ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA

VOL. LXI, 2007 SECTIO A 51–90

ŁUKASZ KRUK

Diffusion approximation for a G/G/1 EDF queue with unbounded lead times

Abstract. We present a heavy traffic analysis for a G/G/1 queue in which customers have unbounded random deadlines correlated with their service times. The customers are processed according to the earliest-deadline-first (EDF) queue discipline. At any time, the customers have a lead time, the time until their deadline lapses. We model the evolution of these lead times as a random measure on the real line. Under suitable scaling, it is proved that the measure-valued lead-time process converges to a deterministic function of the workload process. This work is a generalization of Doytchinov et al. [6], which developed these results for the case of bounded deadlines independent of the service times. Another generalization of the latter results, covering the case of long range dependence, is also discussed.

1. Introduction. Real-time queueing theory is devoted to the study of systems that service customers with individual timing requirements. Such systems arise naturally in manufacturing in which orders have due dates, or in real-time computer and communication networks. To study queueing systems in which the customers have deadlines, we must attach a lead-time variable to each customer in the system (the lead time is the time until the deadline of the customer's job). It is convenient to model the vector

2000 *Mathematics Subject Classification.* 60K25, 60G57, 68M20.
Key words and phrases. Due dates, heavy traffic, queueing, diffusion limit, random measure.
Supported by the KBN Grant No. 2 PO3A 012 23.

of customer lead times at any time t as a counting measure on $\mathbb{R}$ with a unit atom at the current lead-time of each customer and total mass equal to the number of customers in the system at that time. Doytchinov et al. [6] investigated the single queue case under the Earliest-Deadline-First (EDF) queue discipline. They proved that under heavy traffic conditions, a suitably scaled random lead time measure converges to a non-random function of the limit of the scaled workload process. Kruk et al. [11] gave the corresponding results for the First-In-First-Out (FIFO) queue discipline and generalized both the EDF and the FIFO results to the case of a single station with K input streams, queued in separate buffers and served by the head-of-the-line processor sharing (HOL-PS) policy across streams. Yeung and Lehoczky [16] generalized the single server, single customer class analysis for EDF and FIFO to multi-class feedforward networks. Kruk et al. [12] extended these results to multi-class acyclic EDF networks. The accuracy of the approximations of Doytchinov et al. [6] was investigated in Kruk et al. [9, 10].

In all the above-mentioned papers, it was assumed that the (suitably rescaled) customer lead times are bounded above by a constant $y^* < \infty$ and the arguments depended heavily on this assumption. Moreover, independence of the customer service times and initial lead times was always assumed. Both these assumptions may be limiting in some applications, e.g., they do not allow for modelling a regularization of the Shortest-Remaining-Processing-Time-First (SRPT) protocol suggested by Bender et al. [1], in which we use (pseudo-) lead times equal to (suitable positive multiples of) the service times. It is perhaps surprising that the deterministic upper bound for the customer initial lead times seems to be the most important assumption for the analysis of Doytchinov et al. [6] and their result may be generalized considerably with little additional effort as long as we keep this assumption. For a more detailed discussion of this issue, see Appendix, to follow. On the other hand, the need for generalizing the existing theory to more general deadline distributions was already recognized in Doytchinov et al. [6], which provided simulation results suggesting that the main result of that paper should hold also in the case of unbounded lead times.

The aim of this paper is to get a counterpart of the result of Doytchinov et al. [6] for unbounded initial lead times whose positive parts have finite second moments. Our analysis does not require the independence of the customer service times and their initial lead times. It turns out that the approach developed by Doytchinov et al. [6], based on arrival analysis and the observation that the number of partially served customers and the work associated with them are negligible under heavy-traffic scaling, can still be applied. However, in our case, the analysis of the timing requirements of the incoming customers is more difficult and requires different probabilistic tools. Additional difficulties also arise when the workload in the system is small, because then customers with arbitrarily large lead times may receive service. We consider the single queue, single customer class case, but it should be possible to extend our result to HOL-PS stations, feedforward and acyclic networks along the lines of Kruk et al. [11, 12], Yeung and Lehoczky [16].

This paper is organized as follows. Section 2 presents the model, notation and assumptions. It also introduces the measure-valued processes associated with customer lead times and the frontier processes. Section 3 states the main results of the paper. Section 4 is devoted to the analysis of the lead-time profiles of the arriving customers and the work associated with them. In Section 5 we show that the work in the system associated with partially served customers is negligible under diffusion scaling and that the same is true about the number of these customers, provided that the workload is not too small. Section 6 provides proofs of the main results. Section 7 contains two examples illustrating our results. Appendix presents an immediate generalization of the results of Doytchinov et al. [6] to the case of dependent customer arrival times, service times and lead times under the assumption that the customer lead times are bounded from above.

2. The model, assumptions and notation.

2.1. Notation. The following notation will be used throughout the paper. Let $\mathbb{N} = \{1, 2, \ldots\}$ and let $\mathbb{R}$ denote the set of real numbers. For $a, b \in \mathbb{R}$, we write $a \vee b$ for the maximum of a and b, $a \wedge b$ for the minimum of a and b, a^+ for the positive part of a, $\lfloor a \rfloor$ for the largest integer less than or equal to a and $\lceil a \rceil$ for the smallest integer greater than or equal to a. For $a, b \in \mathbb{R}$ such that $a \geq b$, by definition, $(a, b] \triangleq \emptyset$. Let $\overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$ be the two-point compactification of $\mathbb{R}$ with the obvious topology.

A rectangle $(s_1, s_2] \times (t_1, t_2]$, where $s_1, s_2, t_1, t_2 \in \mathbb{R}$, $s_1 < s_2$, $t_1 < t_2$, will be called a *block*. Two blocks B_1, B_2 are called *neighbouring* if they share an edge, i.e., $B_1 = (s_1, s_2] \times (t_1, t_2]$ and either $B_2 = (s_2, s_3] \times (t_1, t_2]$, or $B_2 = (s_1, s_2] \times (t_2, t_3]$ for some $s_i, t_i \in \mathbb{R}$, $i = 1, 2, 3$. For a two-parameter random field $X(s, t)$ and a block $B = (s_1, s_2] \times (t_1, t_2]$, let $X(B) \triangleq X(s_2, t_2) - X(s_1, t_2) - X(s_2, t_1) + X(s_1, t_1)$.

Denote by $\mathcal{M}$ the set of all finite, nonnegative measures on $\mathcal{B}(\mathbb{R})$, the Borel subsets of $\mathbb{R}$. Under the weak topology, $\mathcal{M}$ is a Polish space.

Let A be an arbitrary set. The space $\ell^\infty(A)$ is defined as the set of all uniformly bounded real functions on A, i.e., all functions $z : A \to \mathbb{R}$ such that $||z||_A \triangleq \sup_{a \in A} |z(a)| < \infty$. $(\ell^\infty(A), || \cdot ||_A)$ is a Banach space (not necessarily separable).

We will use the symbol $\Rightarrow$ to denote weak convergence of measures, either on $\mathbb{R}$ (in this case, the same symbol is used for convergence of the corresponding cumulative distribution functions (c.d.f.s)) or $\mathbb{R}^2$, or on $\ell^\infty(A)$

for a suitable set A, or, finally, on the space $D_S[0,\infty)$ of right-continuous functions with left-hand limits (RCLL functions) from $[0,\infty)$ to a Polish space S, equipped with the Skorokhod J_1 topology. See van der Vaart and Wellner [14], Whitt [15] for details. When dealing with $D_S[0,\infty)$, we take $S = \mathbb{R}$ or $\mathbb{R}^d$, with appropriate dimension d for vector-valued functions, unless explicitly stated otherwise. We will also use the space $D([0,T]^2)$ of real, RCLL functions on a square $[0,T]^2$, see, e.g., Bickel and Wichura [2] for its definition and more details.

For functions $f, g : \mathbb{R} \to \mathbb{R}$, where g is RCLL, and for $-\infty < a < b \leq \infty$, we write $\int_a^b f(s)dg(s)$ (or $\int_a^b f(s)g(ds)$) to denote $\int_{(a,b]} f(s)dg(s)$.

Denote by e the identity map on $[0,\infty)$, i.e., $e(t) = t$, $t \geq 0$.

2.2. The basic model. We have a sequence of single-station queueing systems, each serving one class of customers. The queueing systems are indexed by superscript n.

The inter-arrival times for the customer arrival process are $\{u_j^n\}_{j=1}^{\infty}$, a sequence of strictly positive, independent, identically distributed (i.i.d.) random variables (r.v.s) with mean $1/\lambda_n$ and standard deviation α_n. The service times are $\{v_j^n\}_{j=1}^{\infty}$, another sequence of positive, i.i.d. r.v.s with mean $1/\mu_n$ and standard deviation β_n. We assume that each queue is empty at time zero and

$$\lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} \mu_n = \lambda > 0. \tag{2.1}$$

We define the *customer arrival times*

$$S_0^n \triangleq 0, \qquad S_k^n \triangleq \sum_{i=1}^{k} u_i^n, \quad k \geq 1, \tag{2.2}$$

the *customer arrival process*

$$A^n(t) \triangleq \max\{k : S_k^n \leq t\}, \quad t \geq 0, \tag{2.3}$$

and the *work arrival process*

$$V^n(t) \triangleq \sum_{j=1}^{\lfloor t \rfloor} v_j^n, \quad t \geq 0. \tag{2.4}$$

The work which has arrived to the queue by time t is then $V^n(A^n(t))$.

Each customer arrives with an initial lead time L_j^n, the time between the arrival time and the deadline for completion of service for that customer. These initial lead times have common distribution given by

$$\mathbb{P}\{L_j^n \leq \sqrt{n}y\} = G^n(y), \tag{2.5}$$

where

$$G^n \Rightarrow G. \tag{2.6}$$

We assume that the random vectors $\left\{\left(v_j^n, L_j^n\right)\right\}_{j=1}^{\infty}$ are i.i.d. and that

$$G_v^n(y) \triangleq \mathbb{E}\left[v_j^n \mathbb{I}_{\{L_j^n \le \sqrt{n}y\}}\right] \Rightarrow G_v(y), \tag{2.7}$$

$$G_{v^2}^n(y) \triangleq \mathbb{E}\left[\left(v_j^n\right)^2 \mathbb{I}_{\{L_j^n \le \sqrt{n}y\}}\right] \Rightarrow G_{v^2}(y), \tag{2.8}$$

where G_v and G_{v^2} are c.d.f.s of finite positive measures on $\mathbb{R}$ such that G_v has total mass $1/\lambda$ and $G_v(y) < 1/\lambda$ for every $y \in \mathbb{R}$. We also assume that for every n, the sequences $\left\{u_j^n\right\}_{j=1}^{\infty}$ and $\left\{\left(v_j^n, L_j^n\right)\right\}_{j=1}^{\infty}$ are mutually independent.

Customers are served using the EDF queue discipline, i.e., the server always serves the customer with the shortest lead time. Preemption is permitted (we assume preempt-resume). There is no set up, switch-over, or other type of overhead. Late customers (customers with negative lead times) stay in queue until served to completion.

The *netput process*

$$N^n(t) \triangleq V^n\big(A^n(t)\big) - t \tag{2.9}$$

measures the amount of work in queue at time t provided that the server is never idle up to time t. The *cumulative idleness process*

$$I^n(t) \triangleq -\inf_{0 \le s \le t} N^n(s), \tag{2.10}$$

gives the amount of time the server is idle, and adding this to the netput process, we obtain the *workload process*

$$W^n(t) = N^n(t) + I^n(t), \tag{2.11}$$

which records the amount of work in the queue, taking server idleness into account. All the above processes are independent of the queue service discipline, provided that the server is never idle when there are customers in the queue. However, the *queue length process* $Q^n(t)$, which is the number of customers in the queue at time t, depends on the queue discipline. All these processes are RCLL.

2.3. Heavy traffic assumptions. We assume that

$$\lim_{n\to\infty} \alpha_n = \alpha > 0, \quad \lim_{n\to\infty} \beta_n = \beta > 0 \tag{2.12}$$

and that

$$\mathbb{E}\left((L_j^n)^+\right)^2 \le \tilde{C}n \tag{2.13}$$

for some constant $\tilde{C}$ and all n. Define the *traffic intensity* $\rho_n \triangleq \lambda_n/\mu_n$. We make the *heavy traffic assumption*

$$\lim_{n\to\infty} \sqrt{n}(1-\rho_n) = \gamma \tag{2.14}$$

for some $\gamma \in \mathbb{R}$. We impose the Lindeberg condition on the inter-arrival times, service times and the positive parts of the rescaled lead times:

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{E}\left[\left(u_j^n - (\lambda_n)^{-1}\right)^2 \mathbb{I}_{\left\{\left|u_j^n-(\lambda_n)^{-1}\right|>c\sqrt{n}\right\}}\right] \\ &\quad = \lim_{n\to\infty} \mathbb{E}\left[\left(v_j^n - (\mu_n)^{-1}\right)^2 \mathbb{I}_{\left\{\left|v_j^n-(\mu_n)^{-1}\right|>c\sqrt{n}\right\}}\right] \\ &\quad = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\left((L_j^n)^+ - \mathbb{E}(L_j^n)^+\right)^2 \mathbb{I}_{\left\{\left|(L_j^n)^+-\mathbb{E}(L_j^n)^+\right|>cn\right\}}\right] = 0 \end{aligned} \tag{2.15}$$

for all $c > 0$. We extend $G_{v^2}^n$ to $\overline{\mathbb{R}}$ by $G_{v^2}(-\infty) \triangleq 0$, $G_{v^2}(\infty) \triangleq \mathbb{E}(v_j^n)^2$. For every $x, y \in \overline{\mathbb{R}}$, we define a semimetric ρ on $\overline{\mathbb{R}}$ by the formula

$$\rho(x,y) \triangleq \sup_{n\in\mathbb{N}} |G_{v^2}^n(x) - G_{v^2}^n(y)|. \tag{2.16}$$

We assume that $\left(\overline{\mathbb{R}}, \rho\right)$ is a totally bounded semimetric space, i.e., for every $\epsilon > 0$, $\overline{\mathbb{R}}$ may be decomposed into a finite number of sets (or, equivalently, open balls) with radius less than ϵ. This is the case if, for example, G_{v^2} is continuous or $G_{v^2}^n \equiv G_{v^2}$ (see the proof of Lemma 4.2, to follow, for the argument), or, more generally, $G_{v^2}^n = a_n G_{v^2}$, where a_n are real constants converging to 1. The latter is the case if, e.g., $G^n \equiv G$ and the lead times are independent of the service times. However, the assumption (2.8) does not always imply total boundedness of $(\overline{\mathbb{R}}, \rho)$, a counterexample is $G_{v^2}^n(y) = \mathbb{I}_{\{\frac{1}{n}\le y\}}$, $n \in \mathbb{N}$. Finally, we assume that

$$\begin{aligned} &\lim_{y\to\infty} \sup_{n\in\mathbb{N}} \int_y^\infty (1 - G^n(\eta))\, d\eta \\ &\qquad = \lim_{y\to\infty} \sup_{n\in\mathbb{N}} \int_y^\infty \left(\frac{1}{\mu_n} - G_v^n(\eta)\right) d\eta = 0. \end{aligned} \tag{2.17}$$

(2.1), (2.7)–(2.8), (2.17) and Fatou's lemma imply

$$\int_0^\infty (1 - G(\eta))d\eta < \infty, \tag{2.18}$$

$$\int_0^\infty (1 - \lambda\, G_v(\eta))d\eta < \infty. \tag{2.19}$$

We introduce the *heavy traffic scaling* for the idleness, workload and queue length processes

$$\widehat{I}^n(t) = \frac{1}{\sqrt{n}} I^n(nt), \quad \widehat{W}^n(t) = \frac{1}{\sqrt{n}} W^n(nt), \quad \widehat{Q}^n(t) = \frac{1}{\sqrt{n}} Q^n(nt), \tag{2.20}$$

and the *centered heavy traffic scaling* for the arrival processes

$$\widehat{A}^n(t) = \frac{1}{\sqrt{n}}\left[A^n(nt) - \lambda_n nt\right], \qquad \widehat{V}^n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \left(v_j^n - \frac{1}{\mu_n}\right). \tag{2.21}$$

We define also

$$\widehat{N}^n(t) = \frac{1}{\sqrt{n}}\left[V^n\big(A^n(nt)\big) - nt\right]. \tag{2.22}$$

Note that $\widehat{W}^n(t) = \widehat{N}^n(t) + \widehat{I}^n(t)$.

Theorem 3.1 in Prokhorov [13] and Theorem 14.6 in Billingsley [4] imply

$$\widehat{A}^n \Rightarrow A^*, \qquad \widehat{V}^n \Rightarrow V^*, \tag{2.23}$$

where A^* (V^*) is a Brownian motion with no drift and variance $\alpha^2\lambda^3$ (β^2) per unit time. It is also a standard result (see Iglehart and Whitt [7]) that

$$\big(\widehat{N}^n, \widehat{I}^n, \widehat{W}^n\big) \Rightarrow (N^*, I^*, W^*), \tag{2.24}$$

where N^* is a Brownian motion with variance $(\alpha^2+\beta^2)\lambda$ per unit time and drift $-\gamma$, $I^*(t) \stackrel{\Delta}{=} -\min_{0\le s\le t} N^*(s)$, and $W^*(t) = N^*(t) + I^*(t)$. In other words, W^* is a reflected Brownian motion with drift, and I^* causes the reflection.

2.4. Measure-valued processes and frontiers. To study whether customers meet their timing requirements, one must keep track of customer lead times, where the lead time is the time remaining until the deadline elapses, i.e.,

$$\text{lead time} = \text{deadline} - \text{current time}.$$

In this section, we define a collection of measure-valued processes which will be useful in the analysis of the instantaneous lead-time profile of the customers.

Queue length measure:

$$\mathcal{Q}^n(t)(B) \stackrel{\Delta}{=} \left\{\begin{array}{l}\text{Number of customers in the queue at time } t\\ \text{having lead times at time } t \text{ in } B\subset\mathbb{R}\end{array}\right\}.$$

Workload measure:

$$\mathcal{W}^n(t)(B) \stackrel{\Delta}{=} \left\{\begin{array}{l}\text{Work in the queue at time } t \text{ associated with customers}\\ \text{in this queue having lead times at time } t \text{ in } B\subset\mathbb{R}\end{array}\right\}.$$

Customer arrival measure:

$$\mathcal{A}^n(t)(B) \stackrel{\Delta}{=} \left\{\begin{array}{l}\text{Number of all arrivals by time } t,\\ \text{whether or not still in the system at time } t,\\ \text{having lead times at time } t \text{ in } B\subset\mathbb{R}\end{array}\right\}.$$

Workload arrival measure:

$$\mathcal{V}^n(t)(B) \stackrel{\Delta}{=} \left\{\begin{array}{l}\text{Work associated with all arrivals by time } t,\\ \text{whether or not still in the system at time } t,\\ \text{having lead times at time } t \text{ in } B\subset\mathbb{R}\end{array}\right\}.$$

The following relationships easily follow:

$$Q^n(t) = \mathcal{Q}^n(t)(\mathbb{R}), \quad W^n(t) = \mathcal{W}^n(t)(\mathbb{R}), \tag{2.25}$$

$$\mathcal{A}^n(t)(B) = \sum_{j=1}^{A^n(t)} \mathbb{I}_{\{L_j^n-(t-S_j^n)\in B\}} = \sum_{j=1}^{\infty} \mathbb{I}_{\{S_j^n\in B+t-L_j^n,\, S_j^n\le t\}}, \tag{2.26}$$

$$\mathcal{V}^n(t)(B) = \sum_{j=1}^{A^n(t)} v_j^n \mathbb{I}_{\{L_j^n-(t-S_j^n)\in B\}} = \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{S_j^n\in B+t-L_j^n,\, S^n\le t\}}. \tag{2.27}$$

To study the behavior of the EDF queue discipline, it is useful to keep track of the lead time of the customer currently in service and the largest lead time of all customers, whether present or departed, who have ever been in service. We define the *frontier*

$$F^n(t) \triangleq \left\{ \begin{array}{l} \text{Largest lead time of all customers who have ever been} \\ \text{in service, whether still present or not, if } t > S_1^n, \\ \text{or } +\infty, \text{ if } t \le S_1^n \end{array} \right\},$$

the *modified frontier*

$$F_1^n(t) \triangleq \left\{ F^n(t), \text{ if } t \ge n^{\frac{3}{4}}, \text{ or } +\infty, \text{ if } t < n^{\frac{3}{4}} \right\},$$

and the *current lead time*

$$C^n(t) \triangleq \left\{ \begin{array}{l} \text{Lead time of the customer in service,} \\ \text{or } F^n(t) \text{ if the queue is empty} \end{array} \right\}.$$

Under the EDF queue discipline, there is no customer with lead time smaller than $C^n(t)$, and there has never been a customer in service whose lead time, if the customer were still present, would exceed $F^n(t)$. Furthermore, $C^n(t) \le F^n(t) \le F_1^n(t)$ for all $t \ge 0$. F^n, F_1^n and C^n are RCLL.

For the processes just defined, we use the following heavy traffic scalings:

$$\widehat{F}^n(t) \triangleq \frac{1}{\sqrt{n}} F^n(nt), \quad \widehat{F}_1^n(t) \triangleq \frac{1}{\sqrt{n}} F_1^n(nt), \quad \widehat{C}^n(t) \triangleq \frac{1}{\sqrt{n}} C^n(nt), \tag{2.28}$$

$$\widehat{\mathcal{Q}}^n(t)(B) \triangleq \frac{1}{\sqrt{n}} \mathcal{Q}^n(nt)(\sqrt{n}B), \quad \widehat{\mathcal{W}}^n(t)(B) \triangleq \frac{1}{\sqrt{n}} \mathcal{W}^n(nt)(\sqrt{n}B). \tag{2.29}$$

We define also

$$\begin{aligned} \widehat{\mathcal{A}}^n(t)(B) &\triangleq \frac{1}{\sqrt{n}} \mathcal{A}^n(nt)\big(\sqrt{n}B\big) = \frac{1}{\sqrt{n}} \sum_{j=1}^{A^n(nt)} \mathbb{I}_{\{L_j^n-(nt-S_j^n)\in\sqrt{n}B\}} \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} \mathbb{I}_{\{S_j^n\in\sqrt{n}B+nt-L_j^n,\, S_j^n\le nt\}}, \end{aligned} \tag{2.30}$$

$$\begin{aligned}\widehat{\mathcal{V}}^n(t)(B) &\triangleq \frac{1}{\sqrt{n}}\mathcal{V}^n(nt)\big(\sqrt{n}B\big) = \frac{1}{\sqrt{n}}\sum_{j=1}^{A^n(nt)} v_j^n \mathbb{I}_{\{L_j^n-(nt-S_j^n)\in\sqrt{n}B\}}\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{S_j^n\in\sqrt{n}B+nt-L_j^n,\ S_j^n\le nt\}}.\end{aligned}\tag{2.31}$$

For any $y \in \overline{\mathbb{R}}$, define

$$H_v^n(y) \triangleq \lambda_n \int_y^\infty \left(\frac{1}{\mu_n} - G_v^n(\eta)\right) d\eta, \qquad H_v(y) \triangleq \int_y^\infty (1-\lambda G_v(\eta))\, d\eta,$$

$$H^n(y) \triangleq \lambda_n \int_y^\infty (1-G^n(\eta))\, d\eta, \qquad H(y) \triangleq \lambda \int_y^\infty \big(1-G(\eta)\big) d\eta.$$

By (2.18) and (2.19), H and H_v are finite on $\mathbb{R}$. By (2.1), (2.17) and the bounded convergence theorem, H_v^n and H^n are also finite on $\mathbb{R}$ and, moreover, $H_v^n(y) \to H_v(y)$ and $H^n(y) \to H(y)$ uniformly in $y \in [c,\infty)$ for every $c \in \mathbb{R}$. The function H_v maps $\overline{\mathbb{R}}$ onto $[0,\infty]$ and is strictly decreasing and continuous on $\overline{\mathbb{R}}$. Therefore, there exists a continuous inverse function $H_v^{-1} : [0,\infty] \to \overline{\mathbb{R}}$.

The motivation for introducing the modified frontier can be explained as follows. For a queue operating under the EDF discipline, we expect a relationship between the frontier and the workload. For example, if $F^n(t)$ is very negative, there are a lot of customers in the system with lead times greater than $F^n(t)$ and thus $W^n(t)$ is large. Conversely, if $F^n(t)$ is very large, then a customer with a very large lead time must have been served recently, so $W^n(t)$ is likely to be small. In fact, Proposition 3.1, to follow, which is a crucial step in the characterization of the limiting behavior of the processes $\widehat{\mathcal{Q}}^n(t)$ and $\widehat{\mathcal{W}}^n(t)$, asserts that for t not too close to zero,

$$\widehat{F}^n(t) \approx H_v^{-1}\big(\widehat{W}^n(t)\big).\tag{2.32}$$

There is no hope, however, for extending (2.32) to all $t \ge 0$. Indeed, for $t < u_1^n/n$, $\widehat{F}^n(t) = +\infty = H_v^{-1}(0) = H_v^{-1}\big(\widehat{W}^n(t)\big)$, but the random variable $\widehat{F}^n(u_1^n/n) = L_1^n/\sqrt{n}$ has distribution G^n, while $\widehat{W}^n(u_1^n/n) = v_1^n/\sqrt{n} \approx 0$. In fact, the above facts, together with (2.24), imply that the process $\widehat{F}^n$ does *not* converge weakly in $D_{\overline{\mathbb{R}}}[0,\infty)$ equipped with any of the Skorokhod topologies. Therefore, we have introduced the modified frontier process $F_1^n(t)$, which agrees with $F^n(t)$ after an initial time period $\big[0, n^{\frac{3}{4}}\big)$, negligible under heavy-traffic scaling, in which $F_1^n(t) = +\infty$ and $W^n(t)$ is of the order $o\big(\sqrt{n}\big)$, so

$$\widehat{F}_1^n(t) \approx H_v^{-1}\big(\widehat{W}^n(t)\big)$$

for *all* $t \ge 0$ (see Proposition 3.1 and its proof). Intuitively, the modification of the frontier corresponds to giving the system enough time to "warm up" until the relation (2.32) begins to hold. Let us also remark that the exponent

$\frac{3}{4}$ in the definition of $F_1^n(t)$ may be replaced by any $\kappa \in (1/2, 1)$ (the proof of Proposition 3.1 requires that $n^{\kappa-\frac{1}{2}} \to +\infty$ and we want the time interval $[0, n^\kappa)$ to be negligible under heavy-traffic scaling).

3. Main results. We define the *limiting scaled frontier process*

$$F^*(t) \triangleq H_v^{-1}(W^*(t)), \quad t \geq 0, \tag{3.1}$$

where W^* is as in (2.24).

Proposition 3.1. *We have $\widehat{F}_1^n \Rightarrow F^*$ in $D_{\overline{\mathbb{R}}}[0,\infty)$ as $n \to \infty$.*

Let $\mathcal{W}^*$ and $\mathcal{Q}^*$ be the measure-valued processes defined by

$$\mathcal{W}^*(t)(B) \triangleq \int_{B\cap[F^*(t),\infty)} \big(1 - \lambda\, G_v(\eta)\big)\, d\eta, \tag{3.2}$$

$$\mathcal{Q}^*(t)(B) \triangleq \lambda \int_{B\cap[F^*(t),\infty)} \big(1 - G(\eta)\big)\, d\eta, \tag{3.3}$$

for all Borel sets $B \subseteq \mathbb{R}$.

Theorem 3.2. *The processes $\widehat{\mathcal{W}}^n$ and $\widehat{\mathcal{Q}}^n$ converge weakly in $D_{\mathcal{M}}[0,\infty)$ to $\mathcal{W}^*$ and $\mathcal{Q}^*$, respectively.*

Corollary 3.3. *We have $\widehat{Q}^n \Rightarrow Q^* \triangleq H(F^*)$ in $D[0,\infty)$ as $n \to \infty$.*

In particular, the equality $Q^* = \lambda W^*$ does not hold in general, although it does hold if the service times and the lead times are independent.

4. Arrival analysis. In this section, we analyze the limiting behavior of the number of incoming customers with specific timing requirements and the work associated with these customers, without taking departures and service provided by the system into account. We start with Proposition 4.1, a law of large numbers for the distribution function of $\widehat{\mathcal{V}}^n$, which, together with the corresponding Proposition 4.6 for $\widehat{\mathcal{A}}^n$, is the most important auxiliary result of this paper. In its (long and somewhat technical) proof, we use techniques from the theory of empirical processes. Next, in Proposition 4.7, we refine Propositions 4.1 and 4.6 to Glivenko–Cantelli type results. Corollary 4.8, showing that the atoms of $\widehat{\mathcal{V}}^n$ and $\widehat{\mathcal{A}}^n$ are asymptotically negligible, follows.

Proposition 4.1. *Let $T > 0$ and let y be a point of continuity of both G_v and G_{v^2}. Then*

$$\sup_{0\leq t\leq T} \Big| \widehat{\mathcal{V}}^n(t)(y,\infty) - H_v^n(y) + H_v^n\big(y + \sqrt{n}t\big) \Big| \xrightarrow{P} 0. \tag{4.1}$$

To aid the reader, we first provide an outline of the proof. To ease notation throughout this section, let $M_j^n(y) \triangleq v_j^n \mathbb{I}_{\{L_j^n \le \sqrt{n}y\}} - G_v^n(y)$ for $y \in \mathbb{R}$ and $j \in \mathbb{N}$. We also put $M_j^n(-\infty) \triangleq 0$, $M_j^n(\infty) \triangleq v_j^n - \frac{1}{\mu_n}$. By (2.31), we have

$$\widehat{\mathcal{V}}^n(t)(y,\infty) = \frac{1}{\sqrt{n}} \int_y^\infty \sum_{j=1}^\infty \mathbb{I}_{\{nt-\sqrt{n}(l-y)<S_j^n \le nt\}} d\left(v_j^n \mathbb{I}_{\{L_j^n \le \sqrt{n}l\}}\right). \tag{4.2}$$

The main idea of the proof is to approximate $\widehat{\mathcal{V}}^n(t)(y,\infty)$ by

$$I_1^n(t) \triangleq \frac{1}{\sqrt{n}} \int_y^\infty \sum_{j=1}^\infty \mathbb{I}_{\{nt-\sqrt{n}(l-y)<S_j^n \le nt\}} G_v^n(dl), \tag{4.3}$$

i.e., by the process obtained from the RHS of (4.2) by replacing the random variables $v_j^n \mathbb{I}_{\{L_j^n \le \sqrt{n}l\}}$ by their means. It is relatively easy to show that

$$\left| I_1^n(t) - H_v^n(y) + H_v^n\big(y + \sqrt{n}t\big) \right| \Rightarrow 0 \tag{4.4}$$

in $D[0,T]$. Thus, to prove (4.1), it suffices to show that the process

$$\begin{aligned} I_2^n(t) &\triangleq \widehat{\mathcal{V}}^n(t)(y,\infty) - I_1^n(t) \\ &= \frac{1}{\sqrt{n}} \int_y^\infty \sum_{j=1}^\infty \mathbb{I}_{\{nt-\sqrt{n}(l-y)<S_j^n \le nt\}} M_j^n(dl) \end{aligned} \tag{4.5}$$

(i.e., the error in the approximation of $\widehat{\mathcal{V}}^n(t)(y,\infty)$ by $I_1^n(t)$) converges weakly to zero. Intuitively, this should follow from a suitable modification of the law of large numbers. However, a rigorous justification of the fact that $I_2^n \Rightarrow 0$ is rather involved. We break $I_2^n(t)$ into two parts: $I_{2,1}^n(t)$ and $I_{2,2}^n(t)$, corresponding to integration over $\big(y, y+\sqrt{n}t\big]$ and $\big(y+\sqrt{n}t, \infty\big)$ in (4.5), respectively. For $t > 0$, $y+\sqrt{n}t \to \infty$ as $n \to \infty$, so the process $I_{2,2}^n$ may be expected to converge weakly to zero. We show that this is indeed the case with the help of Lemma 4.2 and Corollary 4.3, to follow, using the bracketing central limit theorem from the theory of empirical processes. To deal with $I_{2,1}^n(t)$, we write it in the form

$$I_{2,1}^n(t) = \frac{1}{\sqrt{n}} \int_0^t \sum_{j=A^n(n(t-s))+1}^{A^n(nt)} M_j^n\big(y + \sqrt{n}ds\big) \tag{4.6}$$

and approximate it by the process

$$
\begin{aligned}
U^n(t) &\triangleq \frac{1}{\sqrt{n}} \int_0^t \sum_{j=\lfloor \lambda_n n(t-s)\rfloor+1}^{\lfloor \lambda_n nt\rfloor} M_j^n\big(y+\sqrt{n}ds\big) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda_n nt\rfloor} \left[M_j^n\big(y+\sqrt{n}t\big) - M_j^n\left(y+\sqrt{n}t-\frac{j}{\lambda_n\sqrt{n}}\right)\right]
\end{aligned}
\tag{4.7}
$$

resulting from (4.6) by replacing $A^n(nt)$ by its deterministic approximation $\lambda_n nt$. We then show that both U^n and $I_{2,1}^n - U^n$ converge weakly to zero in $D[0,T]$. The latter task is accomplished, roughly speaking, by using a criterion for tightness of random fields due to Bickel and Wichura [2] and arguing that the finite-dimensional distributions of U^n and $I_{2,1}^n - U^n$ converge to zero.

Proof of Proposition 4.1. For every $t \in [0,T]$,

$$
\widehat{\mathcal{V}}^n(t)(y,\infty) = I_1^n(t) + I_2^n(t),
\tag{4.8}
$$

where $I_1^n(t)$ and $I_2^n(t)$ are given by (4.3) and (4.5), respectively. We have

$$
\begin{aligned}
I_1^n(t) &= \frac{1}{\sqrt{n}} \left(\int_y^{y+\sqrt{n}t} \sum_{j=1}^{\infty} \mathbb{I}_{\{nt-\sqrt{n}(l-y)<S_j^n\le nt\}} \right. \\
&\qquad \left. + \int_{y+\sqrt{n}t}^{\infty} \sum_{j=1}^{\infty} \mathbb{I}_{\{S_j^n \le nt\}} \right) G_v^n(dl) \\
&= \frac{1}{\sqrt{n}} \int_y^{y+\sqrt{n}t} \big(A^n(nt) - A^n\big(nt-\sqrt{n}(l-y)\big)\big)\, G_v^n(dl) \\
&\qquad + \frac{1}{\sqrt{n}} A^n(nt) \left(\frac{1}{\mu_n} - G_v^n\big(y+\sqrt{n}t\big)\right) \\
&= I_{1,1}^n(t) + I_{1,2}^n(t) + I_{1,3}^n(t),
\end{aligned}
\tag{4.9}
$$

$$
I_{1,1}^n(t) = \int_y^{y+\sqrt{n}t} \left(\widehat{A}^n(t) - \widehat{A}^n\left(t-\frac{l-y}{\sqrt{n}}\right)\right) G_v^n(dl),
\tag{4.10}
$$

$$
\begin{aligned}
I_{1,2}^n(t) &= \int_y^{y+\sqrt{n}t} \lambda_n(l-y) G_v^n(dl) \\
&= H_v^n(y) - H_v^n\big(y+\sqrt{n}t\big) - \lambda_n\sqrt{n}t \left(\frac{1}{\mu_n} - G_v^n\big(y+\sqrt{n}t\big)\right),
\end{aligned}
\tag{4.11}
$$

$$
I_{1,3}^n(t) = \Big(\lambda_n\sqrt{n}t + \widehat{A}^n(t)\Big)\left(\frac{1}{\mu_n} - G_v^n\big(y+\sqrt{n}t\big)\right),
\tag{4.12}
$$

$$I_2^n(t) = I_{2,1}^n(t) + I_{2,2}^n(t), \tag{4.13}$$

$$\begin{aligned} I_{2,1}^n(t) &= \frac{1}{\sqrt{n}} \int_y^{y+\sqrt{n}t} \sum_{j=1}^{\infty} \mathbb{I}_{\{nt-\sqrt{n}(l-y)<S_j^n \le nt\}} M_j^n(dl) \\ &= \frac{1}{\sqrt{n}} \int_y^{y+\sqrt{n}t} \sum_{j=A^n(nt-\sqrt{n}(l-y))+1}^{A^n(nt)} M_j^n(dl) \\ &= \frac{1}{\sqrt{n}} \int_0^t \sum_{j=A^n(n(t-s))+1}^{A^n(nt)} M_j^n\big(y+\sqrt{n}ds\big), \end{aligned} \tag{4.14}$$

$$\begin{aligned} I_{2,2}^n(t) &= \frac{1}{\sqrt{n}} \int_{y+\sqrt{n}t}^{\infty} \sum_{j=1}^{\infty} \mathbb{I}_{\{S_j^n \le nt\}} M_j^n(dl) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{A^n(nt)} \left(M_j^n(\infty) - M_j^n\big(y+\sqrt{n}t\big)\right). \end{aligned} \tag{4.15}$$

We begin by showing (4.4). We claim that $I_{1,1}^n \Rightarrow 0$. Indeed, let $\epsilon > 0$. By (2.23), there exists a constant $C > 0$ such that $\mathbb{P}(A_n) \ge 1 - \frac{\epsilon}{2}$ for every n, where $A_n = \left[\sup_{0\le t\le T}\left|\widehat{A}^n(s)\right| \le C\right]$. Let $y_* > y$ be a point of continuity of G_v such that $\frac{1}{\lambda} - G_v(y_*) < \frac{\epsilon}{4C}$. Thus, by (2.1) and (2.7), there exists n_0 such that $\frac{1}{\mu_n} - G_v^n(y_*) < \frac{\epsilon}{4C}$ for $n \ge n_0$. For $\delta > 0$, let $w^n(\delta) \triangleq \sup_{\substack{0\le s_1<s_2\le T \\ s_2-s_1\le\delta}} \left|\widehat{A}^n(s_2) - \widehat{A}^n(s_1)\right|$. By (2.23), there exists $\delta_0 > 0$ such that $\mathbb{P}(B_n) \ge 1 - \frac{\epsilon}{2}$ for every n, where $B_n = \left[w^n(\delta_0) \le \frac{\mu_n\epsilon}{2}\right]$. Then $\mathbb{P}(A_n \cap B_n) \ge 1 - \epsilon$. Moreover, for every $t \in [0,T]$ and $n \ge n_0 \vee \left(\frac{y_*-y}{\delta_0}\right)^2$, on the set $A_n \cap B_n$ we have

$$\begin{aligned} \left|I_{1,1}^n(t)\right| &\le \left(\int_y^{y_*\wedge(t+\sqrt{n}t)} + \int_{y_*\wedge(t+\sqrt{n}t)}^{y+\sqrt{n}t}\right) \left|\widehat{A}^n(t) - \widehat{A}^n\left(t - \frac{l-y}{\sqrt{n}}\right)\right| G_v^n(dl) \\ &\le \frac{1}{\mu_n} \sup_{0\le s\le y_*-y} \left|\widehat{A}^n(t) - \widehat{A}^n\left(\left(t - \frac{s}{\sqrt{n}}\right)^+\right)\right| + 2C\left(\frac{1}{\mu_n} - G_v^n(y_*)\right) \\ &\le \frac{w^n(\delta_0)}{\mu_n} + 2C\frac{\epsilon}{4C} \le \epsilon, \end{aligned}$$

so $I_{1,1}^n \Rightarrow 0$ as claimed. Thus, by (4.9)–(4.12), to show (4.4), it suffices to verify that

$$\widehat{A}^n(t)\left(\frac{1}{\mu_n} - G_v^n\big(y+\sqrt{n}t\big)\right) \Rightarrow 0 \tag{4.16}$$

in $D[0,T]$. Let $\epsilon > 0$. By (2.23) and the fact that $A^*(0) = 0$, there exist $t_0 > 0$ and $n_1 \in \mathbb{N}$ such that $\mathbb{P}(C_n) \geq 1 - \frac{\epsilon}{2}$ for every $n \geq n_1$, where $C_n = \left[\sup_{0 \leq t \leq t_0} \left|\widehat{A}^n(s)\right| \leq \epsilon\mu_n\right]$. By (2.7), the sequence $\{G_v^n\}$ is tight, so there exists $n_2 \in \mathbb{N}$ such that for every $n \geq n_2$, $\frac{1}{\mu_n} - G_v^n\big(y + \sqrt{n}t_0\big) \leq \frac{\epsilon}{C}$, where C is the constant appearing in the definition of A_n. For every $n \geq n_1 \vee n_2$, $\mathbb{P}(A_n \cap C_n) \geq 1 - \epsilon$ and

$$
\begin{aligned}
&\sup_{0 \leq t \leq T} \left\{ \left|\widehat{A}^n(t)\right| \left(\frac{1}{\mu_n} - G_v^n\big(y + \sqrt{n}t\big) \right) \right\} \\
&\qquad = \left(\sup_{0 \leq t \leq t_0} \vee \sup_{t_0 < t \leq T} \right) \left\{ \left|\widehat{A}^n(t)\right| \left(\frac{1}{\mu_n} - G_v^n\big(y + \sqrt{n}t\big) \right) \right\} \\
&\qquad \leq \left(\epsilon\mu_n \frac{1}{\mu_n} \right) \vee \left(C\frac{\epsilon}{C} \right) = \epsilon
\end{aligned}
$$

on $A_n \cap C_n$, so (4.16) holds. We have proved (4.4).

The next step is to show

$$I_{2,2}^n \Rightarrow 0 \tag{4.17}$$

in $D[0,T]$. Let us define a random field

$$Y^n(s,y) \triangleq \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} M_j^n(y) \tag{4.18}$$

with $s \geq 0$, $y \in \overline{\mathbb{R}}$. We need

Lemma 4.2. *There exists a random field Y such that for every $T' > 0$, Y is tight in $\ell^\infty([0,T'] \times \overline{\mathbb{R}})$ and $Y^n \Rightarrow Y$ in $\ell^\infty([0,T'] \times \overline{\mathbb{R}})$.*

Proof of Lemma 4.2. Fix $T' > 0$. Let $\mathcal{F} = [0,T'] \times \overline{\mathbb{R}}$. For each $n \in \mathbb{N}$, let $m_n = \lfloor nT' \rfloor$. For $n \in \mathbb{N}$ and $j = 1, \ldots, m_n$, let us consider random fields

$$Z_{nj}(s,y) \triangleq \frac{1}{\sqrt{n}} v_j^n \mathbb{I}_{\{L_j^n \leq \sqrt{n}y\}} \mathbb{I}_{\{j \leq ns\}}, \qquad (s,y) \in \mathcal{F}.$$

For each n, $Z_{n1}, \ldots, Z_{nm_n}$ are independent with finite second moments. We have $||Z_{nj}||_{\mathcal{F}} = \frac{1}{\sqrt{n}} v_j^n$, so, by (2.15), for every $\eta > 0$

$$
\begin{aligned}
\sum_{j=1}^{m_n} \mathbb{E}\left[||Z_{nj}||_{\mathcal{F}} \mathbb{I}_{\{||Z_{nj}||_{\mathcal{F}} > \eta\}} \right] &\leq \frac{1}{\eta} \sum_{j=1}^{m_n} \mathbb{E}\left[||Z_{nj}||_{\mathcal{F}}^2 \mathbb{I}_{\{||Z_{nj}||_{\mathcal{F}} > \eta\}} \right] \\
&= \frac{\lfloor nT' \rfloor}{n\eta} \mathbb{E}\left[(v_j^n)^2 \mathbb{I}_{\{v_j^n > \sqrt{n}\eta\}} \right] \to 0.
\end{aligned} \tag{4.19}
$$

For $(s,x), (t,y) \in \mathcal{F}$, let $\rho_1((s,x),(t,y)) \triangleq |s-t| + \rho(x,y)$. It is easy to see that total boundedness of $\big(\overline{\mathbb{R}}, \rho\big)$ implies that $(\mathcal{F}, \rho_1)$ is a totally bounded semimetric space. Let $C_1 = \sup_{n\in\mathbb{N}}\left((\beta^n)^2 + \frac{1}{\mu_n^2}\right)$, $C_2 = 2(T' + C_1)$. Let $(s,x),(t,y) \in \mathcal{F}$. To fix ideas, assume $x \le y$. Then

$$\begin{aligned}
&\sum_{j=1}^{m_n} \mathbb{E}\left(Z_{nj}(s,x) - Z_{nj}(t,y)\right)^2 \\
&\qquad \le \frac{2}{n}\sum_{j=1}^{\lfloor nT'\rfloor} \mathbb{E}\left[(v_j^n)^2 \mathbb{I}_{\{L_j^n \le \sqrt{n}y\}} \mathbb{I}_{\{n(s\wedge t) < j \le n(s\vee t)\}}\right] \\
&\qquad\quad + \frac{2\lfloor nT'\rfloor}{n}\mathbb{E}\left[(v_1^n)^2 \mathbb{I}_{\{L_1^n \in \sqrt{n}(x,y]\}}\right] \qquad (4.20)\\
&\qquad \le \frac{2|\lfloor nt\rfloor - \lfloor ns\rfloor|}{n}\mathbb{E}(v_j^n)^2 + 2T'\left(G_{v^2}^n(y) - G_{v^2}^n(x)\right) \\
&\qquad \le C_2\rho_1((s,x),(t,y)) + \frac{2C_1}{n},
\end{aligned}$$

so for every sequence $\delta_n \downarrow 0$, we have

$$\sup_{\rho_1((s,x),(t,y))<\delta_n} \sum_{j=1}^{m_n} \mathbb{E}\left(Z_{nj}(s,x) - Z_{nj}(t,y)\right)^2 \to 0. \qquad (4.21)$$

For every $n \in \mathbb{N}$ and $\epsilon > 0$, define the *bracketing number* $N_{[]}(\epsilon, \mathcal{F}, L_n^2)$ [1] as the minimal number of sets N_ϵ in a partition $\mathcal{F} = \bigcup_{i=1}^{N_\epsilon} \mathcal{F}_{\epsilon,i}^n$ of the set $\mathcal{F}$ such that for every partitioning set $\mathcal{F}_{\epsilon,i}^n$

$$\sum_{j=1}^{m_n} \mathbb{E} \sup_{(s,x),(t,y)\in\mathcal{F}_{\epsilon,i}^n} \left(Z_{nj}(s,x) - Z_{nj}(t,y)\right)^2 \le \epsilon^2.$$

We want to show that every sequence $\delta_n \downarrow 0$,

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_n^2)}\, d\epsilon \to 0. \qquad (4.22)$$

Fix $\epsilon \in (0,1)$. Let $x_1^n, \dots, x_{k_n}^n$ be all the atoms of $G_{v^2}^n$ of size at least $\frac{\epsilon^2}{2C_2}$. Observe that $k_n \le \frac{2C_1C_2}{\epsilon^2}$, because the total mass of $G_{v^2}^n$ is $\mathbb{E}(v_j^n)^2 = (\beta^n)^2 + \frac{1}{\mu_n^2} \le C_1$. Let, for $i = 1,\dots,k_n$, $\tilde{\mathcal{F}}_{\epsilon,i}^n = \{x_i^n\}$ and let $A_n = \{x_1^n,\dots,x_{k_n}^n\}$. For $y \in \overline{\mathbb{R}}$, let $\tilde{G}_{v^2}^n(y) = G_{v^2}^n(y) - \sum_{k=1}^{k_n}\left(G_{v^2}^n(x_k^n) - G_{v^2}^n(x_k^n-)\right)\mathbb{I}_{\{x_k^n \le y\}}$. Let $l_n = \left\lceil \frac{2C_2\tilde{G}_{v^2}^n(\infty)}{\epsilon^2}\right\rceil \vee 1$. We have $1 \le l_n \le \frac{2C_1C_2}{\epsilon^2} + 1$. If $l_n > 1$, take $y_i^n = \left(\tilde{G}_{v^2}^n\right)^{-1}\left(\frac{\epsilon^2}{2C_2}\right)$, $i = 1,\dots,l_n - 1$, where $\left(\tilde{G}_{v^2}^n\right)^{-1}(y) \triangleq \inf\{\theta \in \mathbb{R} : \tilde{G}_{v^2}^n(\theta) \ge y\}$. Observe that $\tilde{G}_{v^2}^n(y_1^n) \le \frac{\epsilon^2}{C_2}$ and $\tilde{G}_{v^2}^n(y_{i+1}^n) - \tilde{G}_{v^2}^n(y_i^n) \le \frac{\epsilon^2}{C_2}$ for $i = 1,\dots,l_n - 2$, because $\tilde{G}_{v^2}^n$ has no atoms of size bigger than or

[1] This notation, although complicated, is used in the theory of empirical processes, see van der Vaart and Wellner [14].

equal to $\frac{\epsilon^2}{2C_2}$. Let $z_1^n < \cdots < z_{k_n+l_n-1}^n$ be such that $\{z_1^n, \ldots, z_{k_n+l_n-1}^n\} = A_n \cup \{y_1^n, \ldots, y_{l_n-1}^n\}$, where $\{y_1^n, \ldots, y_{l_n-1}^n\} = \emptyset$ by definition if $l_n = 1$. Take $\tilde{\mathcal{F}}_{\epsilon,k_n+1}^n = [-\infty, z_1^n] \setminus A_n$, $\tilde{\mathcal{F}}_{\epsilon,k_n+i+1}^n = (z_i^n, z_{i+1}^n] \setminus A_n$, $i = 1, \ldots, k_n + l_n - 2$, $\tilde{\mathcal{F}}_{\epsilon,2k_n+l_n}^n = (z_{k_n+l_n-1}^n, \infty]$. By construction, $\overline{\mathbb{R}} = \bigcup_{i=1}^{2k_n+l_n} \tilde{\mathcal{F}}_{\epsilon,i}^n$ and

$$\sup_{x,y \in \tilde{\mathcal{F}}_{\epsilon,i}^n} \left| G_{v^2}^n(x) - G_{v^2}^n(y) \right| \le \frac{\epsilon^2}{C_2}, \qquad i = 1, \ldots, 2k_n + l_n. \tag{4.23}$$

If $n < \frac{2C_2}{\epsilon^2}$, let $p_n = \lfloor nT' \rfloor + 1$. Then $p_n \le (T'+1)n < \frac{2C_2(T'+1)}{\epsilon^2}$. In this case, let $B_k = [\frac{k-1}{n}, \frac{k}{n}) \cap [0, T']$, $k = 1, \ldots, p_n$. If $n \ge \frac{2C_2}{\epsilon^2}$, then $\frac{1}{n} \le \frac{\epsilon^2}{2C_2}$. In this case, let $p_n = \lfloor \frac{2C_2 T'}{\epsilon^2} \rfloor + 1$, $B_k = \left[\frac{(k-1)\epsilon^2}{2C_2}, \frac{k\epsilon^2}{2C_2}\right) \cap [0, T']$, $k = 1, \ldots, p_n$. Observe that, in any case, $p_n \le \frac{2(C_2 \vee 1)(T'+1)}{\epsilon^2}$, $[0, T'] = \bigcup_{k=1}^{p_n} B_k$ and

$$\sup_{t_1,t_2 \in B_k} \frac{|\lfloor nt_1 \rfloor - \lfloor nt_2 \rfloor|}{n} \le \frac{\epsilon^2}{C_2}, \qquad k = 1, \ldots, p_n. \tag{4.24}$$

Indeed, if $n < \frac{2C_2}{\epsilon^2}$, then the LHS of (4.24) is 0, otherwise for $t_1, t_2 \in B_k$, $\frac{|\lfloor nt_1 \rfloor - \lfloor nt_2 \rfloor|}{n} \le \frac{|nt_1 - nt_2| + 1}{n} = |t_1 - t_2| + \frac{1}{n} \le \frac{\epsilon^2}{2C_2} + \frac{\epsilon^2}{2C_2} = \frac{\epsilon^2}{C_2}$. Now, for $k = 1, \ldots, p_n$, $i = 1, \ldots, 2k_n + l_n$, let $\mathcal{F}_{\epsilon,k,i} = B_k \times \tilde{\mathcal{F}}_{\epsilon,i}^n$. We have $\mathcal{F} = \bigcup_{k=1}^{p_n} \bigcup_{i=1}^{2k_n+l_n} \mathcal{F}_{\epsilon,k,i}$ and $p_n(2k_n + l_n) \le \frac{C_3}{\epsilon^4}$, where $C_3 = 2(C_2 \vee 1)(T' + 1)(6C_1C_2 + 1)$. Proceeding as in (4.20) and using (4.23)–(4.24), we can check that for $k = 1, \ldots, p_n$, $i = 1, \ldots, 2k_n + l_n$,

$$\begin{aligned}
&\sum_{j=1}^{m_n} \mathbb{E} \sup_{(s,x),(t,y) \in \mathcal{F}_{\epsilon,k,i}^n} (Z_{nj}(s,x) - Z_{nj}(t,y))^2 \\
&\qquad \le 2T' \sup_{x,y \in \tilde{\mathcal{F}}_{\epsilon,i}^n} \left| G_{v^2}^n(x) - G_{v^2}^n(y) \right| + 2C_1 \sup_{t_1,t_2 \in B_k} \frac{|\lfloor nt_1 \rfloor - \lfloor nt_2 \rfloor|}{n} \le \epsilon^2.
\end{aligned}$$

Thus, for all n and $\epsilon \in (0,1)$, $N_{[]}(\epsilon, \mathcal{F}, L_n^2) \le \frac{C_3}{\epsilon^4}$, so for δ_n small enough,

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_n^2)}\, d\epsilon \le \int_0^{\delta_n} \sqrt{C_3 - 4 \log \epsilon}\, d\epsilon \le \sqrt{5} \int_0^{\delta_n} \sqrt{|\log \epsilon|}\, d\epsilon,$$

so (4.22) holds.

It is easy to see that the finite-dimensional distributions of Y^n converge. Thus, by (4.19), (4.21), (4.22) and the bracketing central limit theorem (Theorem 2.11.9 in van der Vaart and Wellner [14]), the sequence of random fields

$$Y^n(s,t) = \sum_{j=1}^{\lfloor nT' \rfloor} \frac{1}{\sqrt{n}} \left(v_j^n \mathbb{I}_{\{L_j^n \le \sqrt{n} y\}} - G_v^n(y) \right) \mathbb{I}_{\{j \le ns\}} = \sum_{j=1}^{m_n} (Z_{nj}(s,y) - \mathbb{E} Z_{nj})$$

converges weakly to a tight random field Y in $\ell^\infty([0, T'] \times \overline{\mathbb{R}})$. $\square$

Corollary 4.3. *For every $T' > 0$, the sequence Y^n is asymptotically uniformly ρ_1-equicontinuous in probability on $\mathcal{F}$, i.e., for every $\epsilon, \eta > 0$ there exists $\delta > 0$ such that* $\limsup_{n\to\infty} \mathbb{P}[||Y^n||_{\mathcal{F}_\delta} > \epsilon] < \eta$, *where*

$$||Y^n||_{\mathcal{F}_\delta} \triangleq \sup\{|Y^n(s,x) - Y^n(t,y)| : (s,x),(t,y) \in \mathcal{F},\ \rho_1((s,x),(t,y)) < \delta\}.$$

This follows immediately from the proof of the bracketing central limit theorem (see van der Vaart and Wellner [14], pp. 217–220).

Returning to the proof of Proposition 4.1, let us observe that, by (4.15),

$$I_{2,2}^n(t) = Y^n\left(\frac{1}{n}A^n(nt), \infty\right) - Y^n\left(\frac{1}{n}A^n(nt), y + \sqrt{n}t\right). \tag{4.25}$$

For every $\delta > 0$, we have $||I_{2,2}^n||_{[0,T]} = ||I_{2,2}^n||_{[0,\delta]} \vee ||I_{2,2}^n||_{[\delta,T]}$. By (2.23), $\mathbb{P}(D_n) \to 1$, where $D_n = [A^n(nt) \le n(\lambda+1)t$ for every $t \in [0,T]]$. Let $T' = (\lambda+1)T$. By definition, $Y^n(0,\cdot) \equiv 0$, so, by (4.25), on D_n,

$$\begin{aligned} ||I_{2,2}^n||_{[0,\delta]} &\le \sup_{t\in[0,\delta]} \left|Y^n(0,\infty) - Y^n\left(\frac{1}{n}A^n(nt), \infty\right)\right| \\ &\quad + \sup_{t\in[0,\delta]} \left|Y^n\big(0, y+\sqrt{n}t\big) - Y^n\left(\frac{1}{n}A^n(nt), y+\sqrt{n}t\right)\right| \\ &\le 2||Y^n||_{\mathcal{F}_{(\lambda+1)\delta}}, \end{aligned} \tag{4.26}$$

$$\begin{aligned} ||I_{2,2}^n||_{[\delta,T]} &= \sup_{t\in[\delta,T]} \left|Y^n\left(\frac{1}{n}A^n(nt), \infty\right) - Y^n\left(\frac{1}{n}A^n(nt), y+\sqrt{n}t\right)\right| \\ &\le \sup_{t\in[0,T'], s\ge\delta} \left|Y^n(t,\infty) - Y^n\big(t, y+\sqrt{n}s\big)\right| \\ &\le ||Y^n||_{\mathcal{F}_{\rho(y+\sqrt{n}\delta,\infty)}}. \end{aligned} \tag{4.27}$$

By (2.8) and (2.16), $\lim_{x\uparrow\infty} \rho(x,\infty) = 0$, so, by Corollary 4.3, (4.26), (4.27) and the fact that $\mathbb{P}(D_n) \to 1$, we have (4.17). Therefore, by (4.8), (4.13), (4.4) and (4.17), to prove (4.1), it suffices to show that

$$I_{2,1}^n \Rightarrow 0 \tag{4.28}$$

in $D[0,T]$. To this end, it suffices to prove the following two lemmas concerning the process U^n defined in (4.7).

Lemma 4.4. $U^n \Rightarrow 0$ *in* $D[0,T]$.

Lemma 4.5. $R^n \triangleq I_{2,1}^n - U^n \Rightarrow 0$ *in* $D[0,T]$.

Proof of Lemma 4.4. For $s,t \in [0,T]$, let us define a random field

$$X^n(s,t) \triangleq \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda_n nt \rfloor} M_j^n\left(y + \sqrt{n}s - \frac{j}{\lambda_n\sqrt{n}}\right).$$

By straightforward, but tedious, computations one may check that for every pair of neighbouring blocks $B_1, B_2 \subseteq [0,T]^2$,

$$\mathbb{E}\left[(X^n(B_1))^2\,(X^n(B_2))^2\right] \le m^n(B_1)m^n(B_2), \tag{4.29}$$

where m^n are finite, positive measures on $[0,T]^2$ defined by

$$\begin{aligned} m^n(B) \triangleq \frac{C}{n} \sum_{j=\lfloor \lambda_n n t_1\rfloor+1}^{\lfloor \lambda_n n t_2\rfloor} \Bigg[& G_v^n\left(y+\sqrt{n}s_2-\frac{j}{\lambda_n\sqrt{n}}\right)-G_v^n\left(y+\sqrt{n}s_1-\frac{j}{\lambda_n\sqrt{n}}\right) \\ & +G_{v^2}^n\left(y+\sqrt{n}s_2-\frac{j}{\lambda_n\sqrt{n}}\right)-G_{v^2}^n\left(y+\sqrt{n}s_1-\frac{j}{\lambda_n\sqrt{n}}\right)\Bigg] \end{aligned}$$

for any $B=(s_1,s_2]\times(t_1,t_2]\subseteq[0,T]^2$ and

$$C = 1 \vee \sup_{n\in\mathbb{N}}(1/\mu_n). \tag{4.30}$$

Of course, (4.29) remains true if we replace X^n by $\tilde{X}^n$, where, for $s,t\in[0,T]$, $\tilde{X}^n(s,t)=X^n(s,t)-X^n(0,t)$. We have $\tilde{X}^n(0,\cdot)\equiv\tilde{X}^n(\cdot,0)\equiv 0$. Using (2.7)–(2.8), one can check that $m^n\Rightarrow m$, where, for B as above,

$$m(B)=C\left(1+\lambda\beta^2+\frac{1}{\lambda}\right)\left((s_2\wedge t_2)-(s_1\vee t_1)\right)^+.$$

In particular, m has continuous marginals. By Theorem 3 in Bickel and Wichura [2] (strictly speaking, by its extension described on pp. 1665–1666 of that paper), the sequence $\{\tilde{X}^n\}$ is tight in $D([0,T]^2)$. However, the stochastic processes $X^n(0,t)$ converge weakly in $D([0,T])^2$ by Theorem 3.1 in Prokhorov [13], so the sequence $\{X^n\}$ is also tight in $D([0,T]^2)$. In particular, the sequence $\{X^n(t,t)\}$ is tight in $D[0,T]$. For $0\le t\le T$, $U^n(t)=Y^n\big(\lambda_n t, y+\sqrt{n}t\big)-X^n(t,t)$ (recall that Y^n was defined by (4.18)). As in (4.25)–(4.27), we show that

$$||Y^n(\lambda_n t,\infty)-Y^n\big(\lambda_n t, y+\sqrt{n}t\big)||_{[0,T]}\Rightarrow 0. \tag{4.31}$$

By Theorem 3.1 in Prokhorov [13], $Y^n(\lambda_n t,\infty)$ converges weakly in $D[0,T]$. Thus, the sequence $\{U^n\}$ is tight in $D[0,T]$. Also, by (4.7), for any $t\ge 0$,

$$\begin{aligned} \mathbb{E}(U^n(t))^2 &\le \frac{1}{n}\sum_{j=1}^{\lfloor\lambda_n n t\rfloor}\mathbb{E}\left[(v_j^n)^2\mathbb{I}_{\left\{y+\sqrt{n}t-\frac{j}{\lambda_n\sqrt{n}}<\frac{L_j^n}{\sqrt{n}}\le y+\sqrt{n}t\right\}}\right] \\ &\le \frac{1}{n}\sum_{j=1}^{\lfloor\lambda_n n t\rfloor}\left(G_{v^2}^n(\infty)-G_{v^2}^n\left(y+\sqrt{n}t-\frac{j}{\lambda_n\sqrt{n}}\right)\right) \\ &\approx \lambda_n\int_0^t\left(G_{v^2}^n(\infty)-G_{v^2}^n\left(y+\sqrt{n}(t-s)\right)\right)ds. \end{aligned} \tag{4.32}$$

For a fixed $s \in [0,t)$, $G^n_{v^2}(\infty) - G^n_{v^2}\left(y + \sqrt{n}(t-s)\right) \to 0$ as $n \to \infty$ by (2.8). Thus, by the bounded convergence theorem, (4.32) implies that $U^n(t) \to 0$ in L^2 for any t, so, $U^n \Rightarrow 0$ in $D[0,T]$ as claimed. $\square$

Proof of Lemma 4.5. By (4.14), (4.7) and the definition of R^n, for $t \geq 0$,

$$R^n(t) = \frac{1}{\sqrt{n}} \int_0^t \left(\sum_{j=A^n(n(t-s))+1}^{A^n(nt)} - \sum_{j=\lfloor \lambda_n n(t-s) \rfloor +1}^{\lfloor \lambda_n nt \rfloor} \right) M^n_j\big(y + \sqrt{n} ds\big) \tag{4.33}$$
$$= R^n_1(t) - R^n_2(t),$$

where

$$R^n_1(t) \triangleq \frac{1}{\sqrt{n}} \int_0^t \left(\sum_{j=1}^{A^n(nt)} - \sum_{j=1}^{\lfloor \lambda_n nt \rfloor} \right) M^n_j\big(y + \sqrt{n} ds\big) \tag{4.34}$$
$$= \frac{1}{\sqrt{n}} \left(\sum_{j=1}^{A^n(nt)} - \sum_{j=1}^{\lfloor \lambda_n nt \rfloor} \right) \left(M^n_j\big(y + \sqrt{n}t\big) - M^n_j(y)\right),$$

$$R^n_2(t) \triangleq \frac{1}{\sqrt{n}} \int_0^t \left(\sum_{j=1}^{A^n(n(t-s))} - \sum_{j=1}^{\lfloor \lambda_n n(t-s) \rfloor} \right) M^n_j\big(y + \sqrt{n} ds\big). \tag{4.35}$$

By (4.15) and (4.34), we have

$$R^n_1(t) = \widehat{V}^n\left(\frac{1}{n}A^n(nt)\right) - I^n_{2,2}(t) - Y^n\big(\lambda_n t, y + \sqrt{n}t\big)$$
$$+ \widehat{V}^n_y\left(\frac{1}{n}A^n(nt)\right) - \widehat{V}^n_y\left(\lambda_n t\right),$$

where

$$\widehat{V}^n_y(t) \triangleq \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} M^n_j(y) \Rightarrow V^*_y \tag{4.36}$$

for a Brownian motion V^*_y by Theorem 3.1 in Prokhorov [13]. Thus, by (2.23), (4.17), (4.31), the fact that $Y^n(\lambda_n t, \infty) = \widehat{V}^n(\lambda_n t)$ and the Differencing Theorem,

$$R^n_1(t) = \widehat{V}^n\left(\frac{1}{n}A^n(nt)\right) - \widehat{V}^n(\lambda_n t) \tag{4.37}$$
$$+ \widehat{V}^n_y\left(\frac{1}{n}A^n(nt)\right) - \widehat{V}^n_y\left(\lambda_n t\right) + o(1) \Rightarrow 0.$$

For $0 \le s \le t \le T$, we have $j \le A^n(n(t-s))$ if and only if $S_j^n \le n(t-s)$, which, in turn, is equivalent to $s \le t - S_j^n/n$. Thus,

$$\begin{aligned}
&\int_0^t \sum_{j=1}^{A^n(n(t-s))} M_j^n\big(y+\sqrt{n}ds\big) \\
&= \sum_{j=1}^{A^n(nt)} \left[M_j^n\left(y+\sqrt{n}t-\frac{1}{\sqrt{n}}S_j^n\right)-M_j^n(y)\right] \qquad (4.38)\\
&= \sum_{j=1}^{A^n(nt)} \left[M_j^n\left(y+\sqrt{n}t-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n\sqrt{n}}\right)\right)-M_j^n(y)\right],
\end{aligned}$$

where, for $u \ge 0$, $H^n(u) \triangleq S^n_{\lfloor \lambda_n\sqrt{n}u\rfloor}/\sqrt{n} - u$. Observe that

$$H^n \Rightarrow 0 \qquad (4.39)$$

in $D[0,\infty)$ by the functional law of large numbers and H^n is $\sigma(u_j^n, j = 1,\dots)$-measurable. Also, for $0 \le s \le t$, $j \le \lfloor \lambda_n n(t-s)\rfloor$ iff $s \le t - j/(\lambda_n n)$, so

$$\begin{aligned}
&\int_0^t \sum_{j=1}^{\lfloor \lambda_n n(t-s)\rfloor} M_j^n\big(y+\sqrt{n}ds\big) \\
&\qquad = \sum_{j=1}^{\lfloor \lambda_n nt\rfloor} \left[M_j^n\left(y+\sqrt{n}t-\frac{j}{\lambda_n\sqrt{n}}\right)-M_j^n(y)\right].
\end{aligned} \qquad (4.40)$$

By (4.35), (4.38) and (4.40),

$$\begin{aligned}
R_2^n(t) &= (Z^n-\overline{X}^n)(t,t) + \left(Z^n\left(t,\frac{1}{\lambda_n n}A^n(nt)\right)-Z^n(t,t)\right)\\
&\quad + \left(\widehat{V}_y^n\left(\frac{1}{n}A^n(nt)\right)-\widehat{V}_y^n(\lambda_n t)\right),
\end{aligned} \qquad (4.41)$$

where, for $s,t \ge 0$,

$$\overline{X}^n(s,t) \triangleq \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor \lambda_n nt\rfloor} M_j^n\left(y+\left(\sqrt{n}s-\frac{j}{\lambda_n\sqrt{n}}\right)^+\right),$$

$$Z^n(s,t) \triangleq \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor \lambda_n nt\rfloor} M_j^n\left(y+\left(\sqrt{n}s-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n\sqrt{n}}\right)\right)^+\right).$$

As in the proof of Lemma 4.4, one may check that for every pair of neighbouring blocks $B_1, B_2 \subseteq [0,T+1]^2$,

$$\mathbb{E}\left[\left(\overline{X}^n(B_1)\right)^2\left(\overline{X}^n(B_2)\right)^2\right] \le m_1^n(B_1)m_1^n(B_2), \qquad (4.42)$$

where m_1^n are finite, positive measures on $[0,T+1]^2$ defined by

$$
\begin{aligned}
m_1^n(B) \triangleq \frac{C}{n} \sum_{j=\lfloor \lambda_n n t_1 \rfloor +1}^{\lfloor \lambda_n n t_2 \rfloor} \Bigg[& G_v^n \left(y + \left(\sqrt{n} s_2 - \frac{j}{\lambda_n \sqrt{n}} \right)^+ \right) \\
& - G_v^n \left(y + \left(\sqrt{n} s_1 - \frac{j}{\lambda_n \sqrt{n}} \right)^+ \right) + G_{v^2}^n \left(y + \left(\sqrt{n} s_2 - \frac{j}{\lambda_n \sqrt{n}} \right)^+ \right) \\
& - G_{v^2}^n \left(y + \left(\sqrt{n} s_1 - \frac{j}{\lambda_n \sqrt{n}} \right)^+ \right) \Bigg]
\end{aligned}
$$

for any $B = (s_1, s_2] \times (t_1, t_2] \subseteq [0, T+1]^2$, with C given by (4.30). Also, $m_1^n \Rightarrow m_1$, where, for B as above,

$$
\begin{aligned}
m_1(B) = \lambda C \, & (G_{v^2}(\infty) - G_{v^2}(y) \\
& + G_v(\infty) - G_v(y)) \left((s_2 \wedge t_2) - (s_1 \vee t_1) \right)^+
\end{aligned}
\tag{4.43}
$$

(here we use the assumption that y is a point of continuity of G_v and G_{v^2}). This, as in the proof of Lemma 4.4, implies that the sequence $\{\overline{X}^n\}$ is tight in $D([0,T+1]^2)$. Using Theorem 3.1 in Prokhorov [13], together with the Cramer–Wold device (see, e.g., Billingsley [3]) and the Kolmogorov–Čentsov theorem (see, e.g., Karatzas and Shreve [8]), we can easily check that the finite-dimensional distributions of $\overline{X}^n$ converge to the corresponding finite-dimensional distributions of a Gaussian $\overline{X}$ with continuous sample paths. Thus, $\overline{X}^n \Rightarrow \overline{X}$ in $D([0,T+1]^2)$. Therefore, by (2.23), (4.33), (4.36), (4.37), (4.41) and continuity of the sample paths of $\overline{X}$, to finish the proof of Lemma 4.5, it suffices to show that

$$
Z^n - \overline{X}^n \Rightarrow 0 \tag{4.44}
$$

in $D([0,T+1]^2)$. (4.44) is a statement about *weak convergence* of stochastic processes, so the underlying probability spaces are irrelevant. Thus, without loss of generality we can assume that all the random variables (and thus all the stochastic processes) under consideration are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and, moreover, *all* the arrival times $\left\{u_j^n\right\}_{j,n=1}^{\infty}$ are independent of *all* the random vectors $\left\{\left(v_j^n, L_j^n\right)\right\}_{j,n=1}^{\infty}$. This is not a limiting assumption, because if, for different $n \geq 1$, the probability spaces $(\Omega^n, \mathcal{A}^n, \mathbb{P}^n)$ on which the sequences $\left\{u_j^n\right\}_{j=1}^{\infty}$, $\left\{\left(v_j^n, L_j^n\right)\right\}_{j=1}^{\infty}$ are defined, are different, we can take $(\Omega, \mathcal{A}, \mathbb{P}) = \Pi_{n=1}^{\infty}(\Omega^n, \mathcal{A}^n, \mathbb{P}^n)$. By (4.39), there exists a sequence $\epsilon_n \downarrow 0$ such that $\mathbb{P}(A_n) \to 1$, where $A_n = \left[\sup_{0 \leq s \leq T+1} |H^n(s)| \leq \epsilon_n\right]$. To prove (4.44), it suffices to show

$$
\left(Z^n - \overline{X}^n\right) \mathbb{I}_{A_n} \Rightarrow 0 \tag{4.45}
$$

in $D([0,T+1]^2)$. By straightforward, but tedious, computations we may check that for any neighbouring blocks $B_1, B_2 \subseteq [0,T+1]^2$,

$$\begin{aligned}\mathbb{E}\left[\left(Z^n(B_1)\right)^2\left(Z^n(B_2)\right)^2\mathbb{I}_{A_n}|H^n(\cdot)\right] &\le \overline{m}_2^n\left(B_1\right)\overline{m}_2^n\left(B_2\right)\mathbb{I}_{A_n} \\ &\le m_2^n\left(B_1\right)m_2^n\left(B_2\right),\end{aligned} \tag{4.46}$$

where $\overline{m}_2^n$ and m_2^n are finite, positive random measures on $[0,T+1]^2$ defined by

$$\begin{aligned}\overline{m}_2^n(B) \triangleq \frac{C}{n}\sum_{j=\lfloor\lambda_n n t_1\rfloor+1}^{\lfloor\lambda_n n t_2\rfloor}\Bigg[&G_v^n\left(y+\left(\sqrt{n}s_2-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n n}\right)\right)^+\right) \\ &- G_v^n\left(y+\left(\sqrt{n}s_1-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n n}\right)\right)^+\right) \\ &+ G_{v^2}^n\left(y+\left(\sqrt{n}s_2-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n n}\right)\right)^+\right) \\ &- G_{v^2}^n\left(y+\left(\sqrt{n}s_1-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n n}\right)\right)^+\right)\Bigg]\end{aligned}$$

for any $B=(s_1,s_2]\times(t_1,t_2]\subseteq[0,T+1]^2$ and $m_2^n \triangleq \overline{m}_2^n\mathbb{I}_{A_n}+m_1\mathbb{I}_{A_n^c}$, where C and m_1 are defined by (4.30) and (4.43). Using (2.7)–(2.8), we can check that for every $\omega\in\Omega$, $m_2^n(\omega)\Rightarrow m_1$. As in the proof of Lemma 4.4, this implies conditional tightness of $\{Z^n\mathbb{I}_{A_n}\}$ with respect to $H^n(\cdot)$ $(\mathcal{F})$ in $D([0,T+1]^2)$. The random fields $\overline{X}^n$ are independent on $\mathcal{F}$ and $\overline{X}^n\Rightarrow\overline{X}$, so $\{(Z^n-\overline{X}^n)\mathbb{I}_{A_n}\}$ are also conditionally tight with respect to $H^n(\cdot)$ $(\mathcal{F})$ in $D([0,T+1]^2)$. Moreover, for any $s,t\ge 0$, by (2.8), the bounded convergence theorem and the fact that y is not an atom of G_{v^2}, we have

$$\begin{aligned}&\mathbb{E}\Big[\left(Z^n-\overline{X}^n\right)^2\mathbb{I}_{A_n}|\mathcal{F}\Big] \\ &\qquad = \frac{1}{n}\sum_{j=1}^{\lfloor\lambda_n n t\rfloor}\Bigg|G_{v^2}^n\left(y+\left(\sqrt{n}s-\frac{j}{\lambda_n\sqrt{n}}-H^n\left(\frac{j}{\lambda_n n}\right)\right)^+\right) \\ &\qquad\qquad - G_{v^2}^n\left(y+\left(\sqrt{n}s-\frac{j}{\lambda_n\sqrt{n}}\right)^+\right)\Bigg| \\ &\qquad \approx \lambda_n\int_0^t\Big|G_{v^2}^n\left(y+\left(\sqrt{n}(s-u-H^n(u)/\sqrt{n}\right)^+\right) \\ &\qquad\qquad - G_{v^2}^n\left(y+\left(\sqrt{n}(s-u\right)^+\right)\Big|\,du\end{aligned}$$

$$\leq \lambda_n \int_0^t \left[G_{v^2}^n \left(y + \sqrt{n}(s-u)^+ + \epsilon_n \right) \right.$$
$$\left. - \; G_{v^2}^n \left(y + \sqrt{n}(s-u)^+ - \epsilon_n \right) \right] du \to 0.$$

Thus gives the conditional convergence (4.45) with respect to $\mathcal{F}$, so (4.45) holds unconditionally. □

The proof of Proposition 4.1 is now complete. □

Using a similar (but simpler) argument, we can prove

Proposition 4.6. *Let $T > 0$ and let y be a point of continuity of G. Then*

$$\sup_{0 \leq t \leq T} \left| \widehat{\mathcal{A}}^n(t)(y,\infty) - H^n(y) + H^n\big(y + \sqrt{n}t\big) \right| \xrightarrow{P} 0. \tag{4.47}$$

Propositions 4.1 and 4.6 can be refined to

Proposition 4.7. *For every $T > 0$ and $y_0 \in \mathbb{R}$, we have*

$$\sup_{y \geq y_0} \sup_{0 \leq t \leq T} \left| \widehat{\mathcal{V}}^n(t)(y,\infty) - H_v^n(y) + H_v^n\big(y + \sqrt{n}t\big) \right| \xrightarrow{P} 0, \tag{4.48}$$

$$\sup_{y \geq y_0} \sup_{0 \leq t \leq T} \left| \widehat{\mathcal{A}}^n(t)(y,\infty) - H^n(y) + H^n\big(y + \sqrt{n}t\big) \right| \xrightarrow{P} 0. \tag{4.49}$$

Proof. We will show (4.48), the proof of (4.49) is similar. Without loss of generality we can assume that y_0 is a point of continuity of both G_v and G_{v^2} (if not, change it to $y_0' < y_0$ with this property). Let $\epsilon > 0$ be arbitrary. By (2.17), we can choose $y^0 \in (y_0, \infty)$, a point of continuity of both G_v and G_{v^2}, such that $\sup_{n\geq 1} H_v^n(y^0) \leq \frac{\epsilon}{4}$. Choose a partition $y_0 < y_1 < \cdots < y_M = y^0$ such that y_m is a point of continuity of both G_v and G_{v^2} and $|y_{m+1} - y_m| \leq \frac{\epsilon}{2\sup_{n\geq 1}\rho_n}$ for $m = 0, \ldots, M-1$. Observe that for such m and all n,

$$0 \leq H_v^n(y_m) - H_v^n(y_{m+1}) \leq \rho_n |y_{m+1} - y_m| \leq \frac{\epsilon}{2}. \tag{4.50}$$

By Proposition 4.1, there exists n_0 such that for $m = 0, \ldots, M$ and all $n \geq n_0$, $\mathbb{P}(B_{n,m}) \leq \frac{\epsilon}{4(M+1)}$, where

$$B_{n,m} = \left[\sup_{0\leq t\leq T} \left| \widehat{\mathcal{V}}^n(t)(y_m,\infty) - H_v^n(y_m) + H_v^n\big(y_m + \sqrt{n}t\big) \right| \geq \frac{\epsilon}{2} \right].$$

Using (4.50) and proceeding as in the proof of Proposition 3.4 in Doytchinov et al. [6], we show that $\mathbb{P}(B_n) \leq \frac{\epsilon}{2}$, where

$$B_n = \left[\sup_{y_0 \leq y \leq y^0} \sup_{0\leq t\leq T} \left| \widehat{\mathcal{V}}^n(t)(y,\infty) - H_v^n(y) + H_v^n\big(y + \sqrt{n}t\big) \right| \geq \epsilon \right].$$

For $y \geq y^0$ and $t \in [0,T]$, on $B^c_{n,M}$,

$$
\begin{aligned}
0 &\leq \widehat{\mathcal{V}}^n(t)(y,\infty) \leq \widehat{\mathcal{V}}^n(t)(y_0,\infty) \\
&= \Big(\widehat{\mathcal{V}}^n(t)(y_0,\infty) - H^n_v(y^0) + H^n_v\big(y^0+\sqrt{n}t\big)\Big) + H^n_v(y^0) - H^n_v\big(y^0+\sqrt{n}t\big) \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{4} = \frac{3\epsilon}{4},
\end{aligned}
$$

so

$$
-\frac{\epsilon}{4} \leq \widehat{\mathcal{V}}^n(t)(y,\infty) - H^n_v(y) + H^n_v\big(y+\sqrt{n}t\big) < \frac{3\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.
$$

Let $B = B^c_n \cap B^c_{n,M}$. Then, $P(B) \geq 1-\epsilon$ and

$$
B \subseteq \left[\sup_{y\geq y_0}\sup_{0\leq t\leq T}\left|\widehat{\mathcal{V}}^n(t)(y,\infty) - H^n_v(y) + H^n_v\big(y+\sqrt{n}t\big)\right| < \epsilon\right]. \qquad \square
$$

Corollary 4.8. *For every $T > 0$ and $y_0 \in \mathbb{R}$, we have*

$$
\sup_{y\geq y_0}\sup_{0\leq t\leq T}\widehat{\mathcal{V}}^n(t)\{y\} \xrightarrow{P} 0, \qquad \sup_{y\geq y_0}\sup_{0\leq t\leq T}\widehat{\mathcal{A}}^n(t)\{y\} \xrightarrow{P} 0. \tag{4.51}
$$

This follows from Proposition 4.7 in the same way as Corollary 3.5 in Doytchinov et al. [6] follows from Proposition 3.4 there.

Remark 4.9. In this section, the assumption (2.13) and the Lindeberg condition on $(L^n_j)^+/\sqrt{n}$ were not used.

5. Customers behind the frontiers. In this section, we prove that the work in the n-th system at time nt associated with customers in this system having lead times smaller than or equal to $F^n(nt)$ becomes negligible (Corollary 5.4) and that the number of these customers is also negligible provided that the workload is not too small (Corollary 5.7). (Later we will show that if the workload is small, then the number of customers in the system is also small, see the proof of Theorem 3.2.) This is to be expected, because under the EDF queue discipline these customers form a “high priority class”, arriving at the system at a rate less than 1 and being served with rate 1, thus experiencing low traffic conditions. We formalize this idea with the help of Lemmas 5.3 and 5.6, estimating the last time in which a customer with lead time equal to the frontier was served. Along the way, we establish Lemmas 5.1 and 5.5, showing tightness of the rescaled frontiers (the upper bound, given by Lemma 5.5, requires that the workload is bounded below). The latter results follow from Proposition 4.7, together with the observation that the customers with lead times greater than the current frontier have not received any service. Indeed, if $\widehat{F}^n(t)$ is very large negative, then, by Proposition 4.7, the workload $\widehat{W}^n(t)$, bounded below by $\widehat{\mathcal{V}}^n(t)(\widehat{F}^n(t),\infty)$, is very large, contradicting tightness of the sequence $\{\widehat{W}^n(t)\}$ following from (2.24). If, on the other hand, $\widehat{F}^n(t)$ is very large, then, because the work associated with customers with lead times not greater than the frontier is negligible, the workload $\widehat{W}^n(t) \approx \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t), \infty\big)$ is very small (again by Proposition 4.7), contrary to the assumption that the workload is not too small.

Lemma 5.1. *For every $T > 0$ and $\epsilon > 0$, there exists $C \in \mathbb{R}$ such that for all n,*

$$\mathbb{P}\left[\inf_{0 \le t \le T} \widehat{F}^n(t) \ge C\right] \ge 1 - \epsilon. \tag{5.1}$$

Proof. None of the customers with lead times at time nt greater than $F^n(nt)$ has received service by time nt, so, by (2.24),

$$\sup_{0 \le t \le T} \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t), \infty\big) \le \sup_{0 \le t \le T} \widehat{W}^n(t) \Rightarrow \sup_{0 \le t \le T} W^*(t). \tag{5.2}$$

Recall that $H_v(y) > 0$ for all y, $\lim_{y \to -\infty} H_v(y) = \infty$, the functions H_v^n are decreasing and such that $H_v^n \to H_v$. By these properties, together with (5.2), we can choose $C_1 \in \mathbb{R}$ such that for all n, $H_v^n(C_1) \ge (24/\epsilon)^2$ and $\mathbb{P}(A_n) \ge 1 - \epsilon/6$, where

$$A_n = \left[\sup_{0 \le t \le T} \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t), \infty\big) \le \sqrt{H_v^n(C_1)}\right].$$

By Proposition 4.7, for $n \ge n_0$, $\mathbb{P}(B_n) \ge 1 - \epsilon/6$, where

$$B_n = \left[\inf_{0 \le t \le T} \left\{\widehat{\mathcal{V}}^n(t)(C_1, \infty) + H_v^n\big(C_1 + \sqrt{n}t\big)\right\} \ge H_v^n(C_1)/2\right].$$

Let $c > 0$ be such that $H_v^n(C_1 + c) \le H_v^n(C_1)/4$ for all n. In particular, on B_n,

$$\inf_{c/\sqrt{n} \le t \le T} \widehat{\mathcal{V}}^n(t)(C_1, \infty) \ge H_v^n(C_1)/4.$$

Proceeding as in the proof of Lemma 3.7 in Doytchinov et al. [6], we get, for $n \ge n_0$,

$$\mathbb{P}\left[\left\{\inf_{c/\sqrt{n} \le t \le T} \widehat{F}^n(t) < C_1\right\} \cap A_n \cap B_n\right] \le 4/\sqrt{H_v^n(C_1)},$$

so $\mathbb{P}\left[\inf_{c/\sqrt{n} \le t \le T} \widehat{F}^n(t) < C_1\right] \le \epsilon/2$. We have

$$\inf_{0 \le t \le c/\sqrt{n}} \widehat{F}^n(t) \ge L_1^n/\sqrt{n} - c,$$

so, by (2.5)–(2.6), there exists $C_2 \in \mathbb{R}$ such that for all n,

$$\mathbb{P}\left[\inf_{0 \le t \le c/\sqrt{n}} \widehat{F}^n(t) \ge C_2\right] \ge 1 - \epsilon/2.$$

Let $C = C_1 \wedge C_2$. We have (5.1) for $n \ge n_0$. Decreasing C, if necessary, we get (5.1) for all n. □

Lemma 5.2. *For every $T > 0$ and $\epsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left[\max_{1\le j\le A^n(nT)} L_j^n \ge \epsilon n\right] \to 0. \tag{5.3}$$

Proof. Assume that (5.3) is false, i.e., for some subsequence $n_k \to \infty$,

$$\lim_{k\to\infty} \mathbb{P}\left[\max_{1\le j\le A^{n_k}(n_k T)} L_j^{n_k} \ge \epsilon n_k\right] = c > 0. \tag{5.4}$$

By (2.13), we can assume that $\mathbb{E}\Big[\frac{(L_j^{n_k})^+}{\sqrt{n_k}}\Big]^2 \to c_1 \in [0, \tilde{C}]$ as $k \to \infty$. Thus, by (2.15) and Theorem 3.1 in Prokhorov [13],

$$B^{n_k}(t) \triangleq \frac{1}{\sqrt{n_k}} \sum_{j=1}^{\lfloor \lambda_{n_k} n_k t\rfloor} \left(\frac{(L_j^{n_k})^+}{\sqrt{n_k}} - \mathbb{E}\left[\frac{(L_j^{n_k})^+}{\sqrt{n_k}}\right]\right) \Rightarrow B(t) \tag{5.5}$$

in $D[0,\infty)$, where B is a Brownian motion. Let $A_n = [A^n(nT) \le \lambda_n n(T+1)]$. By (2.23), $P(A_n) \to 1$. On A_{n_k}, we have

$$\begin{aligned}\frac{1}{n_k} \max_{1\le j\le A^{n_k}(n_k T)} L_j^{n_k} \le{}& \frac{1}{\sqrt{n_k}} \max_{1\le j\le \lambda_n n(T+1)} \left\{\frac{(L_j^{n_k})^+}{\sqrt{n_k}} - \mathbb{E}\left[\frac{(L_j^{n_k})^+}{\sqrt{n_k}}\right]\right\}\\ &+ \frac{1}{\sqrt{n_k}}\mathbb{E}\left[\frac{(L_j^{n_k})^+}{\sqrt{n_k}}\right].\end{aligned} \tag{5.6}$$

The first term on the RHS of (5.6) is bounded above by $\sup_{0\le t\le T+1}[B^{n_k}(t) - B^{n_k}(t-)]$, which converges weakly to zero by (5.5) and the continuous mapping theorem. The second term converges to zero by (2.13) and the Schwarz inequality. Thus, the LHS of (5.6) is bounded above by a process converging weakly to zero, which contradicts (5.4). □

Define, for $t \ge 0$,

$$\tau^n(t) \triangleq \sup\big\{s \in [0,t] : \widehat{C}^n(s) = \widehat{F}^n(s)\big\}. \tag{5.7}$$

Observe that $\widehat{C}^n(0) = \widehat{F}^n(0) = \infty$, so the supremum in (5.7) is taken over a nonempty set.

Lemma 5.3. *$\tau^n \Rightarrow e$ in $D[0,\infty)$.*

Proof. We fix $T > 0$ and prove convergence on $[0,T]$. By the definition (5.7) and the fact that the inter-arrival times are strictly positive, we have

$$\begin{aligned}\widehat{\mathcal{W}}^n(\tau^n(t))\big[\widehat{C}^n(\tau^n(t)), \widehat{F}^n(\tau^n(t))\big) &= \frac{1}{\sqrt{n}} v^n_{A^n(n\tau^n(t))}\\ &\le \max_{0\le s\le T}\big[\widehat{N}^n(s) - \widehat{N}^n(s-)\big].\end{aligned} \tag{5.8}$$

As long as there are customers with lead times in $[C^n, F^n)$, F^n decreases at rate 1 per unit time, so for $s \in (n\tau^n(t), nt]$, $F^n(s) = F^n(n\tau^n(t)) - (s - n\tau^n(t))$. Let

$$D^n(t) = \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n\tau^n(t) < S_j^n \le nt\}} \mathbb{I}_{\{L_j^n - (nt - S_j^n) < F^n(n\tau^n(t)) - n(t - \tau^n(t))\}}$$

be the work associated with customers arriving within the time interval $(n\tau^n(t), nt]$ whose lead times upon arrival are to the left of the frontier. On the time interval $(n\tau^n(t), nt]$ the server is never idle, so

$$\begin{aligned} 0 \le \widehat{\mathcal{W}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big) = {} & \widehat{\mathcal{W}}^n(\tau^n(t))\big[\widehat{C}^n(\tau^n(t)), \widehat{F}^n(\tau^n(t))\big) \\ & + \frac{1}{\sqrt{n}} D^n(t) - \sqrt{n}(t - \tau^n(t)). \end{aligned} \tag{5.9}$$

Let $\epsilon > 0$. For a given $t \in [0, T]$, either

$$t - \tau^n(t) < \epsilon, \tag{5.10}$$

or

$$\tau^n(t) + \epsilon \le t. \tag{5.11}$$

Assume (5.11). Let y be a point of continuity of G_v such that for each n, $G_v^n(y) < 1/(2\mu_n)$ (such a choice is possible by (2.7)) and let $A_n = \big[\sup_{0 \le t \le T} F^n(t) \le n\epsilon + \sqrt{n}y\big]$. By Lemma 5.2, $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. On A_n, we have

$$\begin{aligned} \frac{1}{\sqrt{n}} D^n(t) \le {} & \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n\tau^n(t) < S_j^n \le n(\tau^n(t) + \epsilon)\}} \\ & + \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n(\tau^n(t) + \epsilon) < S_j^n \le nt\}} \mathbb{I}_{\{L_j^n < F^n(\tau^n(t)) - n\epsilon\}} \\ \le {} & \frac{1}{\sqrt{n}} V^n(A^n(n(\tau^n(t) + \epsilon))) - \frac{1}{\sqrt{n}} V^n(A^n(n\tau^n(t))) \\ & + \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n(\tau^n(t) + \epsilon) < S_j^n \le nt\}} \mathbb{I}_{\{L_j^n < \sqrt{n}y\}} \\ = {} & \widehat{N}^n(\tau^n(t) + \epsilon) - \widehat{N}^n(\tau^n(t)) + \sqrt{n}\epsilon + Y^n\left(\frac{1}{n} A^n(nt), y\right) \\ & - Y^n\left(\frac{1}{n} A^n(n(\tau^n(t) + \epsilon)), y\right) + \frac{G_v^n(y)}{\sqrt{n}} \left(A^n(nt) - A^n(n(\tau^n(t) + \epsilon))\right) \\ = {} & \widehat{N}^n(\tau^n(t) + \epsilon) - \widehat{N}^n(\tau^n(t)) + \sqrt{n}\epsilon + Y^n\left(\lambda_n t + o(1), y\right) \\ & - Y^n\left(\lambda_n(\tau^n(t) + \epsilon) + o(1), y\right) + G_v^n(y)\left(\widehat{A}^n(t) - \widehat{A}^n(\tau^n(t) + \epsilon)\right) \\ & + \lambda_n \sqrt{n} G_v^n(y)(t - \tau^n(t) - \epsilon). \end{aligned}$$

Thus, by (2.23), (2.24), Lemma 4.2 and the choice of y, on A_n we have

$$\frac{1}{\sqrt{n}} D^n(t) \le O(1) + \sqrt{n}\epsilon + \frac{\rho^n}{2}\sqrt{n}(t - \tau^n(t) - \epsilon).$$

Plugging this into (5.9) and using (5.8), we obtain, under the assumption (5.11),

$$0 \le O(1) + \Big(\frac{\rho_n}{2} - 1\Big)\sqrt{n}(t - \tau^n(t) - \epsilon) \tag{5.12}$$

on A_n. If (5.10) holds, then $(t - \tau^n(t) - \epsilon)^+ = 0$. Thus, by (2.14) and (5.12), $(t - \tau^n(t) - \epsilon)^+ \Rightarrow 0$ in $D[0,T]$. Since $\epsilon > 0$ is arbitrary, $\tau^n \Rightarrow e$ in $D[0,T]$. □

Corollary 5.4. *$\widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big] \Rightarrow 0$ in $D[0,\infty)$.*

Proof. We have

$$\begin{aligned}\widehat{W}^n &= \widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big) + \widehat{\mathcal{W}}^n\big\{\widehat{F}^n\big\} + \widehat{\mathcal{W}}^n\big(\widehat{F}^n, \infty\big)\\ &= \widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big) + \widehat{\mathcal{V}}^n\big(\widehat{F}^n, \infty\big) + o(1),\end{aligned}$$

because, by Corollary 4.8 and Lemma 5.1,

$$0 \le \widehat{\mathcal{W}}^n\big\{\widehat{F}^n\big\} \le \widehat{\mathcal{V}}^n\big\{\widehat{F}^n\big\} = o(1). \tag{5.13}$$

Thus,

$$\begin{aligned}\widehat{W}^n(t) - \widehat{W}^n(\tau^n(t)) &= \widehat{\mathcal{W}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big)\\ &\quad - \widehat{\mathcal{W}}^n(\tau^n(t))\big[\widehat{C}^n(\tau^n(t)), \widehat{F}^n(\tau^n(t))\big)\\ &\quad + \widehat{\mathcal{V}}(t)\big(\widehat{F}^n(t), \infty\big)\\ &\quad - \widehat{\mathcal{V}}(\tau^n(t))\big(\widehat{F}^n(\tau^n(t)), \infty\big) + o(1).\end{aligned} \tag{5.14}$$

By (2.24) and Lemma 5.3, the LHS of (5.14) is $o(1)$. The second term on the RHS of (5.14) is $o(1)$ by (2.24) and (5.8). Finally, on the time interval $[\tau^n(t), t]$, $\widehat{\mathcal{V}}(\cdot)\big(\widehat{F}^n(\cdot), \infty\big)$ is nondecreasing, so (5.14) implies that $\widehat{\mathcal{W}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big) \Rightarrow 0$ in $D[0,\infty)$. This, together with (5.13), shows that $\widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big] \Rightarrow 0$ in $D[0,\infty)$. □

Lemma 5.5. *Let $\epsilon > 0$, $\eta > 0$ be arbitrary. There exists $C > 0$ such that for all n,*

$$\mathbb{P}\left[\sup_{0\le t\le T}\Big\{\widehat{F}^n(t)\mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}}\Big\} \le C\right] \ge 1 - \eta. \tag{5.15}$$

Proof. By Corollary 5.4, there exists n_1 such that for $n \ge n_1$, $\mathbb{P}(A_n) \ge 1 - \eta/3$, where $A_n = \Big[\sup_{0\le t\le T} \widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big] \le \epsilon/2\Big]$. We have

$$\widehat{W}^n = \widehat{\mathcal{W}}^n\big[\widehat{C}^n, \widehat{F}^n\big] + \widehat{\mathcal{V}}^n\big(\widehat{F}^n, \infty\big), \tag{5.16}$$

so on A_n, for any $t \in [0,T]$, $\widehat{\mathcal{V}}^n\big(\widehat{F}^n(t),\infty\big) \geq \widehat{W}^n(t) - \epsilon/2$. Using Lemma 5.1, choose C_1 such that $\mathbb{P}(B_n) \geq 1-\eta/3$ for all n, where $B_n = \Big[\inf_{0\leq t\leq T} \widehat{F}^n(t) \geq C_1\Big]$. By Proposition 4.7, there exists n_2 such that for $n \geq n_2$, $\mathbb{P}(C_n) \geq 1-\eta/3$, where

$$C_n = \left[\sup_{y\geq C_1}\sup_{0\leq t\leq T}\left|\widehat{\mathcal{V}}^n(t)(y,\infty) - H_v^n(y) + H_v^n\big(y+\sqrt{n}t\big)\right| \leq \epsilon/4\right].$$

On $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) \geq \epsilon\big]$, we have

$$\epsilon/2 \leq \widehat{\mathcal{V}}^n\big(\widehat{F}^n(t),\infty\big) \leq H_v^n\big(\widehat{F}^n(t)\big) - H_v^n\big(\widehat{F}^n(t)+\sqrt{n}t\big) + \epsilon/4.$$

In particular, on this set $\epsilon/4 \leq H_v^n\big(\widehat{F}^n(t)\big)$, so $\widehat{F}^n(t) \leq (H_v^n)^{-1}(\epsilon/4) \leq C$, where $C \triangleq \sup_{n\in\mathbb{N}}(H_v^n)^{-1}(\epsilon/4) < \infty$, because, by (2.17),

$$\lim_{y\to\infty}\sup_{n\in\mathbb{N}} H_v^n(y) = 0.$$

Thus, on $A_n \cap B_n \cap C_n$,

$$\sup_{0\leq t\leq T}\left\{\widehat{F}^n(t)\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}}\right\} \leq C$$

and $\mathbb{P}(A_n \cap B_n \cap C_n) \geq 1-\eta$ for $n \geq n_1 \vee n_2$. Increasing C, if necessary, we get (5.15) for all n. □

Lemma 5.6. *For every $\epsilon > 0$,*

$$\sqrt{n}(t-\tau^n(t))\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}} \Rightarrow 0 \tag{5.17}$$

in $D[0,\infty)$.

Proof. We fix $T > 0$ and prove convergence on $[0,T]$. Choose $\eta > 0$. By (2.24), there exist $\delta > 0$ and $n_1 \in \mathbb{N}$ such that for $n \geq n_1$, $\mathbb{P}(A_n) \geq 1-\eta/3$, where $A_n = \big[\omega_{\widehat{W}^n}(\delta) \leq \epsilon/2\big]$ and for a function $f \in D[0,\infty)$,

$$\omega_f(\delta) = \sup_{\substack{0\leq s_1<s_2\leq T\\ s_2-s_1\leq\delta}} |f(s_2)-f(s_1)|.$$

By Lemma 5.3, there exists n_2 such that for $n \geq n_2$, $\mathbb{P}(B_n) \geq 1-\eta/3$, where $B_n = \big[\sup_{0\leq t\leq T}(t-\tau^n(t)) \leq \delta\big]$. For $t \in [0,T]$, on $A_n \cap B_n \cap \big[\widehat{W}^n(t) \geq \epsilon\big]$ we have $\widehat{W}^n(\tau^n(t)) \geq \epsilon/2$. By Lemma 5.5, we can choose $C > 0$ such that for all n, $\mathbb{P}(C_n) \geq 1-\eta/3$, where $C_n = \Big[\sup_{0\leq s\leq T}\Big\{\widehat{F}^n(s)\mathbb{I}_{\{\widehat{W}^n(s)\geq\epsilon/2\}}\Big\} \leq C\Big]$. Thus, for $t \in [0,T]$, on $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) \geq \epsilon\big]$,

$$\widehat{F}^n(\tau^n(t)) \leq C. \tag{5.18}$$

Let $c > 0$ be such that $C - c$ is a point of continuity of G_v. Either

$$t - \tau^n(t) < c/\sqrt{n}, \tag{5.19}$$

or

$$n\tau^n(t) + c\sqrt{n} \le nt. \tag{5.20}$$

Assume (5.20). Then, on $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) \ge \epsilon\big]$,

$$\begin{aligned}
\frac{1}{\sqrt{n}} D^n(t) &\le \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n\tau^n(t) < S_j^n \le n\tau^n(t) + c\sqrt{n}\}} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n\tau^n(t)+c\sqrt{n} < S_j^n \le nt\}} \mathbb{I}_{\{L_j^n < F^n(\tau^n(t)) - c\sqrt{n}\}} \\
&\le \frac{1}{\sqrt{n}} V^n\big(A^n\big(n\tau^n(t) + c\sqrt{n}\big)\big) - \frac{1}{\sqrt{n}} V^n(A^n(n\tau^n(t))) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} v_j^n \mathbb{I}_{\{n\tau^n(t)+c\sqrt{n} < S_j^n \le nt\}} \mathbb{I}_{\{L_j^n < \sqrt{n}(C-c)\}} \\
&= \widehat{N}^n\big(\tau^n(t) + c/\sqrt{n}\big) - \widehat{N}^n(\tau^n(t)) + c \\
&\quad + Y^n\left(\frac{1}{n} A^n(nt), C - c\right) - Y^n\left(\frac{1}{n} A^n\big(n\tau^n(t) + c\sqrt{n}\big), C - c\right) \\
&\quad + \frac{G_v^n(C-c)}{\sqrt{n}} \left(A^n(nt) - A^n\big(n\tau^n(t) + c\sqrt{n}\big)\right) \\
&= \widehat{N}^n\big(\tau^n(t) + c/\sqrt{n}\big) - \widehat{N}^n(\tau^n(t)) + c + Y^n\left(\lambda_n t + o(1), C - c\right) \\
&\quad - Y^n\left(\lambda_n \tau^n(t) + o(1), C - c\right) \\
&\quad + G_v^n(C-c)\left(\widehat{A}^n(t) - \widehat{A}^n(\tau^n(t) + o(1)) + \lambda_n\sqrt{n}(t - \tau^n(t)) - \lambda_n c\right),
\end{aligned}$$

where the second inequality follows from (5.18). By Theorem 3.1 in Prokhorov [13], $Y^n(\cdot, C-c)$ converge weakly to a Brownian motion. Thus, by (2.24) and Lemma 5.3, on $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) \ge \epsilon\big]$, under the assumption (5.20),

$$\frac{1}{\sqrt{n}} D^n(t) \le o(1) + \lambda_n G_v^n(C-c)\sqrt{n}(t - \tau^n(t)) + (1 - \lambda_n G_v^n(C-c))c. \tag{5.21}$$

Plugging (5.21) into (5.9) and using (5.8), we get

$$0 \le o(1) - (1 - \lambda_n G_v^n(C-c))\sqrt{n}(t - \tau^n(t)) + (1 - \lambda_n G_v^n(C-c))c. \tag{5.22}$$

Thus, on $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) \ge \epsilon\big]$, under the assumption (5.20),

$$0 \le \sqrt{n}(t - \tau^n(t)) \le c + \frac{o(1)}{1 - \lambda_n G_v^n(C-c)}. \tag{5.23}$$

Of course, if (5.19) holds, then (5.23) holds as well. However, $G_v^n(C-c) \to G_v(C-c) < 1/\lambda$ and $c > 0$ may be arbitrarily small, so (5.23) and the fact that $\mathbb{P}(A_n \cap B_n \cap C_n) \ge 1 - \eta$ imply (5.17). □

Corollary 5.7. *For every $\epsilon > 0$,*

$$\widehat{\mathcal{Q}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big]\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}} \Rightarrow 0 \tag{5.24}$$

in $D[0,\infty)$.

Proof.

$$\begin{aligned}
&\widehat{\mathcal{Q}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big)\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}} \\
&\leq \frac{1}{\sqrt{n}}\left[1 + A^n(nt) - A^n(n\tau^n(t))\right]\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}} \qquad (5.25)\\
&\leq \frac{1}{\sqrt{n}} + \big[\widehat{A}^n(t) - \widehat{A}^n(\tau^n(t)) + \lambda_n\sqrt{n}(t-\tau^n(t))\big]\mathbb{I}_{\{\widehat{W}^n(t)\geq\epsilon\}} \Rightarrow 0
\end{aligned}$$

by (2.23) and Lemma 5.6. Using the inequality $0 \leq \widehat{\mathcal{Q}}^n\{\widehat{F}^n\} \leq \widehat{\mathcal{A}}^n\{\widehat{F}^n\}$, Corollary 4.8 and Lemma 5.1, we upgrade (5.25) to (5.24). □

6. Proofs of the main results. In this section, we prove Proposition 3.1 and Theorem 3.2. Proposition 3.1, the fact that $\widehat{\mathcal{W}}^n \Rightarrow \mathcal{W}^*$ and the limiting behavior of $\widehat{\mathcal{Q}}^n$ as long as $\widehat{W}^n$ is bounded away from zero follow quickly from the results of Sections 4 and 5. Therefore, to show that $\widehat{\mathcal{Q}}^n \Rightarrow \mathcal{Q}^*$, it suffices to show that for large n, $\widehat{Q}^n$ is small when $\widehat{W}^n$ is small (the latter is not obvious, because, in principle, there may be many partially served customers present in the system). This is implied by Lemma 6.1, to follow.

Proof of Proposition 3.1. Fix $T > 0$. We will show that

$$\sup_{0\leq t\leq T}\left|\widehat{W}^n(t) - H_v\big(\widehat{F}_1^n(t)\big)\right| \xrightarrow{P} 0. \tag{6.1}$$

By (2.24) and the definition of $\widehat{F}_1^n$,

$$\sup_{0\leq t\leq n^{-\frac{1}{4}}}\left|\widehat{W}^n(t) - H_v\big(\widehat{F}_1^n(t)\big)\right| = \sup_{0\leq t\leq n^{-\frac{1}{4}}}\left|\widehat{W}^n(t)\right| \xrightarrow{P} 0,$$

so it suffices to prove

$$\sup_{n^{-\frac{1}{4}}\leq t\leq T}\left|\widehat{W}^n(t) - H_v\big(\widehat{F}_1^n(t)\big)\right| \xrightarrow{P} 0. \tag{6.2}$$

Let $\epsilon > 0$. Using Lemma 5.1, we choose $C > 0$ such that $\mathbb{P}(A_n) \geq 1 - \epsilon/3$, where $A_n = \Big[\inf_{0\leq t\leq T}\widehat{F}^n(t) \geq C\Big]$. By Proposition 4.7 and Corollary 5.4, there exists n_0 such that for $n \geq n_0$, $\mathbb{P}(B_n) \geq 1 - \epsilon/3$ and $\mathbb{P}(C_n) \geq 1 - \epsilon/3$, where

$$B_n = \left[\sup_{y\geq C}\sup_{0\leq t\leq T}\left|\widehat{\mathcal{V}}^n(t)(y,\infty) - H_v^n(y) + H_v^n\big(y+\sqrt{n}t\big)\right| \leq \epsilon/4\right],$$

$$C_n = \left[\sup_{0\le t\le T} \widehat{\mathcal{W}}^n(t)\big[\widehat{C}^n(t), \widehat{F}^n(t)\big] \le \epsilon/4 \right].$$

On $A_n \cap B_n \cap C_n$,

$$\sup_{0\le t\le T} \left| \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t), \infty\big) - H_v^n\big(\widehat{F}^n(t)\big) + H_v^n\big(\widehat{F}^n(t) + \sqrt{n}t\big) \right| \le \epsilon/4,$$

so, by (5.16),

$$\sup_{0\le t\le T} \left| \widehat{W}^n(t) - H_v^n\big(\widehat{F}^n(t)\big) + H_v^n\big(\widehat{F}^n(t) + \sqrt{n}t\big) \right| \le \epsilon/2.$$

On $A_n \cap B_n \cap C_n$,

$$\sup_{n^{-\frac14}\le t\le T} H_v^n\big(\widehat{F}^n(t) + \sqrt{n}t\big) \le H_v^n\big(C + n^{\frac14}\big) \to 0$$

by (2.17), so there exists $n_1 \ge n_0$ such that for $n \ge n_1$,

$$\sup_{n^{-\frac14}\le t\le T} \left| \widehat{W}^n(t) - H_v^n\big(\widehat{F}_1^n(t)\big) \right| \le 3\epsilon/4$$

(recall that, by definition, $\widehat{F}_1^n(t) = \widehat{F}^n(t)$ for $t \ge n^{-\frac14}$). The functions H_v^n converge to H_v uniformly on $[C, \infty)$, so there exists $n_2 \ge n_1$ such that for $n \ge n_2$,

$$\sup_{n^{-\frac14}\le t\le T} \left| \widehat{W}^n(t) - H_v\big(\widehat{F}_1^n(t)\big) \right| \le \epsilon$$

on $A_n \cap B_n \cap C_n$ and $\mathbb{P}(A_n \cap B_n \cap C_n) \ge 1 - \epsilon$. Thus, (6.2) holds, so (6.1) holds also. In particular, by (2.24), $H_v\big(\widehat{F}_1^n\big) \Rightarrow W^*$ in $D[0,T]$. Applying a continuous map H_v^{-1}, we get $\widehat{F}_1^n \Rightarrow F^*$ in $D_{\overline{\mathbb{R}}}[0,T]$. $\square$

Proof of Theorem 3.2. Fix $T > 0$. From Proposition 4.7, for any $y_0 \in \mathbb{R}$,

$$\sup_{y\ge y_0} \sup_{0\le t\le T} \Big| \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t) \vee y, \infty\big) - H_v^n\big(\widehat{F}^n(t) \vee y\big) + H_v^n\Big(\big(\widehat{F}^n(t) \vee y\big) + \sqrt{n}t\Big) \Big| \xrightarrow{P} 0.$$

From this, using the inequality $\widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t)\vee y, \infty\big) \le \widehat{W}^n(t)$ and proceeding as in the proof of (6.1), we get

$$\sup_{y\ge y_0} \sup_{0\le t\le T} \Big| \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t) \vee y, \infty\big) - H_v\big(\widehat{F}_1^n(t) \vee y\big) \Big| \xrightarrow{P} 0. \tag{6.3}$$

By Lemma 5.1, (6.3) may be extended to

$$\sup_{y\in\mathbb{R}} \sup_{0\le t\le T} \Big| \widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t) \vee y, \infty\big) - H_v\big(\widehat{F}_1^n(t) \vee y\big) \Big| \xrightarrow{P} 0. \tag{6.4}$$

Similarly, we can show

$$\sup_{y\in\mathbb{R}} \sup_{n^{-\frac{1}{4}}\le t\le T} \left|\widehat{\mathcal{A}}^n(t)\big(\widehat{F}_1^n(t)\vee y,\infty\big) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \xrightarrow{P} 0. \tag{6.5}$$

(6.4), together with Corollary 5.4, implies that

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \sup_{0\le t\le T} \left|\widehat{\mathcal{W}}^n(t)(y,\infty) - H_v\big(\widehat{F}_1^n(t)\vee y\big)\right| \\
&\quad\le \sup_{0\le t\le T} \widehat{\mathcal{W}}^n(t)\big[\widehat{C}^n(t),\widehat{F}^n(t)\big] \\
&\qquad + \sup_{y\in\mathbb{R}} \sup_{0\le t\le T} \left|\widehat{\mathcal{V}}^n(t)\big(\widehat{F}^n(t)\vee y,\infty\big) - H_v\big(\widehat{F}_1^n(t)\vee y\big)\right| \xrightarrow{P} 0.
\end{aligned} \tag{6.6}$$

Define $\psi,\psi_1 : \overline{\mathbb{R}} \to \mathcal{M}$ by

$$\psi(x)(B) = \int_{B\cap[x,\infty)} (1-\lambda G_v(\eta))d\eta, \quad \psi_1(x)(B) = \lambda\int_{B\cap[x,\infty)} (1-G(\eta))d\eta$$

for any $B\in\mathcal{B}(\mathbb{R})$. It is easy to check that ψ,ψ_1 are continuous. In particular, by Proposition 3.1,

$$\psi\big(\widehat{F}_1^n\big) \Rightarrow \psi(F^*) = \mathcal{W}^*, \qquad \psi_1\big(\widehat{F}_1^n\big) \Rightarrow \psi_1(F^*) = \mathcal{Q}^* \tag{6.7}$$

in $D_{\mathcal{M}}[0,\infty)$. Moreover, for any $y\in\mathbb{R}$, $\psi\big(\widehat{F}_1^n(t)\big)(y,\infty) = H_v\big(\widehat{F}_1^n(t)\vee y\big)$, so (6.6) and (6.7) imply that $\widehat{\mathcal{W}}^n \Rightarrow \mathcal{W}^*$ in $D_{\mathcal{M}}[0,\infty)$.

We claim that for every $\epsilon>0$,

$$\sup_{y\in\mathbb{R}} \sup_{0\le t\le T} \left|\widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}} \xrightarrow{P} 0. \tag{6.8}$$

Indeed, for $n^{-\frac{1}{4}}\le t\le T$,

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \left|\widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}} \\
&\quad\le \widehat{\mathcal{Q}}^n(t)\big[\widehat{C}^n(t),\widehat{F}^n(t)\big]\mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}} \\
&\qquad + \sup_{y\in\mathbb{R}} \left|\widehat{\mathcal{Q}}^n(t)\big(\widehat{F}^n(t)\vee y,\infty\big) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \\
&\quad= \widehat{\mathcal{Q}}^n(t)\big[\widehat{C}^n(t),\widehat{F}^n(t)\big]\mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}} \\
&\qquad + \sup_{y\in\mathbb{R}} \left|\widehat{\mathcal{A}}^n(t)\big(\widehat{F}_1^n(t)\vee y,\infty\big) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right|
\end{aligned}$$

because the customers with lead times greater than $F^n(nt)$ have not received service by time nt and $\widehat{F}^n(t) = \widehat{F}_1^n(t)$ for all t under consideration. This, together with Corollary 5.7 and (6.5), shows

$$\sup_{y\in\mathbb{R}} \sup_{n^{-\frac{1}{4}}\le t\le T} \left|\widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon\}} \xrightarrow{P} 0. \tag{6.9}$$

However, by (2.24) and the equality $W^*(0) = 0$,

$$\mathbb{P}\left[\sup_{0\le t\le n^{-\frac{1}{4}}} \widehat{W}^n(t) \ge \epsilon\right] \to 0,$$

which, together with (6.9), proves (6.8). We need

Lemma 6.1. *For every $\epsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left[\widehat{Q}^n(t) < \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon \textit{ for every } t \in [0,T]\right] \to 1. \tag{6.10}$$

The main idea of the proof of Lemma 6.1 is, roughly speaking, the observation that if

$$\widehat{Q}^n(t) \ge \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon \tag{6.11}$$

for some t, then, by (4.49) and the fact that $\widehat{\mathcal{Q}}^n \le \widehat{\mathcal{A}}^n$, there exists a constant C such that customers with lead times at most $C\sqrt{n}$ have been present at the system for some period of time before nt. Therefore, the newcoming customers with lead times greater than $C\sqrt{n}$ have not received any service during this time period and, consequently, $\widehat{W}^n(t)$ is bounded away from zero. This, however, is unlikely because of (6.8) and (6.11).

Proof of Lemma 6.1. Suppose that (6.10) is false. Then, there exist $\eta > 0$ and a subsequence (still denoted by n) such that along this subsequence, $P(A_n) \ge \eta$, where

$$A_n = \left[\widehat{Q}^n(t) \ge \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon \text{ for some } t \in [0,T]\right].$$

Let

$$\sigma^n = \begin{cases} \inf\left\{t \in [0,T] : \widehat{Q}^n(t) \ge \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon\right\} & \text{on } A_n, \\ +\infty & \text{on } A_n^c. \end{cases}$$

On A_n, $\widehat{Q}^n(\sigma^n) \ge \lambda H\big(\widehat{F}_1^n(\sigma^n)\big) + \epsilon$. By (6.8), we have, for every $\epsilon_1 > 0$, $\mathbb{P}\big[\sigma^n < \infty, \widehat{W}^n(\sigma^n) \ge \epsilon_1\big] \to 0$, so

$$\widehat{W}^n(\sigma^n)\mathbb{I}_{\{\sigma^n<\infty\}} \Rightarrow 0. \tag{6.12}$$

Let

$$\nu^n = \begin{cases} \sup\left\{t \in [0,\sigma^n] : \widehat{Q}^n(t) \le \epsilon/2\right\} & \text{on } A_n, \\ +\infty & \text{on } A_n^c. \end{cases}$$

The inter-arrival times are strictly positive, so on A_n

$$\epsilon/2 < \widehat{Q}^n(\nu^n) \le \epsilon/2 + 1/\sqrt{n}. \tag{6.13}$$

Thus, on A_n, $\widehat{Q}^n(t) > \epsilon/2$ for $t \in [\nu^n, \sigma^n]$. In what follows, all random quantities under consideration are evaluated at some $\omega \in A_n$. Our next step is to show

$$\sup_{\nu^n \le t \le \sigma^n} \widehat{W}^n(t) \Rightarrow 0 \tag{6.14}$$

If $\tau^n(\sigma^n) < n^{-\frac{1}{4}}$, then, by Lemma 5.3, $\sigma^n = o(1)$ and thus, by (2.24) and $W^*(0) = 0$, (6.14) holds. Assume $\tau^n(\sigma^n) \ge n^{-\frac{1}{4}}$. By (2.24), Lemma 5.3 and (6.12), $\widehat{W}^n(\tau^n(\sigma^n)) \Rightarrow 0$, so, by (6.1), $\widehat{F}_1^n(\tau^n(\sigma^n)) \Rightarrow \infty$. Thus, by Proposition 4.7 and the fact that $\lim_{y\to\infty} \sup_{n\in\mathbb{N}} H^n(y) = 0$,

$$\widehat{\mathcal{Q}}^n(\tau^n(\sigma^n))\big(\widehat{F}_1^n(\tau^n(\sigma^n)), \infty\big) \le \widehat{\mathcal{A}}^n(\tau^n(\sigma^n))\big(\widehat{F}_1^n(\tau^n(\sigma^n)), \infty\big) \Rightarrow 0.$$

Also, $\widehat{F}_1^n(\tau^n(\sigma^n)) = \widehat{F}^n(\tau^n(\sigma^n))$, so, by Corollary (4.8) and the definition of τ^n,

$$\begin{aligned}
\widehat{\mathcal{Q}}^n(\tau^n(\sigma^n))&\big[\widehat{C}^n(\tau^n(\sigma^n)), \widehat{F}_1^n(\tau^n(\sigma^n))\big] \\
&\le \widehat{\mathcal{Q}}^n(\tau^n(\sigma^n))\big[\widehat{C}^n(\tau^n(\sigma^n)), \widehat{F}^n(\tau^n(\sigma^n))\big) \\
&\quad + \widehat{\mathcal{A}}^n(\tau^n(\sigma^n))\big\{\widehat{F}_1^n(\tau^n(\sigma^n))\big\} = \frac{1}{\sqrt{n}} + o(1) \Rightarrow 0.
\end{aligned}$$

Thus, $\widehat{Q}^n(\tau^n(\sigma^n)) \Rightarrow 0$, so $\mathbb{P}[\tau^n(\sigma^n) \ge \nu^n] \to 0$. This implies that, with probability arbitrarily close to 1,

$$\sup_{\nu^n \le t \le \sigma^n} \widehat{W}^n(t) \le \sup_{\tau^n(\sigma^n) \le t \le \sigma^n} \widehat{W}^n(t) \Rightarrow 0$$

by (2.24), Lemma 5.3 and (6.12). In any case, (6.14) holds.

Let $C > 0$ be such that $\sup_{n\in\mathbb{N}} H^n(C) \le \epsilon/6$. We claim that there exists n_0 such that for every $n \ge n_0$, $\mathbb{P}(B_n) \ge 3\eta/4$, where $B_n \subseteq A_n$ is the set on which at any time $t' \in [n\nu^n, n\sigma^n]$ customers with lead times at most $C\sqrt{n}$ are present in the system. Indeed, suppose that $\mathbb{P}(A_n \setminus B_n) > \eta/4$ for infinitely many n. On the set $A_n \setminus B_n$ there exists $t' \in [n\nu^n, n\sigma^n]$ such that all customers in the system at time t' have lead times greater than $C\sqrt{n}$, so, for $t = t'/n$, by Proposition 4.6,

$$\begin{aligned}
\frac{\epsilon}{2} &\le \widehat{Q}^n(t) = \widehat{\mathcal{Q}}^n(t)(C, \infty) \le \widehat{\mathcal{A}}^n(t)(C, \infty) \\
&\le H^n(C) - H^n\big(C + \sqrt{n}t\big) + \frac{\epsilon}{6} \le H^n(C) + \frac{\epsilon}{6} \le \frac{\epsilon}{3}
\end{aligned}$$

with probability at least $\eta/6$ for n sufficiently large, contradiction. By (2.23), we can choose $\delta > 0$ and $n_1 \in \mathbb{N}$ such that for $n \ge n_1$, $\mathbb{P}(C_n) \ge 1 - \eta/4$, where $C_n = [\omega_{\widehat{A}^n}(\delta) \le \epsilon/6]$ and $\omega_{\widehat{A}^n}$ is the modulus of continuity of $\widehat{A}^n$ on $[0, T]$. Let $c = \epsilon/(6\lambda)$. On $B_n \cap C_n$ we have, for n sufficiently large,

$$\sqrt{n}(\sigma^n - \nu^n) \ge c. \tag{6.15}$$

Indeed, if (6.15) is false, then, by (2.1), (6.13), for arbitrarily large n, on $B_n \cap C_n$,

$$\begin{aligned}\epsilon &\le \widehat{Q}^n(\sigma^n) \le \widehat{Q}^n(\nu^n) + \widehat{A}^n(\sigma^n) - \widehat{A}^n(\nu^n) + \lambda_n\sqrt{n}(\sigma^n - \nu^n)\\ &\le \epsilon/2 + 1/\sqrt{n} + \lambda_n\epsilon/(6\lambda) + \epsilon/6 < \epsilon,\end{aligned}$$

contradiction. The customers who have entered the system in the time interval $\left[n\nu^n, n\nu^n + c\sqrt{n}\right]$ with initial lead times greater than $\sqrt{n}(C+c)$ have lead times greater than $\sqrt{n}C$ at all times in $\left[n\nu^n, n\nu^n + c\sqrt{n}\right]$, so, by (6.15) and the definition of B_n, on the event $B_n \cap C_n$, none of them has received any service by time $n\nu^n + c\sqrt{n}$. Thus, on this event,

$$\begin{aligned}&\widehat{W}^n\big(\nu^n + c/\sqrt{n}\big)\\ &\quad\ge \frac{1}{\sqrt{n}}\Big(V^n\Big(A^n\big(n\nu^n + c\sqrt{n}\big)\Big) - V^n(A^n(n\nu^n))\Big)\\ &\qquad - \left(Y^n\left(\frac{1}{n}A^n\big(n\nu^n + c\sqrt{n}\big), C+c\right) - Y^n\left(\frac{1}{n}A^n(n\nu^n), C+c\right)\right)\\ &\qquad - \frac{1}{\sqrt{n}}\left(A^n\big(n\nu^n + c\sqrt{n}\big) - A^n(n\nu^n)\right)G_v^n(C+c)\\ &\quad= \widehat{N}^n\big(\nu^n + c/\sqrt{n}\big) - \widehat{N}^n(\nu^n) + c\\ &\qquad - Y^n\left(\lambda_n\big(\nu^n + c/\sqrt{n}\big) + o(1), C+c\right) + Y^n\left(\lambda_n\nu^n + o(1), C+c\right)\\ &\qquad - \Big(\widehat{A}^n\big(\nu^n + c/\sqrt{n}\big) - \widehat{A}^n(\nu^n) + \lambda_n c\Big)G_v^n(C+c)\\ &\quad= c(1 - \lambda_n G_v^n(C+c)) + o(1),\end{aligned}$$

because (2.23)–(2.24) hold and, as we have already explained, $Y^n(\cdot, C+c)$ converges weakly to a Brownian motion. This, however, contradicts (6.14)–(6.15), because $\mathbb{P}(B_n \cap C_n) \ge \eta/2$ for $n \ge n_0 \vee n_1$ and $\limsup_{n\to\infty} \lambda_n G_v^n(C+c) \le \lambda G_v(d) < 1$ for any $d \ge C+c$ being a point of continuity of G_v. □

Returning to the proof of Theorem 3.2, we want to upgrade (6.8) to

$$\sup_{y\in\mathbb{R}}\sup_{0\le t\le T}\left|\widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right| \xrightarrow{P} 0. \tag{6.16}$$

Let $\epsilon > 0$ be arbitrary and let $\epsilon_1 > 0$ be such that $\lambda H(H_v^{-1}(2\epsilon_1)) + \epsilon_1 \le \epsilon/2$. Let

$$\begin{aligned}A_n &= \left[\widehat{Q}^n(t) < \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon_1 \text{ for every } t \in [0,T]\right],\\ B_n &= \left[\sup_{0\le t\le T}\left|\widehat{W}^n(t) - H_v\big(\widehat{F}_1^n(t)\big)\right| \le \epsilon_1\right],\\ C_n &= \left[\sup_{y\in\mathbb{R}}\sup_{0\le t\le T}\left|\widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t)\vee y\big)\right|\mathbb{I}_{\{\widehat{W}^n(t)\ge\epsilon_1\}} \le \epsilon_1\right].\end{aligned}$$

By (6.1), (6.8) and Lemma 6.1, $\mathbb{P}(A_n \cap B_n \cap C_n) \to 1$. On $A_n \cap B_n$, if $\widehat{W}^n(t) < \epsilon_1$, then $H_v\big(\widehat{F}_1^n(t)\big) < 2\epsilon_1$, so $\lambda H\big(\widehat{F}_1^n(t)\big) < \lambda H(H_v^{-1}(2\epsilon_1))$. Thus, on $A_n \cap B_n \cap C_n \cap \big[\widehat{W}^n(t) < \epsilon_1\big]$,

$$0 \le \widehat{Q}^n(t) < \lambda H\big(\widehat{F}_1^n(t)\big) + \epsilon_1 < \lambda H(H_v^{-1}(2\epsilon_1)) + \epsilon_1 \le \epsilon/2,$$

so on $A_n \cap B_n \cap C_n$,

$$\sup_{y\in\mathbb{R}} \sup_{0\le t\le T} \Big| \widehat{\mathcal{Q}}^n(t)(y,\infty) - \lambda H\big(\widehat{F}_1^n(t) \vee y\big) \Big| \le \epsilon.$$

Thus, (6.16) holds. This, together with (6.7) and the equality

$$\psi_1\big(\widehat{F}_1^n(t)\big)(y,\infty) = \lambda H\big(\widehat{F}_1^n(t) \vee y\big)$$

for all $y \in \mathbb{R}$, shows that $\widehat{\mathcal{Q}}^n \Rightarrow \mathcal{Q}^*$ in $D_{\mathcal{M}}[0,\infty)$. □

7. Examples. In this section, we provide two examples illustrating our theory. In the first one, customer service times and initial lead times are independent. Thus, we get a counterpart of the results of Doytchinov et al. [6] for the case of unbounded lead times. In the second example, customer initial lead times are equal to their (suitably rescaled) service times. This case may be thought of as a regularization of the SRPT service discipline, in which small jobs get preferential treatment, but the priorities of large jobs increase as they wait in queue. For more information on the latter issue, see Bender et al. [1], Crovella et al. [5].

7.1. Independence of initial lead times and service times. Suppose that, in addition to the independence assumptions made in Section 2.2, v_j^n and L_j^n are independent for all $j, n \in \mathbb{N}$ and $G^n \equiv G$ is such that $G(y) < 1$ for all $y \in \mathbb{R}$. Assume that (2.1) and (2.18) hold. Then $G_v^n = \frac{1}{\mu_n} G \Rightarrow G_v = \frac{1}{\lambda} G$ and $G_{v^2}^n = \Big(\beta_n^2 + \frac{1}{\mu_n^2}\Big) G \Rightarrow G_{v^2} = \big(\beta^2 + \frac{1}{\lambda^2}\big) G$. Assume also (2.12), (2.14), the Lindeberg condition on u_j^n, v_j^n, and $\mathbb{E}\big((L_j^n)^+/\sqrt{n}\big)^2 = \int_0^\infty \eta^2 dG(\eta) < \infty$. Then (2.13) holds and $(L_j^n)^+/\sqrt{n}$ satisfy the Lindeberg condition. Moreover, (2.18) implies (2.17) and, as we have observed in Section 2.3, $\big(\overline{\mathbb{R}}, \rho\big)$ is totally bounded. Here we have $H_v = H$ so, by (3.1), $F^* = H^{-1}(W^*)$. By (3.2)–(3.3), for all Borel sets $B \subseteq \mathbb{R}$,

$$\mathcal{W}^*(t)(B) = \int_{B\cap[F^*(t),\infty)} \big(1 - G(\eta)\big)\, d\eta$$

and $\mathcal{Q}^* = \lambda \mathcal{W}^*$. By Theorem 3.2, $\widehat{\mathcal{W}}^n \Rightarrow \mathcal{W}^*$ and $\widehat{\mathcal{Q}}^n \Rightarrow \mathcal{Q}^*$ in $D_{\mathcal{M}}[0,\infty)$. This generalizes the result of Doytchinov et al. [6] to the case of unbounded lead times whose positive parts have finite second moments.

7.2. Initial lead times equal to multiples of service times. Now assume that $L_j^n = \sqrt{n} v_j^n$ for each $j, n \in \mathbb{N}$ and that (2.1), (2.12) hold. For simplicity, we also assume that the distribution $G^n \equiv G$ of v_j^n does not depend on n. In particular, $\mu_n \equiv \lambda$, $\beta_n \equiv \beta$ and for $y \geq 0$,

$$G_v^n(y) = E\left[v_j^n \mathbb{I}_{\{v_j^n \leq y\}}\right] \equiv G_v(y) = \int_0^y \eta dG(\eta),$$

$$G_{v^2}^n(y) = \mathbb{E}\left[\left(v_j^n\right)^2 \mathbb{I}_{\{v_j^n \leq y\}}\right] \equiv G_{v^2}(y) = \int_0^y \eta^2 dG(\eta).$$

Also, $\int_0^\infty (1 - G(\eta))d\eta = \mathbb{E}v_j^n < \infty$ and

$$\begin{aligned}\int_0^\infty (1 - \lambda G_v(\eta))d\eta &= \lambda \int_0^\infty \int_y^\infty \eta dG(\eta) dy \\ &= \frac{\lambda}{2} \int_0^\infty \eta^2 dG(\eta) = \frac{\lambda}{2}\mathbb{E}(v_j^n)^2 < \infty,\end{aligned}$$

so (2.17) holds. We assume (2.14) and the Lindeberg condition on u_j^n. Then, (2.13) and (2.15) hold. Also, as we have observed in Section 2.3, $\left(\overline{\mathbb{R}}, \rho\right)$ is totally bounded. After simple computations, we get

$$H_v(y) = \begin{cases} \frac{\lambda}{2} \int_y^\infty (z^2 - y^2) dG(z), & y \geq 0, \\ \left(\frac{\lambda\beta^2}{2} + \frac{1}{2\lambda}\right) - y, & y < 0. \end{cases}$$

The limiting distributions for $\widehat{\mathcal{W}}^n$ and $\widehat{\mathcal{Q}}^n$ are given by

$$\mathcal{W}^*(t)(B) = \lambda \int_{B \cap [F^*(t), \infty)} \int_{\eta^+}^\infty z dG(z)\, d\eta$$

for all Borel sets $B \subseteq \mathbb{R}$ and (3.3) respectively, where F^* is defined by (3.1).

Appendix. In the Appendix, we observe that the original argument of Doytchinov et al. [6] works, without any substantial modification, as long as functional central limit theorems for the customer arrival and service times hold (see (A.2) and (A.3), to follow), with the limiting processes having continuous sample paths, and the customer lead times are bounded from above. These assumptions hold in a much more general setting than the one that was originally considered in Doytchinov et al. [6]. For example, all the stochastic primitives may be correlated and may exhibit short- or long-range dependence. In the latter case, we can consider heavy-tailed interarrival and/or service time distributions. For more information, see Whitt [15].

Consider a sequence of single-station queueing systems, indexed by superscript n, each with one customer class. Assume that $\left\{u_j^n\right\}_{j=1}^\infty$, the customer inter-arrival times, are strictly positive, i.i.d. r.v.s with mean $1/\lambda_n$ and $\left\{v_j^n\right\}_{j=1}^\infty$, the customer service times, are positive, i.i.d. r.v.s with mean

$1/\mu_n$. We assume that each queue is empty at time zero and (2.1) holds. Let c_n be a sequence of constants such that $c_n \to \infty$, $n/c_n \to \infty$. Let L_j^n denote the customer initial lead times. Assume that their distribution is given by $\mathbb{P}\{L_j^n \le c_n y\} = G^n(y)$ and such that $y^* \triangleq \min\{y \in \mathbb{R} : G^n(y) = 1\}$ is finite and independent of n. Moreover, we assume $G^n \Rightarrow G$. We assume that the random vectors $\{(v_j^n, L_j^n)\}_{j=1}^{\infty}$ are i.i.d. and that $G_v^n(y) \triangleq \mathbb{E}\left[v_j^n \mathbb{I}_{\{L_j^n \le c_n y\}}\right] \Rightarrow G_v(y)$, where G_v is a c.d.f. of a finite, positive measure on $\mathbb{R}$ with total mass $1/\lambda$. We make the *heavy traffic assumption*

$$\lim_{n\to\infty} n(1-\rho_n)/c_n = \gamma \tag{A.1}$$

for some $\gamma \in \mathbb{R}$, where $\rho_n \triangleq \lambda_n/\mu_n$. Let S^n, A^n, V^n, W^n be defined by (2.2)–(2.4), (2.9)–(2.11), and let, for $t \ge 0$,

$$\widehat{S}^n(t) \triangleq c_n^{-1}\left[S^n(nt) - \lambda_n^{-1}nt\right], \qquad \widehat{V}^n(t) \triangleq c_n^{-1}\left[V^n(nt) - \mu_n^{-1}nt\right].$$

We assume that

$$\left(\widehat{S}^n, \widehat{V}^n\right) \Rightarrow (S^*, V^*) \tag{A.2}$$

in $D_{\mathbb{R}^2}[0,\infty)$, where S^* and V^* have continuous sample paths. Thus, by Theorem 9.3.4 in Whitt [15], $\widehat{W}^n(t) \triangleq c_n^{-1}W^n(nt) \Rightarrow W^*(t)$ in $D[0,\infty)$, where, for any $t \ge 0$, $W^*(t) = N^*(t) - \inf_{0\le s\le t} N^*(s)$ and $N^* = (V^* - S^*)\circ \lambda e - \gamma e$. Finally, we assume that, for every $-\infty \le a < b \le y^*$,

$$\tilde{V}_{a,b}^n(t) \triangleq \frac{1}{c_n}\sum_{j=1}^{\lfloor nt\rfloor}\left(v_j^n \mathbb{I}_{\{c_n a < L_j^n \le c_n b\}} - (G_v^n(b) - G_v^n(a))\right) \Rightarrow V_{a,b}^*(t) \tag{A.3}$$

in $D[0,\infty)$, where $V_{a,b}^*$ has continuous sample paths. For $y \le y^*$, let $H_v(y) \triangleq \int_y^{y^*}(1-\lambda G_v(\eta))d\eta$. Let

$$F^n(t) \triangleq \left\{\begin{array}{l}\text{Largest lead time of all customers who have ever been in}\\ \text{service, whether still present or not, or } c_n\, y^* - t, \text{ if this}\\ \text{quantity is larger than the former one}\end{array}\right\},$$

and let $\mathcal{W}^n$ and $\mathcal{Q}^n$ be as in Section 2.4. Let $\widehat{F}^n(t) = c_n^{-1}F^n(nt)$ and let, for any Borel set $B \subseteq \mathbb{R}$,

$$\widehat{\mathcal{Q}}^n(t)(B) \triangleq c_n^{-1}\mathcal{Q}^n(nt)(c_n B), \qquad \widehat{\mathcal{W}}^n(t)(B) \triangleq c_n^{-1}\mathcal{W}^n(nt)(c_n B).$$

Let F^*, $\mathcal{W}^*$, $\mathcal{Q}^*$ be as in (3.1)–(3.3).

Proposition A.1. *$\widehat{F}^n \Rightarrow F^*$ in $D_{\mathbb{R}}[0,\infty)$ as $n\to\infty$.*

Theorem A.2. *$\widehat{\mathcal{W}}^n \Rightarrow \mathcal{W}^*$ and $\widehat{\mathcal{Q}}^n \Rightarrow \mathcal{Q}^*$ in $D_{\mathcal{M}}[0,\infty)$ as $n\to\infty$.*

Proposition A.1 and Theorem A.2 can be proved by a straightforward generalization of the arguments of Doytchinov et al. [6].

References

[1] Bender, M., Chakrabarti, S. and Muthukrishnan, S., *Flow and stretch metrics for scheduling continuous job streams*, Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 1998), ACM, New York, 1998, 270–279.

[2] Bickel, P. J., Wichura, M. J., *Convergence for multiparameter stochastic processes and some applications*, Ann. Math. Statist. **42** (1971), 1656–1670.

[3] Billingsley, P., *Probability and Measure*, 2nd ed., Wiley, New York, 1986.

[4] Billingsley, P., *Convergence of Probability Measures*, 2nd ed., Wiley, New York, 1999.

[5] Crovella, M. E., Frangioso, R. and Harchol-Balter, M., *Connection scheduling in Web servers*, USENIX Symposium on Internet Technologies and Systems (USITS '99), Boulder, Colorado, October 1999, 243–254.

[6] Doytchinov, B., Lehoczky, J. P. and Shreve, S. E., *Real-time queues in heavy traffic with earliest-deadline-first queue discipline*, Ann. Appl. Probab. **11** (2001), 332–378.

[7] Iglehart, D., Whitt, W., *Multiple channel queues in heavy traffic I*, Advances in Appl. Probability **2** (1970), 150–177.

[8] Karatzas, I., Shreve, S. E., *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.

[9] Kruk, Ł., Lehoczky, J. P. and Shreve, S. E., *Second order approximation for the customer time in queue distribution under the FIFO service discipline*, Ann. Univ. Mariae Curie-Skłodowska Sect. AI Inform. **1** (2003), 37–48.

[10] Kruk, Ł., Lehoczky, J. P. and Shreve, S. E., *Accuracy of state space collapse for earliest-deadline-first queues*, Ann. Appl. Probab. **16** (2006), 516–561.

[11] Kruk, Ł., Lehoczky, J. P., Shreve, S. E. and Yeung, S. N., *Multiple-input heavy-traffic real-time queues*, Ann. Appl. Probab. **13** (2003), 54–99.

[12] Kruk, Ł., Lehoczky, J. P., Shreve, S. E. and Yeung, S. N., *Earliest-deadline-first service in heavy-traffic acyclic networks*, Ann. Appl. Probab. **14** (2004), 1306–1352.

[13] Prokhorov, Yu., *Convergence of random processes and limit theorems in probability theory*, Theory Probab. Appl. **1** (1956), 157–214.

[14] van der Vaart, A., Wellner, J. A., *Weak Convergence and Empirical Processes*, Springer-Verlag, New York, 1996.

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

[16] Yeung, S. N., Lehoczky, J. P., *Real-time queueing networks in heavy traffic with EDF and FIFO queue discipline*, preprint (2004).

Łukasz Kruk
Institute of Mathematics
M. Curie-Skłodowska University
pl. Marii Curie-Skłodowskiej 1
20-031 Lublin, Poland
e-mail: `lkruk@hektor.umcs.lublin.pl`

Received April 4, 2006