variable d_{n+1} is independent of $\widehat{\mathcal{F}}_n$.

Proof. We begin with a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ and a sequence of random variables (v_i) and (d_i) on this space with the following properties: the sequences of random variables (v_i) and (d_i) are i.i.d. and independent of each other. Each d_i has the given distribution function F , while each v_i has mean 1 and variance $\sigma^2 > 0$. We define the σ -algebra $\overset{\circ}{\mathcal{G}}_0$ by $\overset{\circ}{\mathcal{G}}_0 \equiv \sigma((v_i, d_i) : i = 1, 2, \dots)$. Let $A(\cdot)$ be a unit-rate Poisson process which is independent of the σ -algebra $\overset{\circ}{\mathcal{G}}_0$. We introduce the offered waiting time process $V(\cdot)$ using the sequence $(v_i, d_i)_{i \geq 1}$ and the process $A(\cdot)$ as in (2.1). Next, introduce the filtrations $(\mathcal{G}_t)_{t \geq 0}$ and $(\overset{\circ}{\mathcal{G}}_t)_{t \geq 0}$ by

$$\mathcal{G}_t \equiv \sigma(A(s), V(s) : 0 \leq s \leq t) \quad \text{and} \quad \overset{\circ}{\mathcal{G}}_t \equiv \mathcal{G}_t \vee \overset{\circ}{\mathcal{G}}_0, \quad (7.1)$$

for all $t \geq 0$, where $\overset{\circ}{\mathcal{G}}_t$ is the σ -algebra generated by the sets in $\mathcal{G}_t \cup \overset{\circ}{\mathcal{G}}_0$.

Next, consider the discrete filtration $(\overset{\circ}{\mathcal{F}}_n)_{n \geq 0}$ given by $\overset{\circ}{\mathcal{F}}_0 \equiv \{\emptyset, \Omega\}$ and

$$\overset{\circ}{\mathcal{F}}_n = \sigma((t_1, v_1, d_1), \dots, (t_n, v_n, d_n))$$

for $n \geq 1$ as in (2.9). For each $t \geq 0$, it is easy to verify that $A(t)$ is a $(\overset{\circ}{\mathcal{F}}_n)_{n \geq 0}$ -stopping time. Then, we introduce the filtration $(\tilde{\mathcal{G}}_t)_{t \geq 0}$ by

$$\tilde{\mathcal{G}}_t \equiv \overset{\circ}{\mathcal{F}}_{A(t)} \quad \text{for all } t \geq 0. \quad (7.2)$$

Following the discussion below (2.12) in Section 2, it yields that

$$\mathcal{G}_t \subseteq \tilde{\mathcal{G}}_t \subseteq \overset{\circ}{\mathcal{G}}_t \quad \text{for all } t \geq 0. \quad (7.3)$$

Next, we introduce the left-continuous version $\tilde{V}(\cdot)$ of the offered waiting time process by

$$\tilde{V}(t) = \begin{cases} V(t), & \text{if } t \neq t_i, \\ V(t_i-), & \text{if } t = t_i. \end{cases} \quad (7.4)$$

Thus $\tilde{V}(\cdot)$ in (7.4) is (\mathcal{G}_t) -adapted left-continuous process with left limits. Consequently, $\tilde{V}(\cdot)$ is a *predictable* process with respect to $(\mathcal{G}_t)_{t \geq 0}$, $(\tilde{\mathcal{G}}_t)_{t \geq 0}$ as well as $(\overset{\circ}{\mathcal{G}}_t)_{t \geq 0}$ (see Section 3 of Chapter 1 in [7]). Next, we intend to use a result on change of intensities for point processes (cf. Section 2 of Chapter 6 in [7]). We introduce the process $L(\cdot)$ by

$$L(t) = \begin{cases} \exp\left(\int_0^t (1 - \lambda(\tilde{V}(s))) ds\right), & \text{if } 0 \leq t < t_1, \\ \exp\left(\int_0^t (1 - \lambda(\tilde{V}(s))) ds + \int_0^t \log(\lambda(\tilde{V}(s))) dA(s)\right), & \text{if } t \geq t_1. \end{cases}$$

Since $\lambda(\cdot)$ is Borel measurable and $0 < \epsilon < \lambda(x) < C$ for all $x \geq 0$, we can use theorems T2, T3 and T4 in pages 165–168 of [7] to verify that $L(\cdot)$ is a $(\mathring{\mathcal{G}}_t)$ -martingale and $\mathbb{E}_{\mathbb{Q}}[L(t)] = 1$ for each $t \geq 0$. Also, since $L(\cdot)$ is adapted to $(\mathcal{G}_t)_{t \geq 0}$, it is also a (\mathcal{G}_t) -martingale.

Let $T > 0$ be fixed and introduce the probability measure \mathbb{P}_T on $\mathring{\mathcal{G}}_T$ by

$$\frac{d\mathbb{P}_T}{d\mathbb{Q}} = L(T).$$

Then by Theorem T3 in Chapter 6 of [7], the process $A(\cdot)$ has $(\mathbb{P}_T, \mathring{\mathcal{G}}_t)$ -intensity $\lambda(\tilde{V}(t))$ for $0 \leq t \leq T$. Using Theorem T9 in page 28 of [7] and following a straightforward computation, it follows that $\{A(t) - \int_0^t \lambda(\tilde{V}(s))ds : 0 \leq t \leq T\}$ is a $(\mathring{\mathcal{G}}_t)_{0 \leq t \leq T}$ -martingale with respect to \mathbb{P}_T . Since this process is adapted to $(\mathcal{G}_t)_{t \geq 0}$ as well as $(\tilde{\mathcal{G}}_t)_{t \geq 0}$, using (7.3) it follows that $\{A(t) - \int_0^t \lambda(\tilde{V}(s))ds : 0 \leq t \leq T\}$ is a \mathbb{P}_T -martingale with respect to $(\mathcal{G}_t)_{0 \leq t \leq T}$ as well as $(\tilde{\mathcal{G}}_t)_{0 \leq t \leq T}$.

It is evident that $\int_0^t \lambda(\tilde{V}(s))ds = \int_0^t \lambda(V(s))ds$ for all $t \geq 0$ and the probability measures $(\mathbb{P}_T)_{T > 0}$ are consistent and thus there is a probability measure \mathbb{P} on $\mathring{\mathcal{G}}_\infty$ such that \mathbb{P}_T and \mathbb{P} agree on $\mathring{\mathcal{G}}_T$ for each $T > 0$. Hence, we conclude that with respect to the probability measure \mathbb{P} , the process $\{A(t) - \int_0^t \lambda(\tilde{V}(s))ds : t \geq 0\}$ is a martingale with respect to each of the filtrations $(\mathcal{G}_t)_{t \geq 0}$, $(\tilde{\mathcal{G}}_t)_{t \geq 0}$ and $(\mathring{\mathcal{G}}_t)_{t \geq 0}$. Next, notice that \mathbb{P}_T and \mathbb{Q} agree on $\mathring{\mathcal{G}}_0$, hence the assertion (i) of the proposition follows.

To establish (ii), we intend to show that the sequence $\{(d_{n+j}, v_{n+j}) : j = 1, 2, \dots\}$ is independent of $\hat{\mathcal{F}}_n$ for each $n \geq 0$, with respect to the probability measure \mathbb{P} . Let $m \geq 1$, $T > 0$ and C_1, \dots, C_m and K_1, \dots, K_m be positive constants. We also pick arbitrary Borel sets A_1, A_2, \dots, A_n in \mathbb{R}^3 . Introduce the sets

$$G_1 \equiv \{d_{n+1} \leq K_1, \dots, d_{n+m} \leq K_m, v_{n+1} \leq C_1, \dots, v_{n+m} \leq C_m\}$$

and

$$G_2 \equiv \{(t_1, v_1, d_1) \in A_1, \dots, (t_n, v_n, d_n) \in A_n, t_{n+1} \leq T\}.$$

Notice that G_2 is a basic set in $\hat{\mathcal{F}}_n$ and it is also in $\mathring{\mathcal{G}}_T$. The set G_1 is in $\mathring{\mathcal{G}}_0$ and the probability measures \mathbb{P} and \mathbb{Q} agree on $\mathring{\mathcal{G}}_0$. Thus,

$$\mathbb{P}(G_1) = \mathbb{Q}(G_1) = \prod_{j=1}^m \mathbb{Q}[d_{n+j} \leq K_j] \mathbb{Q}[v_{n+j} \leq C_j] = \prod_{j=1}^m F(K_j) \mathbb{Q}[v_{n+j} \leq C_j].$$

Notice that $G_1 \cap G_2$ is in $\mathring{\mathcal{G}}_T$ and $L(T)\mathbf{1}_{G_2}$ is $\hat{\mathcal{F}}_n$ -measurable. Hence,

$$\mathbb{P}[G_1 \cap G_2] = \mathbb{E}_{\mathbb{Q}}[L(T)\mathbf{1}_{G_1 \cap G_2}] = \mathbb{E}_{\mathbb{Q}}[L(T)\mathbf{1}_{G_2} \mathbb{E}_{\mathbb{Q}}[\mathbf{1}_{G_1} | \hat{\mathcal{F}}_n]].$$

But G_1 is independent of $\hat{\mathcal{F}}_n$ with respect to \mathbb{Q} -probability. Hence, $\mathbb{E}_{\mathbb{Q}}[\mathbf{1}_{G_1} | \hat{\mathcal{F}}_n] = \mathbb{Q}(G_1) = \mathbb{P}(G_1)$. Therefore,

$$\begin{aligned} \mathbb{P}[G_1 \cap G_2] &= \mathbb{P}(G_1) \mathbb{E}_{\mathbb{Q}}[L(T)\mathbf{1}_{G_2}] = \mathbb{P}(G_1) \mathbb{P}(G_2) \\ &= \mathbb{P}(G_2) \prod_{j=1}^m F(K_j) \mathbb{Q}[v_1 \leq C_j]. \end{aligned}$$

Consequently, the random variables $\{(d_{n+j}, v_{n+j}) : j = 1, 2, \dots\}$ are independent of $\hat{\mathcal{F}}_n$ for each $n \geq 0$. This completes the proof of part (ii). \blacksquare

The following lemma was used in the proofs of Proposition 4.12 and Lemma 6.4. We include it for completeness.

Lemma 7.2. *Let $T > 0$ be fixed. Consider a sequence of non-negative i.i.d. random variables (X_n) with $\mathbb{E}(X_n) < \infty$. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\max_{1 \leq j \leq [nT]} X_j \right] = 0. \quad (7.5)$$

Proof. Without loss of generality, we can simply take $T > 2$. Let G be the distribution function of X_n and introduce $u = \sup\{x \geq 0 : G(x) < 1\}$. Notice that $0 \leq u \leq +\infty$. Since (X_n) is i.i.d., we have

$$\mathbb{P} \left[\max_{1 \leq j \leq [nT]} X_j \leq x \right] = G(x)^{[nT]}$$

and therefore,

$$\begin{aligned} \mathbb{E} \left[\max_{1 \leq j \leq [nT]} X_j \right] &= \int_0^u (1 - G(x)^{[nT]}) dx \\ &= \int_0^u \int_x^u [nT] G(y)^{[nT]-1} dG(y) dx \\ &= [nT] \int_0^u y G(y)^{[nT]-1} dG(y), \end{aligned}$$

by using Fubini's theorem. Consequently,

$$0 \leq \frac{1}{n} \mathbb{E} \left[\max_{1 \leq j \leq [nT]} X_j \right] \leq T \int_0^u y G(y)^{[nT]-1} dG(y) \leq T \int_0^u y dG(y). \quad (7.6)$$

Since $[nT] \geq 2$, $0 \leq y G(y)^{[nT]-1} \leq y$ for all $y \geq 0$ and $\lim_{n \rightarrow \infty} y G(y)^{[nT]-1} = 0$ for $0 \leq y \leq u$. On the other hand, $\int_0^u y dG(y) < \infty$ since $\mathbb{E}(X_n) < \infty$. Therefore, by the dominated convergence theorem, we conclude $\lim_{n \rightarrow \infty} \int_0^u y G(y)^{[nT]-1} dG(y) = 0$. Hence, using this in (7.6) we obtain the desired conclusion (7.5). This completes the proof. \blacksquare

Acknowledgment

We are thankful to anonymous referees for their comments which helped us to improve the content and the presentation of this article. The research work of A. W. was supported in part by the Army Research Office under the grant no. W911NF0710424 and he is grateful for their support.

References

- [1] B. Ata. Dynamic control of a multiclass queue with thin arrival streams. *Oper. Res.*, 54(5):876–892, 2006.
- [2] B. Ata, J. M. Harrison, and L. A. Shepp. Drift rate control of a Brownian processing system. *Ann. Appl. Probab.*, 15(2):1145–1160, 2005.
- [3] B. Ata and S. Kumar. Heavy traffic analysis of open processing networks with complete resource pooling: asymptotic optimality of discrete review policies. *Ann. Appl. Probab.*, 15(1A):331–391, 2005.

- [4] R. Atar. Scheduling control for queueing systems with many servers: asymptotic optimality in heavy traffic. *Ann. Appl. Probab.*, 15(4):2606–2650, 2005.
- [5] R. Atar, A. Mandelbaum, and M. I. Reiman. Scheduling a multi class queue with many exponential servers: asymptotic optimality in heavy traffic. *Ann. Appl. Probab.*, 14(3):1084–1134, 2004.
- [6] P. Billingsley. *Convergence of Probability Measures, Second Edition*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., New York, 1999.
- [7] P. Brémaud. *Point processes and queues. Martingale dynamics*. Springer Series in Statistics. Springer-Verlag, New York, 1981.
- [8] K. Burdzy, W. Kang, and K. Ramanan. The Skorokhod problem in a time-dependent interval. *Stochastic Process. Appl.*, 119(2):428–452, 2009.
- [9] H. Chen and D. D. Yao. *Fundamentals of queueing networks*, volume 46 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001.
- [10] J. G. Dai and S. He. Customer abandonment in many-server queues. *Math. Oper. Res.*, 35(2):347–362, 2010.
- [11] J. G. Dai, S. He, and T. Tezcan. Many-server diffusion limits for $G/Ph/n + GI$ queues. *Ann. Appl. Probab.*, 20(5):1854–1890, 2010.
- [12] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and Convergence*. Wiley, New York, 1986.
- [13] O. Garnett, A. Mandelbaum, and M. I. Reiman. Designing a call center with impatient customers. *Manufacturing and Service Oper. Management*, 4(1A):208–227, 2002.
- [14] J. M. George and J. M. Harrison. Dynamic control of a queue with adjustable service rate. *Oper. Res.*, 49(5):720–731, 2001.
- [15] A. Ghosh and A. Weerasinghe. Optimal buffer size and dynamic rate control for a queueing network with reneging in heavy traffic. *Stochastic Process. Appl.*, 120(11):2103–2141, 2010.
- [16] J. M. Harrison. *Brownian motion and stochastic flow systems*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1985.
- [17] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic. I. *Adv. in Appl. Probab.*, 2:150–177, 1970.
- [18] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Graduate Texts in Mathematics 113. Springer, New York, second edition, 1991.
- [19] Y. L. Kocaga and A. R. Ward. Admission control for a multi-server queue with abandonment. *Queueing Syst.*, 65(3):275–323, 2010.
- [20] L. Kruk, J. Lehoczky, K. Ramanan, and S. Shreve. An explicit formula for the Skorokhod map on $[0, a]$. *Ann. Probab.*, 35(5):1740–1768, 2007.
- [21] A. Mandelbaum and P. Momčilović. Queues with many servers: the virtual waiting-time process in the QED regime. *Math. Oper. Res.*, 33(3):561–586, 2008.

- [22] A. Mandelbaum and A. Weerasinghe. A many server controlled queuing system with impatient customers. *Preprint*, 2010.
- [23] C. Palm. Etude des delais d'attente. *Ericsson Technics*, 5:37–56, 1937.
- [24] P. E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, second edition, 2004.
- [25] J. E. Reed and T. Tezcan. Hazard rate scaling for the $GI/M/n + GI$ queue. *Preprint*, 2009.
- [26] J. E. Reed and A. R. Ward. Approximating the $GI/GI/1 + GI$ queue with a nonlinear drift diffusion: hazard rate scaling in heavy traffic. *Math. Oper. Res.*, 33(3):606–644, 2008.
- [27] M. I. Reiman. Some diffusion approximations with state space collapse. In *Modelling and performance evaluation methodology (Paris, 1983)*, volume 60 of *Lecture Notes in Control and Inform. Sci.*, pages 209–240. Springer, Berlin, 1984.
- [28] Y. Ren and F. Tian. On the Rosenthal's inequality for locally square integrable martingales. *Stochastic Process. Appl.*, 104(1):107–116, 2003.
- [29] R. Talreja and W. Whitt. Heavy-traffic limits for waiting times in many-server queues with abandonment. *Ann. Appl. Probab.*, 19(6):2137–2175, 2009.
- [30] A. R. Ward and P. W. Glynn. A diffusion approximation for a Markovian queue with reneging. *Queueing Syst.*, 43(1-2):103–128, 2003.
- [31] A. R. Ward and P. W. Glynn. A diffusion approximation for a $GI/GI/1$ queue with balking or reneging. *Queueing Syst.*, 50(4):371–400, 2005.
- [32] A. R. Ward and S. Kumar. Asymptotically optimal admission control of a queue with impatient customers. *Math. Oper. Res.*, 33(1):167–202, 2008.
- [33] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002.
- [34] W. Whitt. Proofs of the martingale FCLT. *Probab. Surv.*, 4:268–302, 2007.
- [35] K. Yamada. Diffusion approximation for open state-dependent queueing networks in the heavy traffic situation. *Ann. Appl. Probab.*, 5(4):958–982, 1995.
- [36] S. Zeltyn and A. Mandelbaum. Call centers with impatient customers: many-server asymptotics of the $M/M/n + G$ queue. *Queueing Syst.*, 51(3-4):361–402, 2005.