Two-Parameter Heavy-Traffic Limits for Infinite-Server Queues With Dependent Service Times

Guodong Pang* and Ward Whitt**

*Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University, University Park, PA, 16802
Email: gup3@psu.edu

**Department of Industrial Engineering and Operations Research
Columbia University, New York, NY, 10027
Email: ww2040@columbia.edu

October 5, 2011

Abstract

This paper is a sequel to our 2010 paper in this journal in which we established heavy-traffic limits for two-parameter processes in infinite-server queues with an arrival process that satisfies a FCLT and i.i.d. service times with a general distribution. The arrival process can have a time-varying arrival rate. In particular, a FWLLN and a FCLT were established for the two-parameter process describing the number of customers in the system at time t that have been so for a duration y. The present paper extends the previous results to cover the case in which the service times are weakly dependent. The deterministic fluid limit obtained from the new FWLLN is unaffected by the dependence, whereas the Gaussian process limit (random field) obtained from the FCLT has a term resulting from the dependence. Explicit expressions are derived for the time-dependent means, variances and covariances for the common case in which the limit process for the arrival process is a (possibly time scaled) Brownian motion.

Key words: infinite-server queues, two-parameter processes, time-varying arrivals, martingales, weakly dependent service times, ϕ-mixing, S-mixing, functional central limit theorems, Gaussian (random field) approximation, generalized Kiefer process

1 Introduction

This paper is a sequel to Pang and Whitt (2010), in which we established heavy-traffic limits for the stochastic processes describing performance of the $G_t/GI/\infty$ infinite-server (IS) model, allowing a non-Poisson arrival process with time-varying arrival rate and a non-exponential service-time distribution. Extending Krichagina and Puhalskii (1997), we established heavy-traffic limits for two-parameter stochastic processes, such as $\{Q^e(t,y) : t \geq 0, 0 \leq y \leq t\}$ and $\{Q^r(t,y) : t \geq 0, y \geq 0\}$, where $Q^e(t,y)$ represents the number of customers in the system at time t with elapsed service times less than or equal to y, and $Q^r(t,y)$ represents the number of customers in the system at time t with residual service time strictly greater than y. Moreover, we showed that these limit processes are Markov processes. A key assumption was that the arrival process satisfy a functional central limit theorem (FCLT), which includes many cases with dependence among the interarrival times. Including time-varying arrival rates is important too, because that allows applications to approximate the performance of large-scale service systems, which usually have time-varying arrival rates; see Green et al. (2007).

In the present paper we establish new heavy-traffic limits that extend our previous results by allowing the service times to be weakly dependent; we refer to our model as $G_t/G^D/\infty$ queue. Roughly, weak dependence means that the dependence among the service times is limited so that the CLT remains valid, but the variability constant in the CLT is affected by the cumulative correlations; see §4.4 of Whitt (2002). In the present context, weak dependence among the service times is especially interesting, because, as shown by Krichagina and Puhalskii (1997) and Pang and Whitt (2010), in the iid case the service times affect the heavy-traffic limit through the sequential empirical process (see (2.5) below) rather than the conventional CLT.

We are motivated to study dependence among service times by several applications. First, in hospitals, several patients can have similar medical conditions, requiring similar treatment. That occurs with seasonal or epidemic diseases and with multi-person transportation accidents, as with cars or trains. Second, in technical support customer contact centers, new products may have defects that lead to many customers calling with similar needs. These patients or customers will have service requests that are highly dependent upon each other. Third, service times can be affected by common events in the service mechanism. For instance, service interruptions are inevitable in many large-scale service systems, e.g., Pang and Whitt (2009), and interruptions can cause all service times to become longer or stimulate the servers to interact with each other in order to reduce the effect.

There has been considerable work on IS models. Since we already reviewed earlier work on IS queues in Pang and Whitt (2010), here we only discuss models with dependent service times. Very few articles have studied the IS models with dependent interarrival times and dependent service times. Falin (1994) considers the $M^k/G/\infty$ batch arrival queue with heterogeneous dependent demands. Liu and Templeton (1993) give explicit formulas for the autocorrelation in IS queues with batch arrivals including dependence structure. Our paper is evidently the first to establish heavy-traffic limits for IS models with dependent service times. These limits are useful, not only to yield direct approximations for large-scale queues when the arrival rate is high, but also they aid in establishing associated many-server heavy-traffic limits for queues with finitely many-servers; e.g., see Liu and Whitt (2011).

We analyze this $G_t/G^D/\infty$ model in the heavy-traffic regime by scaling up the arrival rates while fixing the service-time distributions. We consider the two-parameter stochastic process $\{Q^e(t,y) : t \geq 0, 0 \leq y \leq t\}$, where $Q^e(t,y)$ represents the number of customers in the system at time t with elapsed service times less than or equal to y. (As shown in Pang and Whitt (2010), equivalent results can be obtained for the process $Q^r(t,y)$, and thus we only focus on $Q^e(t,y)$ here.) We prove a functional weak law of large numbers (FWLLN, Theorem 3.1) and an FCLT (Theorem 3.2) for this process jointly with the departure process from the system. The FWLLN limits are simple deterministic two-parameter functions and the FCLT limits are continuous two-parameter Gaussian processes (random fields). Propositions 3.2 and 3.3 provide explicit variance formulas for the Gaussian limit processes when the arrival limit process is a Brownian motion (BM). Dependence among the service times has no impact upon the fluid limit (the mean), but has a clear impact upon the variances; we study this impact further in Pang and Whitt (2011a, 2011b).

In order to allow dependence among the service times, we exploit previous FCLT's for the sequential empirical process of weakly dependent random variables satisfying the ϕ-mixing or S-mixing conditions, by Berkes and Philipp (1977) and Berkes, Hörmann and Schauer (2009), respectively. One key step in proving our limits is to show that the sequential empirical processes with the underlying weakly dependent service times converge in distribution to a continuous generalized Kiefer process, in the space of $D_D \equiv D([0,\infty)D([0,\infty),\mathbb{R}))$ endowed with the Skorohod J_1 topology, see Theorem 2.1. The previous results were established in the space of $D([0,1]\times[0,1],\mathbb{R})$ endowed with the generalized Skorohod topology by Bickel and Wichura (1971) and Straf (1971). Here we need to extend the convergence to the larger space D_D because the two-parameter queueing process $Q^e(t,y)$ are not in the space $D([0,T] \times [0,T], \mathbb{R})$.

To establish the FCLT limit of $Q^e(t,y)$ under the assumptions of service times satisfying ϕ-mixing or S-mixing conditions, we employ the same approach as Pang and Whitt (2010) by proving the tightness of the processes together with the convergence of their finite-dimensional distributions. However, the methods to prove tightness and convergence of finite-dimensional distributions here are completely different from those in Pang and Whitt (2010). Before, following Krichagina and Puhalskii (1997), we were able to apply a semimartingale decomposition for the sequential empirical processes with underlying iid random variables and standard Kiefer processes. However, now we do not have an analogous semimartingale decomposition for sequential empirical processes with underlying weakly dependent random variables and generalized Kiefer processes. Instead, we construct martingale difference sequences from the weakly dependent sequences, see §4.3.

Here is how the rest of this paper is organized. In §2, we give the detailed model description and assumptions. We also establish some preliminary results including the FCLT for the sequential empirical processes with underlying weakly dependent sequences in D_D, see Theorem 2.1, and the representation of the process Q^e_n in terms of the sequential empirical processes, Lemma 2.1. In §3, we state our main results, the FWLLN in §3.1, the FCLT in §3.2, and the characterization of Gaussian properties in §3.3. We collect the proofs for the main results in §4. In §4.1, we prove Theorem 2.1; in §4.2, we prove the Gaussian characterization of the FCLT limits; in §4.3, we prove the FCLT. We draw conclusions in §5.

2 The Model and Preliminaries

2.1 The Model Assumptions

We consider a sequence of $G_t/G^D/\infty$ queueing models indexed by n and then let $n \to \infty$, where the arrival rate increases in n. We assume the system starts empty at time 0. As in Pang and Whitt (2010), we would analyze other initial content separately, which can be done because capacity is unlimited. For the n^{th} system, the i^{th} customer arrives at the time τ_i^n with the service time η_i and receives service upon arrival. Let $A_n \equiv \{A_n(t) : t \geq 0\}$ be the arrival counting process in the n^{th} system. We assume that the sequence of arrival processes satisfies a FCLT as in Pang and Whitt (2010). All single-parameter continuous-time processes are assumed to be random elements in the function space $D \equiv D([0,\infty), \mathbb{R})$ with the Skorohod J_1 topology (Billingsley (1999), Whitt(2002)).

Assumption 1: FCLT for arrivals. There exist: (i) a continuous nondecreasing deterministic real-valued function $\bar{a}$ on $[0,\infty)$ with $\bar{a}(0) = 0$ and (ii) a stochastic process $\hat{A}$ in D with continuous sample paths, such that

$$\hat{A}_n(t) \equiv n^{-1/2}(A_n(t) - n\bar{a}(t)) \Rightarrow \hat{A}(t) \quad \text{in} \quad D \quad \text{as} \quad n \to \infty. \tag{2.1}$$

■

As an immediate consequence of Assumption 1, we have the associated FWLLN

$$\bar{A}_n \equiv n^{-1}A_n(t) \Rightarrow \bar{a}(t) \quad \text{in} \quad D \quad \text{as} \quad n \to \infty. \tag{2.2}$$

The Standard Case. The standard case concerns a stationary model in which the limit of the arrival process FCLT is Brownian motion. In the assumed arrival FWLLN, $\bar{a} = \lambda t$, $t \geq 0$, for some positive constant λ. The limit in the FCLT is $\hat{A} = \sqrt{\lambda c_a^2} B_a$, i.e., a Brownian motion (BM), where c_a^2 is variability parameter, which for a renewal arrival process is the squared coefficient of variation (SCV) of an interarrival times, and B_a is a standard BM. ■

Here we emphasize that the assumption in (2.1) on the arrival processes A_n includes the cases where the interarrival times are correlated, see Theorem 4.4.1 and 7.3.2 of Whitt (2002). In the standard case, the variability parameter c_a^2 will capture the correlation effect among inter arrival times.

We will allow the service times to be weakly dependent and consider two types of weak dependence for stationary stochastic sequences: ϕ-mixing and S-mixing. The ϕ-mixing is a common condition for weakly dependent stationary sequence, see Billingsley (1999) and Whitt (2002). Here we restate the definition of S-mixing, first introduced by Berkes, Hörmann and Schauer (2009). A stationary stochastic sequence $\{x_i : i \geq 1\}$ is called S-mixing if (i) for any $i, m \geq 1$, there exists a random variable x_{im} such that $P(|x_i - x_{im}| \geq \beta_m) \leq \epsilon_m$ for some constant sequences $\beta_m \to 0$ and $\epsilon_m \to 0$ as $m \to \infty$; (ii) for any disjoint intervals $I_1, ..., I_r$ of positive integers and any positive integers $m_1, ..., m_r$, the vectors $\{x_{im_1} : i \in I_1\}$,...,$\{x_{im_r} : i \in I_r\}$ are independent provided that the separation between $I_{r'}$ and $I_{r''}$, $1 \leq r', r'' \leq r$, is greater than $m_{r'} + m_{r''}$. Berkes, Hörmann and Schauer (2009) show that neither of the two mixing condition includes the other, but the S-mixing condition is relatively easy to verify because it is restricted to random sequences $\{x_i : i \geq 1\}$ with representations that $x_i = \psi(y_i, y_{i+1}, ...)$ for iid sequences $\{y_i : i \geq 1\}$ and Borel measurable functions $\psi : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$.

Assumption 2: weakly dependent service times. We assume that the successive service times $\{\eta_i : i \geq 1\}$ are weakly dependent and constitute a one-sided stationary sequence. We also assume that η_i's have the same continuous c.d.f. F and p.d.f. f with $F(0) = 0$, and $E[\eta_1^2] < \infty$, and

$$\sum_{i=1}^{\infty}(E[(E[\eta_{i+k}|\mathcal{F}_k^s])^2])^{1/2} < \infty, \quad k = 1, 2, ...,$$

where $\mathcal{F}_k^s \equiv \sigma\{\eta_i : 1 \le i \le k\}$. We let

$$\mu \equiv E[\eta_1], \quad \sigma^2 \equiv Var(\eta_1) + 2\sum_{i=1}^{\infty} Cov(\eta_1, \eta_{1+i}) < \infty.$$

Moreover, we assume that one of the following two types of mixing conditions holds for both $\{\eta_i : i \ge 1\}$:

(i) (ϕ-mixing) Define

$$\phi_k \equiv \sup\{|P(B|A) - P(B)| : A \in \mathcal{F}_m^s, P(A) > 0, B \in \mathcal{G}_{m+k}^s, m \ge 1\},$$

where $\mathcal{G}_k^s = \sigma\{\eta_i : i \ge k\}$. The two sequences satisfy the ϕ-mixing condition:

$$\sum_{k=1}^{\infty} \phi_k < \infty.$$

(ii) (S-mixing) Each of the two sequences is S-mixing.

■

2.2 Preliminaries

Let $Q_n^e(t, y)$ represent the number of new arrivals in the system at time t in the n^{th} model that have elapsed service times less than or equal to y, $0 \le y \le t$. Then we can express $Q_n^e(t, y)$ as

$$Q_n^e(t, y) = \sum_{i=A_n(t-y)}^{A_n(t)} \mathbf{1}(\tau_i^n + \eta_i > t), \quad t \ge 0, \quad 0 \le y \le t, \tag{2.3}$$

Note that $Q_n^e(t, t)$ counts the total number of customers receiving service in the system at time t. Evidently, we have the balance equation

$$A_n(t) = Q_n^e(t, t) + D_n(t), \quad t \ge 0,$$

where $D_n \equiv \{D_n(t) : t \ge 0\}$ is the departure process in the n^{th} system.

The processes Q_n^e and their limits (after scaling) to be established lie in the space $D_D \equiv D([0, \infty), D([0, \infty), \mathbb{R}))$, where $D \equiv D([0, \infty), S)$, for a separable metric space S, is the space of all right-continuous S-valued functions with left-limits in $(0, \infty)$; see Billingsley (1999) and Whitt

(2002) for background. We will be using the standard Skorohod J_1 topologies on both D spaces in D_D. For a discussion of D_D, see Talreja and Whitt (2008) and Pang and Whitt (2010).

Following Krichagina and Puhalskii (1998) and Pang and Whitt (2010), we can rewrite the random sum in (2.3) as integrals with respect to the random field

$$Q^e_n(t,y) = n\int_{t-y}^{t}\int_0^\infty \mathbf{1}(s+x>t)d\bar{K}_n(\bar{A}_n(s),x), \quad t\ge 0, \quad 0\le y\le t, \tag{2.4}$$

where the two-parameter random fields $\bar{K}_n$ in D_D are defined by

$$\bar{K}_n(t,x) \equiv \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\mathbf{1}(\eta_i\le x), \quad t\ge 0, \quad x\ge 0. \tag{2.5}$$

The integral in (2.4) is well defined as a Stieltjes integral for functions of bounded variation as integrators.

These two-parameter random fields are often called sequential empirical processes. For the case of iid service times for IS queues, the FWLLN and FCLT for such random fields is discussed in Pang and Whitt (2010). Here, for weakly dependent service times, the corresponding FCLT was established by Berkes and Philipp (1977) for ϕ-mixing sequences and by Berkes, Hörmann and Schauer (2009) for S-mixing sequences, where the convergence is in the space of $D([0,1]\times[0,1],\mathbb{R})$ with the generalized Skorohod J_1 topology on two-parameter processes (Bickel and Wichura (1971) and Straf (1971)). Here we first extend their results to the space D_D with the Skorohod J_1 topology on both D spaces (recall that the space $D([0,1]\times[0,1],\mathbb{R})\subset D([0,1],D([0,1],\mathbb{R}))$). The proof is in §4.1.

For a sequence of random variables $\{\xi_k : k\ge 1\}$, each uniformly distributed on $[0,1]$, let $\gamma_k(x)\equiv \mathbf{1}(\xi_k\le x)-x$ and

$$\Gamma(x,y) = E[\gamma_1(x)\gamma_1(y)] + \sum_{k=2}^{\infty}\Big(E[\gamma_1(x)\gamma_k(y)] + E[\gamma_1(y)\gamma_k(x)]\Big), \quad x,y\in[0,1]. \tag{2.6}$$

Let the diffusion-scaled sequential empirical processes $\hat{U}_n(t,x)$ be defined by

$$\hat{U}_n(t,x) \equiv \frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor}\gamma_k(x), \quad t\ge 0, \quad x\in[0,1]. \tag{2.7}$$

Theorem 2.1 (*FCLT in D_D for the sequential empirical process with weakly dependent random variables*) *Let $\{\xi_k : k\ge 1\}$ be a weakly dependent stationary sequence of random variables uniformly distributed on $[0,1]$, either (i) ϕ-mixing or (ii) S-mixing. Assume that*

$$\sum_{i=1}^{\infty}\|E[\xi_{i+k}|\mathcal{F}_k]\|_{L^2} = \sum_{i=1}^{\infty}(E[(E[\xi_{i+k}|\mathcal{F}_k])^2])^{1/2} < \infty \tag{2.8}$$

where $\mathcal{F}_k \equiv \sigma\{\xi_i : 1 \le i \le k\}$ for each $k \ge 1$. Then, the series $\Gamma(x,y)$ in (2.6) converges absolutely and

$$\hat{U}_n \Rightarrow \hat{U} \quad \textit{in} \quad D([0,\infty), D([0,1],\mathbb{R})) \quad \textit{as} \quad n \to \infty, \tag{2.9}$$

for $\hat{U}_n$ in (2.7), where $\hat{U}$ is a generalized Kiefer process (continous two-parameter Gaussian process) with $E[\hat{U}(t,x)] = 0$ and $E[\hat{U}(t,x)\hat{U}(s,y)] = (t\wedge s)\Gamma(x,y)$ with $\Gamma(x,y)$ defined in (2.6) for any $t,s \ge 0$ and $x,y \in [0,1]$. Moreover, the convergence is uniform in the second parameter $x \in [0,1]$.

The convergence in (2.9) implies that the fluid-scaled sequential processes satisfy the FWLLN:

$$\bar{U}_n(t,x) \equiv \frac{1}{n}\sum_{k=1}^{\lfloor nt \rfloor} \mathbf{1}(\xi_k \le x) \Rightarrow \bar{u}(t,x) \equiv tx, \quad \text{in} \quad D([0,\infty), D([0,1],\mathbb{R})) \quad \text{as} \quad n \to \infty \tag{2.10}$$

Moreover, in Theorem 2.1, when the sequence $\{\xi_k\}$ is iid, the limit process $\hat{U}$ becomes a standard Kiefer process, where $\Gamma(x,y) = x \wedge y - xy$ for $x,y \in [0,1]$.

Theorem 2.1 for uniform random variables implies associated results for the random variables η_i, using the fact that $F(\eta_i)$ is distributed the same as ξ_i, implying that $\mathbf{1}(\eta_i \le x) = \mathbf{1}(F(\eta_i) \le F(x) \stackrel{\mathrm{d}}{=} \mathbf{1}(\xi_i \le F(x))$. Thus, the two-parameter random fields in (2.5) satisfy the FWLLN:

$$\bar{K}_n \Rightarrow \bar{k} \quad \text{in} \quad D_D \quad \text{as} \quad n \to \infty,$$

where $\bar{k}(t,x) = tF(x)$, and the convergence is uniform over sets of the form $[0,T]\times[0,\infty)$ and there is uniformity in the second argument x over $[0,\infty)$. Define the scaled processes

$$\hat{K}_n(t,x) \quad \equiv \quad \sqrt{n}(\bar{K}_n(t,x) - \bar{k}(t,x)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor}(\mathbf{1}(\eta_i \le x) - F(x)) \stackrel{\mathrm{d}}{=} \hat{U}_n(t,F(x)), \quad t,x \ge 0,$$

where $\hat{U}_n(t,x)$ is defined in (2.7). This implies that the FCLT for $\hat{K}_n$ holds; in particular,

$$\hat{K}_n \Rightarrow \hat{K} \quad \text{in} \quad D_D \quad \text{as} \quad n \to \infty, \tag{2.11}$$

where $\hat{K}$ is a time-changed generalized Kiefer process

$$\hat{K}(t,x) = \hat{U}(t,F(x)), \quad t,x \ge 0,$$

independent of $\hat{A}$, with mean 0 and covariance

$$E[\hat{K}(t,x)\hat{K}(s,y)] = (t\wedge s)\Gamma_K(x,y), \quad t,s,x,y \ge 0, \tag{2.12}$$

$$\Gamma_K(x,y) = [F(x)\wedge F(y) - F(x)F(y)] + \Gamma_K^c(x,y) < \infty, \tag{2.13}$$

$$\Gamma_K^c(x,y) = \sum_{k=2}^{\infty} \Big(E[\bar{\gamma}_1(x)\bar{\gamma}_k(y)] + E[\bar{\gamma}_1(y)\bar{\gamma}_k(x)] \Big) < \infty, \tag{2.14}$$

for each $x, y \geq 0$, where $\bar{\gamma}_k(x) \equiv \mathbf{1}(\eta_k \leq x) - F(x)$ for $k \geq 1$.

In the case of iid service times, $\hat{U}(t,x)$ is the standard Kiefer process, and $U(t,x) = W(t,x) - xW(t,1)$ for standard Brownian sheet W, so that $\hat{K}$ is a standard Kiefer process with the second parameter having a time change by the service-time distribution, $\Gamma_K(x,y) = F(x) \wedge F(y) - F(x)F(y)$.

Then we obtain the following representation of the processes Q_n^e. The proof follows from the same argument as Lemma 2.1 in Pang and Whitt (2010) and thus is omitted.

Lemma 2.1 (*Queue-length representation by sequential empirical processes*) *The processes Q_n^e defined in* (2.3) *can be represented as*

$$Q_n^e(t,y) = n \int_{t-y}^{t} F^c(t-s) d\bar{a}(s) + \sqrt{n}(\hat{X}_{n,1}^e(t,y) + \hat{X}_{n,2}^e(t,y)), \tag{2.15}$$

where

$$\begin{aligned} \hat{X}_{n,1}^e(t,y) &= \int_{t-y}^{t} F^c(t-s) d\hat{A}_n(s) \qquad (2.16) \\ &= \hat{A}_n(t) - F^c(y)\hat{A}_n(t-y) - \int_{t-y}^{t} \hat{A}_n(s-) dF^c(t-s), \end{aligned}$$

$$\hat{X}_{n,2}^e(t,y) = \int_{t-y}^{t} \int_{0}^{\infty} \mathbf{1}(s+x>t) d\hat{R}_n(s,x) = -\int_{t-y}^{t} \int_{0}^{\infty} \mathbf{1}(s+x \leq t) d\hat{R}_n(s,x), \tag{2.17}$$

with the integrals in (2.16) *and* (2.17) *defined as Stieltjes integrals for functions of bounded variation as integrators, and*

$$\begin{aligned} \hat{R}_n(t,x) &= \hat{K}_n(\bar{A}_n(t),x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{A_n(t)} \gamma_i(x) \\ &= \sqrt{n}\bar{K}_n(\bar{A}_n(t),x) - \hat{A}_n(t)F(x) - \sqrt{n}\bar{a}(t)F(x). \qquad (2.18) \end{aligned}$$

3 Main Results

In this section, we will present the main results, the heavy-traffic FWLLN and FCLT limits for the queue-length process, and also give explicit Gaussian characterizations of the limit processes.

3.1 FWLLN Limits

We first define the LLN-scaled processes $(\bar{D}_n, \bar{Q}_n^e) \equiv n^{-1}(D_n, Q_n^e)$. By Lemma 2.1, these LLN-scaled processes can be represented as

$$\bar{Q}_n^e(t,y) = \int_{t-y}^{t} F^c(t-s) d\bar{a}(s) + \frac{1}{\sqrt{n}}(\hat{X}_{n,1}^e(t,y) + \hat{X}_{n,2}^e(t,y)), \quad t \geq 0, \quad 0 \leq y \leq t, \tag{3.1}$$

$$\bar{D}_n(t) = \bar{A}_n(t) - \bar{Q}^e_n(t,t), \quad t \geq 0. \tag{3.2}$$

The FWLLN limits for these processes are given in the following theorem. The proof for the convergence of the processes Q^e_n simply follows from tightness of the processes $\hat{X}^e_{n,1}$ and $\hat{X}^e_{n,2}$ to be established as a main component in proving the FCLT limits. The convergence of other processes follows from applying the continuous mapping theorem (CMT). Thus, the proof for the following theorem is omitted.

Theorem 3.1 (FWLLN with weakly dependent service times) *Under Assumptions 1 - 2,*

$$(\bar{A}_n, \bar{D}_n, \bar{Q}^e_n) \Rightarrow (\bar{a}, \bar{d}, \bar{q}^e) \quad in \quad D^2 \times D_D \quad as \quad n \to \infty \tag{3.3}$$

where the limits are all deterministic functions,

$$\bar{q}^e(t,y) = \int_{t-y}^{t} F^c(t-s)d\bar{a}(s), \quad t \geq 0, \quad 0 \leq y \leq t, \tag{3.4}$$

$$\bar{d}(t) = \bar{a}(t) - \bar{q}^e(t,t) = \int_0^t F(t-s)d\bar{a}(s), \quad t \geq 0, \tag{3.5}$$

We remark that the weak dependence among service times does not affect the fluid limits, which are the same as the case of iid service times.

Corollary 3.1 (FWLLN in the standard case) *In the standard case, the limits in* (3.3) *simplify as follows,*

$$\bar{q}^e(t,y) = \lambda \int_{t-y}^{t} F^c(t-s)ds = \lambda \int_0^y F^c(s)ds = (\lambda/\mu)F_e(y) \equiv \bar{q}^e(\infty, y), \tag{3.6}$$

$$\bar{d}(t) = \lambda \int_0^t F(t-s)ds = \lambda \int_0^t F(s)ds, \quad t \geq 0, \tag{3.7}$$

where F_e is the stationary-excess (or residual-lifetime) cdf associated with the service-time cdf F, defined by $F_e(x) = \mu \int_0^x F^c(s)ds$ for each $x \geq 0$, and $\bar{d}'(t) = \lambda F(t) \to \lambda$ as $t \to \infty$.

3.2 FCLT Limits

We first define the FCLT-scaled processes associated with (D_n, Q^e_n):

$$\hat{D}_n \equiv \sqrt{n}(\bar{D}_n - \bar{d}), \quad \hat{Q}^e_n \equiv \sqrt{n}(\bar{Q}^e_n - \bar{q}^e), \tag{3.8}$$

where $\bar{d}$ and $\bar{q}^e$ are defined in Theorem 3.1. By Lemma 2.1, the processes Q^e_n can be represented as

$$\hat{Q}^e_n(t,y) = \hat{X}^e_{n,1}(t,y) + \hat{X}^e_{n,2}(t,y), \quad t \geq 0, \quad 0 \leq y \leq t, \tag{3.9}$$

and it is clear that

$$\hat{D}_n(t) = \hat{A}_n(t) + \hat{Q}_n^e(t,t), \quad t \geq 0. \tag{3.10}$$

The limit of the processes $\hat{X}^e_{n,2}(t,y)$ are given as mean-square integrals of the time-changed generalized Kiefer process $\hat{K}(t,x)$ in (2.11). Here we first give the definition of the limit.

Definition 3.1 *The two-parameter process $\hat{X}^e_2$, written as*

$$\hat{X}^e_2(t,y) = \int_{t-y}^t \int_0^\infty \mathbf{1}(s+x>t)d\hat{K}(\bar{a}(s),x) = -\int_{t-y}^t \int_0^\infty \mathbf{1}(s+x\leq t)d\hat{K}(\bar{a}(s),x), \tag{3.11}$$

is defined by a mean-square integral, i.e.,

$$\lim_{k\to\infty} E[(\hat{X}^e_2(t,y) - \hat{X}^e_{2,k}(t,y))^2] = 0, \quad t\geq 0, \quad 0\leq y\leq t, \tag{3.12}$$

with

$$\hat{X}^e_{2,k}(t,y) = \int_{t-y}^t \int_0^\infty \mathbf{1}_{k,t,y}(s,x)d\hat{K}(\bar{a}(s),x), \quad t\geq 0, \quad 0\leq y\leq t, \tag{3.13}$$

$$\mathbf{1}_{k,t,y}(s,x) \equiv \sum_{i=1}^k [\mathbf{1}(s^k_{i-1} < s \leq s^k_i)\mathbf{1}(t - s^k_i < x \leq t)] \tag{3.14}$$

$t-y = s^k_0 < s^k_1 < \cdots < s^k_k = t$ *and* $\max_{1\leq i\leq k}|s^k_i - s^k_{i-1}| \to 0$ *as* $k\to\infty$. *Write* $\hat{X}^e_2(t,y) = l.i.m._{k\to\infty}\hat{X}^e_{2,k}(t,y)$

Theorem 3.2 (FCLT with weakly dependent service times) *Under Assumptions 1-2,*

$$(\hat{A}_n, \hat{D}_n, \hat{Q}^e_n) \Rightarrow (\hat{A}, \hat{D}, \hat{Q}^e) \quad in \quad D^2\times D_D \quad as \quad n\to\infty, \tag{3.15}$$

where

$$\hat{Q}^e(t,y) \equiv \hat{X}^e_1(t,y) + \hat{X}^e_2(t,y), \quad t\geq 0, \quad 0\leq y\leq t, \tag{3.16}$$

$$\hat{X}^e_1(t,y) = \int_{t-y}^t F^c(t-s)d\hat{A}(s) = \hat{A}(t) - F^c(y)\hat{A}(t-y) - \int_{t-y}^t \hat{A}(s)dF^c(t-s), \tag{3.17}$$

$\hat{X}^e_2$ *is defined in* (3.11)

$$\begin{aligned}\hat{D}(t) &= \hat{A}(t) - \hat{Q}^e(t,t) = \int_0^t F(t-s)d\hat{A}(s) - \hat{X}^e_2(t,t) \\ &= \int_0^t \hat{A}(s)dF^c(t-s) - \hat{X}^e_2(t,y),\end{aligned} \tag{3.18}$$

where $\hat{A}$ is given in Assumption 1, $\hat{X}^e_1$ and $\hat{D}$ take the first expression in (3.17) *and* (3.18) *respectively if $\hat{A}$ is a BM and the second if $\hat{A}$ is a general Gaussian process.*

We remark about the impact of weak dependence of service times upon various processes. Weak dependence of service times affects the FCLT limits of the number of customers in the system and departure process, in particular, in the $\hat{X}_2^e$ term with $\hat{K}$ capturing the effect, see its covariance formula $\Gamma^c_K(x,y)$ in (2.14). These effects are all captured in the variance formulas for these processes when the arrival limit process is a BM, see Propositions 3.2 and 3.3.

Special Case I: EARMA(1,1) Service Times. Jacobs and Lewis (1977) proposed an approach to generate a stationary sequence of dependent random variables from a sequence of iid exponential random variables, the so-called EARMA(1,1) sequence, and Jacobs (1980) applied such stationary sequences to study single server queues with dependent service and interarrival times. The stationary EARMA(1,1) sequence satisfies the ϕ-mixing condition, see Jacobs (1980). We apply this to the IS and many-server models with dependent service times and conducted simulations to evaluate their performance in Pang and Whitt (2011a). ■

Special Case II: Batch Arrivals. Suppose that at each arrival time τ_i^n, $i = 1, 2, ...$, there are a random number $\mathcal{B}_i$ of service requests entering the system at the same time, where $\{\mathcal{B}_i : i = 1, 2, ...\}$ is a sequence of iid random variables with a common distribution. Let $p_{\mathcal{B},k} = P(\mathcal{B}_i = k)$ and $\sum_{k=1}^{\infty} p_{\mathcal{B},k} = 1$. Suppose that $E[\mathcal{B}_i] = \sum_{k=1}^{\infty} k p_{\mathcal{B},k} < \infty$ and $E[\mathcal{B}_i^2] = \sum_{k=1}^{\infty} k^2 p_{\mathcal{B},k} < \infty$. The stationary excess distribution of $\mathcal{B}_i$ is given by $p^*_{\mathcal{B},k} = (E[\mathcal{B}_i])^{-1} \sum_{j=k}^{\infty} p_{\mathcal{B},j}$ for $k = 1, 2, ...$, and $E[\mathcal{B}_i^*] = (E[\mathcal{B}_i^2] + E[\mathcal{B}_i])/(2E[\mathcal{B}_i])$.

For the arrivals in the i^{th} batch, the service requirements $\{\eta_{i_1}, \eta_{i_2}, ..., \eta_{i_{\mathcal{B}_i}}\}$ are correlated, and moreover, for any i^{th} and j^{th} batches of arrivals, the service requirements $\{\eta_{i_1}, \eta_{i_2}, ..., \eta_{i_{\mathcal{B}_i}}\}$ and $\{\eta_{j_1}, \eta_{j_2}, ..., \eta_{j_{\mathcal{B}_j}}\}$ are independent. Then, the covariance function $\Gamma^c_K(x,y)$ in (2.14) becomes

$$\begin{aligned}
\Gamma^c_K(x,y) &= \sum_{i=1}^{\infty} \Big[p^*_{\mathcal{B},i} \sum_{k=2}^{i} \Big(E[\gamma_{1_1}(x)\gamma_{1_k}(y)] + E[\gamma_{1_1}(y)\gamma_{1_k}(x)] \Big) \Big] \\
&= \sum_{i=1}^{\infty} \Big[p^*_{\mathcal{B},i} \sum_{k=2}^{i} \Big(F_k(x,y) + F_k(y,x) - 2F(x)F(y) \Big) \Big], \qquad (3.19)
\end{aligned}$$

where $\gamma_{i_k}(x) = \mathbf{1}(\eta_{i_k} \le x) - F(x)$ for each service requirement $k = 1, ..., \mathcal{B}_i$ in the i^{th} batch, and $F_k(x,y)$ is the joint distribution function for each pair (η_{i_1}, η_{i_k}) of the i^{th} batch. Note that the job 1 in batch i is not necessarily the first job in the batch, but instead an arbitrary job in the batch, and thus we use the stationary-excess batch size distribution. For a comparison of the difference between the first job delay and an arbitrary job delay in a batch for single-server queues, see Whitt

(1983). It is easy to see that such sequences of service times form a stationary sequence satisfying the ϕ-mixing condition and the S-mixing condition.

Suppose, in addition, that the dependence between any two service requests among the arrivals in a batch is the same, that is, $F(x,y) = F_k(x,y)$ for each pair (η_{i_1}, η_{i_k}) of the i^{th} batch. Then the covariance function $\Gamma_K^c(x,y)$ in (3.19) can be simplified as

$$\Gamma_K^c(x,y) \quad = \quad \Big(F(x,y) + F(y,x) - 2F(x)F(y)\Big)(E[\mathcal{B}_1^*] - 1). \tag{3.20}$$

In Pang and Whitt (2011b), we study this special model in more detail. ■

3.3 Characterizing the FCLT Limit Processes

In this section, we give the Gaussian characterizations of the limit processes in Theorem 3.2. First, we give the Gaussian property of the process $\hat{X}_2^e$ defined in Definition 3.1. Recall that this process does not involve the limit process $\hat{A}$. In §4.2, we will prove Proposition 3.1. The Gaussian property of $\hat{Q}^e(t,y)$ is then simply obtained by combining the Gaussian property of $\hat{X}_1^e$ together with that of $\hat{X}_2^e$ since $\hat{X}_1^e$ and $\hat{X}_2^e$ are independent. The Gaussian property of $\hat{X}_1^e$ follows from applying Ito's isometry property, see Karatzas and Shreve (1991).

Proposition 3.1 (Gaussian property of $\hat{X}_2^e$) *Under Assumptions 1 and 2, the two-parameter process $\hat{X}_2^e$ in* (3.11) *is a well-defined continuous Gaussian process with mean* 0 *and covariance*

$$\begin{aligned} E[\hat{X}_2^e(t_1,y_1)\hat{X}_2^e(t_2,y_2)] \quad &= \quad \int_{(t_1-y_1)\vee(t_2-y_2)}^{t_1\wedge t_2} \Big(F(t_1\wedge t_2 - s) - F(t_1-s)F(t_2-s) \\ &\qquad + \Gamma_K^c(t_1-s, t_2-s)\Big) d\bar{a}(s), \end{aligned} \tag{3.21}$$

Proposition 3.2 (Gaussian property with time-varying arrivals) *If, in addition to the Assumptions in Theorem 3.2, $\hat{A}(t) = \sqrt{c_a^2} B_a(\bar{a}(t))$, where B_a is a standard BM, $\bar{a}(t) = \int_0^t \lambda(s)ds$ and c_a^2 is a constant* (*the variability parameter*), *then the limit processes are all continuous Gaussian processes with*

$$\hat{Q}^e(t,y) \stackrel{\mathrm{d}}{=} N(0, \sigma^2_{Q,e}(t,y)), \quad \hat{D}(t) \stackrel{\mathrm{d}}{=} N(0, \sigma^2_D(t)), \quad t \ge 0, \quad 0 \le y \le t, \tag{3.22}$$

where

$$\sigma^2_{Q,e}(t,y) = \int_{t-y}^t \lambda(s)\Big(F^c(t-s) + (c_a^2 - 1)(F^c(t-s))^2 + \Gamma_K^c(t-s, t-s)\Big)ds, \tag{3.23}$$

$$\sigma^2_D(t) \quad = \quad \int_0^t \lambda(s)\Big(F(t-s) + (c_a^2 - 1)(F(t-s))^2 + \Gamma_K^c(t-s, t-s)\Big)ds. \tag{3.24}$$

Proposition 3.3 (Gaussian property in the standard case) *If, in addition to the Assumptions in Theorem 3.2, we have the standard case, then* (3.22) *holds with*

$$\begin{aligned}\sigma^2_{Q,e}(t,y) &= \lambda\int_{t-y}^{t}\Big(F^c(t-s)+(c_a^2-1)(F^c(t-s))^2+\Gamma_K^c(t-s,t-s)\Big)ds \\ &= \lambda\int_0^y\Big(F^c(s)+(c_a^2-1)(F^c(s))^2+\Gamma_K^c(s,s)\Big)ds=\sigma^2_{Q,e}(\infty,y), \end{aligned}\tag{3.25}$$

$$\sigma^2_D(t) = \lambda\int_0^t\Big(F(s)+(c_a^2-1)(F(s))^2+\Gamma_K^c(s,s)\Big)ds, \tag{3.26}$$

and

$$\lim_{t\to\infty}\frac{\sigma^2_D(t)}{t}=\lim_{t\to\infty}\lambda\Big[F(t)+(c_a^2-1)(F(t))^2+\Gamma_K^c(t,t)\Big)\Big]=\lambda c_a^2. \tag{3.27}$$

We observe from Propositions 3.2 and 3.3 that the dependence among service times affects the variance functions of the number of customers in the system and the departure process by simply adding an additional term involving Γ_K^c to the expressions in the case of iid service times. Moreover, in the standard case, the variability of the departure process is not affected by the dependence among service times, which is the same as the variability of the arrival process, c_a^2, as shown in (3.27).

4 Proofs

4.1 Proof of Theorem 2.1

We first show the convergence of the finite-dimensional distributions (f.d.d.'s) and then we show tightness of $\{\hat{U}_n : n\geq 1\}$ in the space $D([0,\infty), D([0,1],\mathbb{R}))$.

For the convergence of f.d.d.'s, we can apply Theorem 1 of Berkes and Philipp (1977) under the ϕ-mixing condition and Theorem A of Berkes, Hormann and Schauer (2009) under the S-mixing condition to deduce that, for $0\leq t_1<t_2<\cdots<t_k$,

$$(\hat{U}_n(t_1,\cdot),...,\hat{U}_n(t_k,\cdot))\Rightarrow(\hat{U}(t_1,\cdot),...,\hat{U}(t_k,\cdot)) \quad \text{in} \quad D([0,1],\mathbb{R})^k \quad \text{as} \quad n\to\infty, \tag{4.1}$$

where the k elements in the limit are random elements in the functional space $D([0,1],\mathbb{R})$. Then, by those two theorems above, for each t_i, we have that for each $x_{t_i,1},...,x_{t_i,j_{t_i}}$,

$$\begin{aligned}&(\hat{U}_n(t_1,x_{t_1,1}),...,\hat{U}_n(t_1,x_{t_1,j_{t_i}}),\cdots,\hat{U}_n(t_k,x_{t_k,1}),...,\hat{U}_n(t_k,x_{t_k,j_{t_k}})) \\ \Rightarrow\ &(\hat{U}(t_1,x_{t_1,1}),...,\hat{U}(t_1,x_{t_1,j_{t_i}}),\cdots,\hat{U}(t_k,x_{t_k,1}),...,\hat{U}(t_k,x_{t_k,j_{t_k}})) \quad \text{in} \quad \mathbb{R}^{j_{t_1}+\cdots+j_{t_k}} \quad \text{as} \quad n\to\infty.\end{aligned}\tag{4.2}$$

We next show the tightness of $\{\hat{U}_n : n \geq 1\}$ in $D([0,\infty), D([0,1],\mathbb{R}))$ by applying the tightness criteria in Theorem 6.2 in Pang and Whitt (2010). First, the stochastic boundedness of $\{\hat{U}_n : n \geq 1\}$ in $D([0,\infty), D([0,1],\mathbb{R}))$ follows easily from the convergence in $D([0,1]^2,\mathbb{R})$ under either the ϕ-mixing condition or the S-mixing condition.

Then, it suffices to show that

$$\lim_{\vartheta\to 0}\limsup_{n\to\infty}\sup_{\kappa_n} P\left(\sup_{t\leq\vartheta} d_{J_1}(\hat{U}_n(\kappa_n+t,\cdot),\hat{U}_n(\kappa_n,\cdot)) \geq \varsigma\right) = 0 \tag{4.3}$$

where $\{\kappa_n : n \geq 1\}$ is a sequence of uniformly bounded stopping times with respect to the natural filtration $\mathbf{G}_n \equiv \{\mathcal{G}_n(t) : t \in [0,T]\}$ with $\mathcal{G}_n(t) = \sigma\{\hat{U}_n(s,\cdot) : 0 \leq s \leq t\} \vee \mathcal{N}$ satisfying the usual conditions (complete, increasing and right continuous). Due to the fact that the Shokorod J_1 metric for any two functions in D is less than the uniform metric (§3.3, Whitt (2002)), and moreover, by easily observing that

$$\begin{aligned} & P\left(\sup_{t\leq\vartheta}\sup_{x\in[0,1]}\left|\hat{U}_n(\kappa_n+t,x)-\hat{U}_n(\kappa_n,x)\right| \geq \varsigma\right) \\ \leq\ & 2P\left(\sup_{t\leq\vartheta}\sup_{x\in[0,1/2]}\left|\hat{U}_n(\kappa_n+t,x)-\hat{U}_n(\kappa_n,x)\right| \geq \varsigma\right), \end{aligned}$$

we only need to prove that

$$\lim_{\vartheta\to 0}\limsup_{n\to\infty}\sup_{\kappa_n} P\left(\sup_{t\leq\vartheta}\sup_{x\in[0,1/2]}\left|\hat{U}_n(\kappa_n+t,x)-\hat{U}_n(\kappa_n,x)\right| \geq \varsigma\right) = 0. \tag{4.4}$$

The sequence $\{\gamma_k(x) : k \geq 1\}$ for each $x \in [0,1]$ is stationary and ergodic, because $\{\xi_k : k \geq 1\}$ is stationary and ergodic under either the ϕ-mixing condition or the S-mixing condition, and moreover,

$$E[\gamma_k(x)] = 0, \quad E[\gamma_k(x)^2] = x(1-x) \leq \frac{1}{4}, \quad \text{for all} \quad x \in [0,1].$$

We now construct a martingale difference sequence from the sequence $\{\gamma_k(\cdot) : k \geq 1\}$. We follow the idea in the proof of Theorem 19.1 in Billingsley (1999). Let $\mathbf{F} \equiv \{\mathcal{F}_k : k \geq 1\}$ be the natural filtration generated by the sequence $\{\xi_k : k \geq 1\}$, defined by $\mathcal{F}_k \equiv \sigma\{\xi_i : i \leq k\}$. Define

$$\hat{\gamma}_k(x) \equiv \sum_{i=1}^{\infty} E[\gamma_{k+i}(x)|\mathcal{F}_k], \quad x \in [0,1], \quad k = 1,2,... \tag{4.5}$$

and

$$\tilde{\gamma}_k(x) \equiv \gamma_k(x) + \hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x), \quad x \in [0,1], \quad k = 1,2,.... \tag{4.6}$$

Then, the sequence $\{\tilde{\gamma}_k(x) : k \geq 1\}$ for each $x \in [0,1]$ is a martingale difference sequence, because for each $k \geq 1$,

$$\begin{aligned} E[\tilde{\gamma}_{k+1}(x)|\mathcal{F}_k] &= E[\gamma_{k+1}(x) + \hat{\gamma}_{k+1}(x) - \hat{\gamma}_k(x)|\mathcal{F}_k] \\ &= E[\gamma_{k+1}(x)|\mathcal{F}_k] + E\Big[\sum_{i=1}^{\infty} E[\gamma_{k+1+i}(x)|\mathcal{F}_{k+1}]\Big|\mathcal{F}_k\Big] - E[\hat{\gamma}_k(x)|\mathcal{F}_k] \\ &= E[\gamma_{k+1}(x)|\mathcal{F}_k] + \sum_{i=1}^{\infty} E[\gamma_{k+1+i}(x)|\mathcal{F}_k] - \sum_{i=1}^{\infty} E[\gamma_{k+i}(x)|\mathcal{F}_k] \\ &= E[\gamma_{k+1}(x)|\mathcal{F}_k] - E[\gamma_{k+1}(x)|\mathcal{F}_k] = 0, \end{aligned}$$

and $E[|\tilde{\gamma}_k(x)|] < \infty$ since $E[|\hat{\gamma}_k(x)|] < \infty$ by (2.8).

Define the processes $\tilde{U}_n \equiv \{\tilde{U}_n(t,x) : t, x \geq 0\}$ by

$$\tilde{U}_n(t,x) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \tilde{\gamma}_k(x). \tag{4.7}$$

Then it follows that for each $t \geq 0$ and $x \in [0,1]$, (see the proof of Theorem 19.1 in Billingsley (1999))

$$\|\tilde{U}_n(t,x) - \hat{U}_n(t,x)\|_{L^2} = \left\| \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} (\hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x)) \right\|_{L^2} \to 0 \quad \text{as} \quad n \to \infty. \tag{4.8}$$

Hence, for each $x \in [0,1]$, κ_n and $n \geq 1$, the process $\{\tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x) : t \geq 0\}$ defined by

$$\tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x) \equiv \frac{1}{\sqrt{n}} \sum_{k=\lfloor n\kappa_n \rfloor + 1}^{\lfloor n(\kappa_n + t) \rfloor} \tilde{\gamma}_k(x) \tag{4.9}$$

is a locally square integrable martingale with respect to the filtration $\{\mathcal{G}_{\kappa_n + t} : t \geq 0\}$ by Doob's sampling theorem. The difference between $\hat{U}_n(\kappa_n + t, x) - \hat{U}_n(\kappa_n, x)$ and $\tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x)$ is asymptotically negligible as $n \to \infty$ because for $t < \vartheta$ small,

$$\begin{aligned} &\left\| \frac{1}{\sqrt{n}} \sum_{k=\lfloor n\kappa_n \rfloor + 1}^{\lfloor n(\kappa_n + t) \rfloor} (\tilde{\gamma}_k(x) - \gamma_k(x)) \right\|_{L^2} = \left\| \frac{1}{\sqrt{n}} \sum_{k=\lfloor n\kappa_n \rfloor + 1}^{\lfloor n(\kappa_n + t) \rfloor} (\hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x)) \right\|_{L^2} \\ = &\left\| \frac{1}{\sqrt{n}} (\hat{\gamma}_{\lfloor n(\kappa_n + t) \rfloor}(x) - \hat{\gamma}_{\lfloor n\kappa_n \rfloor}(x)) \right\|_{L^2} \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Thus, it suffices to show that

$$\lim_{\vartheta \to 0} \limsup_{n \to \infty} \sup_{\kappa_n} P\left(\sup_{t \leq \vartheta} \sup_{x \in [0,1/2]} \left| \tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x) \right| \geq \varsigma \right) = 0. \tag{4.10}$$

For each $x \in [0,1]$, κ_n and $n \geq 1$, the process $\{\tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x) : t \geq 0\}$ is a locally square integrable martingale with respect to the filtration $\{\mathcal{G}_{\kappa_n + t} : t \geq 0\}$, and then, by Doob's maximal inequality,

$$
\begin{aligned}
& P\left(\sup_{t\leq\vartheta}\sup_{x\in[0,1/2]}\left|\tilde{U}_n(\kappa_n+t,x)-\tilde{U}_n(\kappa_n,x)\right|\geq\varsigma\right) \\
\leq\ & \frac{1}{\varsigma^2}E\left[\sup_{x\in[0,1/2]}\left|\tilde{U}_n(\kappa_n+\vartheta,x)-\tilde{U}_n(\kappa_n,x)\right|^2\right]=\frac{1}{\varsigma^2}E\left[\sup_{x\in[0,1/2]}\left(\frac{1}{\sqrt{n}}\sum_{k=\lfloor n\kappa_n\rfloor+1}^{\lfloor n(\kappa_n+\vartheta)\rfloor}\tilde{\gamma}_k(x)\right)^2\right].
\end{aligned}
$$

Then, it is obvious that for each fixed n and k, $\{\tilde{\gamma}_k(x) : x \in [0,1]\}$ is a square integrable martingale, and so is $\{\tilde{U}_n(\kappa_n + t, x) - \tilde{U}_n(\kappa_n, x) : x \in [0,1]\}$, and thus, by Doob's maximal inequality again,

$$
\begin{aligned}
& P\left(\sup_{t\leq\vartheta}\sup_{x\in[0,1/2]}\left|\tilde{U}_n(\kappa_n+t,x)-\tilde{U}_n(\kappa_n,x)\right|\geq\varsigma\right) \\
\leq\ & \frac{1}{\varsigma^2}E\left[\left(\frac{1}{\sqrt{n}}\sum_{k=\lfloor n\kappa_n\rfloor+1}^{\lfloor n(\kappa_n+\vartheta)\rfloor}\tilde{\gamma}_k(1/2)\right)^2\right]\leq\frac{1}{\varsigma^2}(\vartheta+1/n)M_\gamma
\end{aligned}
$$

where $M_r = \sum_{k=1}^{\infty} E[\tilde{\gamma}_k(1/2)^2] < \infty$. This upper bound goes to zero as $\vartheta \to 0$ and $n \to \infty$ and thus, (4.10) holds. The proof is complete. ■

4.2 Proof of Proposition 3.1

First, since the process $\hat{K}_1$ is continuous Gaussian, the process $\hat{X}^e_{2,k}$ defined in (3.13) and (3.14) is also continuous Gaussian for each $k \geq 1$, and thus the limit as $k \to \infty$ is also Gaussian. Next, we want to calculate

$$
E[(\hat{X}^e_2(t_1,y_1)-\hat{X}^e_2(t_2,y_2))^2]=\lim_{k\to\infty}E[(\hat{X}^e_{2,k}(t_1,y_1)-\hat{X}^e_{2,k}(t_2,y_2))^2] \tag{4.11}
$$

for each $t_1 \leq t_2$ and $y_1 \leq y_2$.

Define for $t_1 \leq t_2$ and $x_1 \leq x_2$,

$$
\Delta_K(t_1,t_2,x_1,x_2)\equiv\hat{K}(\bar{a}(t_2),x_2)-\hat{K}(\bar{a}(t_1),x_2)-\hat{K}(\bar{a}(t_2),x_1)+\hat{K}(\bar{a}(t_1),x_1). \tag{4.12}
$$

Then, for $t_1 \leq t_2$ and $x_1 \leq x_2$,

$$
\begin{aligned}
& E[(\Delta_K(t_1,t_2,x_1,x_2))^2] \\
=\ & E[\hat{K}(\bar{a}(t_2),x_2)^2]+E[\hat{K}(\bar{a}(t_1),x_2)^2]+E[\hat{K}(\bar{a}(t_2),x_1)^2]+E[\hat{K}(\bar{a}(t_1),x_1)^2] \\
& -2E[\hat{K}(\bar{a}(t_2),x_2)\hat{K}(\bar{a}(t_1),x_2)]-2E[\hat{K}(\bar{a}(t_2),x_2)\hat{K}(\bar{a}(t_2),x_1)] \\
& +2E[\hat{K}(\bar{a}(t_2),x_2)\hat{K}(\bar{a}(t_1),x_1)]+2E[\hat{K}(\bar{a}(t_1),x_2)\hat{K}(\bar{a}(t_2),x_1)]
\end{aligned}
$$

$$
\begin{aligned}
& -2E[\hat{K}(\bar{a}(t_1),x_2)\hat{K}(\bar{a}(t_1),x_1)] - 2E[\hat{K}(\bar{a}(t_2),x_1)\hat{K}(\bar{a}(t_1),x_1)] \\
= \;& \bar{a}(t_2)\Gamma_K(x_2,x_2) + \bar{a}(t_1)\Gamma_K(x_2,x_2) + \bar{a}(t_2)\Gamma_K(x_1,x_1) + \bar{a}(t_1)\Gamma_K(x_1,x_1) \\
& -2\bar{a}(t_1)\Gamma_K(x_2,x_2) - 2\bar{a}(t_2)\Gamma_K(x_2,x_1) + 2\bar{a}(t_1)\Gamma_K(x_2,x_1) + 2\bar{a}(t_1)\Gamma_K(x_2,x_1) \\
& -2\bar{a}(t_1)\Gamma_K(x_2,x_1) - 2\bar{a}(t_1)\Gamma_K(x_1,x_1) \\
= \;& (\bar{a}(t_2)-\bar{a}(t_1))[\Gamma_K(x_2,x_2)+\Gamma_K(x_1,x_1)-2\Gamma_K(x_2,x_1)] \\
= \;& (\bar{a}(t_2)-\bar{a}(t_1))(F(x_2)-F(x_1))(1+F(x_1)-F(x_2)) \\
& \quad +(\bar{a}(t_2)-\bar{a}(t_1))[\Gamma^c_K(x_2,x_2)+\Gamma^c_K(x_1,x_1)-2\Gamma^c_K(x_2,x_1)] \qquad (4.13)
\end{aligned}
$$

and for $t_1 \le t_2$ and $x_1 \le x_2$, $t'_1 \le t'_2$ and $x'_1 \le x'_2$ and $t_2 < t'_1$,

$$E[\Delta_K(t_1,t_2,x_1,x_2)\Delta_K(t'_1,t'_2,x'_1,x'_2)] = 0 \qquad (4.14)$$

We choose the same set $\{s^k_i : 0 \le i \le k\}$ for $t_1 \le t_2$ and $y_1 \le y_2$ so that $t_2 - y_2 = s^k_0 < \cdots < s^k_k = t_2$ for each $k \ge 1$. Without loss of generality, assume that $t_2 - y_2 < t_1 - y_1$. Then, we can write

$$\hat{X}^e_{2,k}(t_1,y_1) - \hat{X}^e_{2,k}(t_2,y_2) = \sum_{i=1}^{k} \Delta_K(s^k_{i-1}, s^k_i, t_1 - s^k_i, t_2 - s^k_i), \qquad (4.15)$$

and by (4.13) and (4.14), we obtain

$$
\begin{aligned}
& E[(\hat{X}^e_{2,k}(t_1,y_1) - \hat{X}^e_{2,k}(t_2,y_2))^2] = \sum_{i=1}^{k} E[(\Delta_K(s^k_{i-1}, s^k_i, t_1 - s^k_i, t_2 - s^k_i))^2] \\
= \;& \sum_{i=1}^{k} (\bar{a}(s^k_i) - \bar{a}(s^k_{i-1}))\Big[(F(t_2 - s^k_i) - F(t_1 - s^k_i))(1 + F(t_1 - s^k_i) - F(t_2 - s^k_i)) \\
& \quad +[\Gamma^c_K(t_2 - s^k_i, t_2 - s^k_i) + \Gamma^c_K(t_1 - s^k_i, t_1 - s^k_i) - 2\Gamma^c_K(t_2 - s^k_i, t_1 - s^k_i)]\Big]. \qquad (4.16)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& E[(\hat{X}^e_2(t_1,y_1) - \hat{X}^e_2(t_2,y_2))^2] \\
= \;& \int_{t_2-y_2}^{t_2} \Big[(F(t_2-u) - F(t_1-u))(1 + F(t_1-u) - F(t_2-u)) \\
& \quad +[\Gamma^c_K(t_2-u,t_2-u) + \Gamma^c_K(t_1-u,t_1-u) - 2\Gamma^c_K(t_2-u,t_1-u)]\Big] d\bar{a}(u) \qquad (4.17)
\end{aligned}
$$

for each $t_1 \le t_2$ and $y_1 \le y_2$ with $t_2 - y_2 < t_1 - y_1$. The continuity property of $\hat{X}^e_2(t,y)$ in both t and y w.p.1 follows from (4.17) by applying Chebyshev's inequality and the continuity of $\bar{a}$. The covariance of $\hat{X}^e_2(t,y)$ follows from a similar argument. The proof is completed. ■

4.3 Proofs for the FCLT

Proof of Theorem 3.2 Here we outline the main steps to prove the joint convergence of the processes in (3.15). Once we prove the convergence of $\hat{Q}^e_n$, the convergence of $\hat{D}_n$ follows from applying the CMT to the addition mapping. Thus, The main task is to prove the convergence of $\hat{Q}^e_n$, for which the convergence of $\hat{X}^e_{n,1}$ follows from applying CMT to the following mapping $\phi : D \times D \to D_D$

$$\phi(x,z)(t,y) = x(t) - z(y)x(t-y) - \int_{t-y}^{t} x(s)dz(t-s), \quad t, y \geq 0 \tag{4.18}$$

where $x, z \in D$. The continuity of the mapping ϕ in the Skorohod J_1 topology follows from a similar argument as in the proof of Lemma 6.1 in Pang and Whitt (2010), and thus, is omitted. Thus, it suffices to prove the joint convergence of $\hat{X}^e_{n,1}$ and $\hat{X}^e_{n,2}$. We will take two steps: tightness (Lemma 4.1) and convergence of f.d.d.'s (Lemma 4.2). ∎

Lemma 4.1 (*Tightness*) *Under the assumptions of Theorem 3.2, the processes* $\{(\hat{A}_n, \hat{X}^e_{n,1}, \hat{X}^e_{n,2}, \hat{D}_n : n \geq 1\}$ *are tight in* $D \times D_D^2 \times D$*, and so are the processes* $\{(\hat{A}_n, \hat{Q}^e_n, \hat{D}_n) : n \geq 1\}$ *in* $D \times D_D \times D$.

Proof. The tightness of the processes $\{\hat{A}_n : n \geq 1\}$ and $\{\hat{X}^e_{n,1} : n \geq 1\}$ follows from the Assumption 1, and from applying the CMT to the mapping in (4.18).

For the tightness of $\{\hat{X}^e_{n,2} : n \geq 1\}$, we first construct a martingale difference sequence from the sequence $\{\eta_i : i \geq 1\}$. As in the proof of Theorem 2.1, we follow the idea in the proof of Theorem 19.1 in Billingsley (1999). Let $\mathbf{F} \equiv \{\mathcal{F}_k : k \geq 1\}$ be the natural filtration generated by the sequence $\{\eta_i : i \geq 1\}$, defined by $\mathcal{F}_k = \sigma\{\eta_i : i \leq k\} \vee \mathcal{N}$. Define

$$\hat{\gamma}_k(x) \equiv \sum_{i=1}^{\infty} E[\gamma_{k+i}(x)|\mathcal{F}_k], \quad x \geq 0, \quad k \geq 1, \tag{4.19}$$

and

$$\tilde{\gamma}_k(x) \equiv \gamma_k(x) + \hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x), \quad x \geq 0, \quad k \geq 1, \tag{4.20}$$

where

$$\gamma_k(x) \equiv \mathbf{1}(\eta_k \leq x) - F(x) = -(\mathbf{1}(\eta_k > x) - F^c(x)), \quad x \geq 0, \quad k \geq 1. \tag{4.21}$$

Then, it is easy to check that for each $x \geq 0$, the sequence $\{\tilde{\gamma}_k(x) : k \geq 1\}$ is a martingale difference sequence. Define

$$\tilde{K}_n(t,x) \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \tilde{\gamma}_k(x), \quad t, x \geq 0, \tag{4.22}$$

and

$$\tilde{R}_n(t,x) = \tilde{K}_n(\bar{A}_n, x) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n\bar{A}_n(t)\rfloor} \tilde{\gamma}_k(x), \quad t, x \geq 0. \tag{4.23}$$

Moreover, define the processes $\tilde{X}^e_{n,2}$ by

$$\tilde{X}^e_{n,2}(t,y) \equiv \int_{t-y}^{t} \int_0^\infty \mathbf{1}(s+x>t) d\tilde{R}_n(s,x), \quad t \geq 0, \quad 0 \leq y \leq t. \tag{4.24}$$

We now show that the difference between $\hat{X}^e_{n,2}$ and $\tilde{X}^e_{n,2}$ becomes negligible as $n \to \infty$. By the definitions of $\hat{X}^e_{n,2}$ and $\tilde{X}^e_{n,2}$, we have

$$\tilde{X}^e_{n,2}(t,y) - \hat{X}^e_{n,2}(t,y) = \int_{t-y}^{t} \int_0^\infty \mathbf{1}(s+x>t) d(\tilde{R}_n(s,x) - \hat{R}_n(s,x)), \tag{4.25}$$

where

$$\tilde{R}_n(s,x) - \hat{R}_n(s,x) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n\bar{A}_n(s)\rfloor} (\hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x)). \tag{4.26}$$

By Assumption 2,

$$E[(\hat{\gamma}_k(x))^2] = E[(\hat{\gamma}_{k-1}(x))^2] < \infty, \quad k \geq 1, \quad x \geq 0, \tag{4.27}$$

and similar to (4.8),

$$\left\| \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt\rfloor} (\hat{\gamma}_k(x) - \hat{\gamma}_{k-1}(x)) \right\|_{L^2} \to 0 \quad \text{as} \quad n \to \infty, \quad \text{for} \quad t, x \geq 0. \tag{4.28}$$

By Assumption 1, $\bar{A}_n \Rightarrow \bar{a}$ with $\bar{a}$ being a deterministic and continuous function, it follows that

$$E[(\tilde{R}_n(s,x) - \hat{R}_n(s,x))^2] \to 0 \quad \text{as} \quad n \to \infty, \quad for \quad s, x \geq 0, \tag{4.29}$$

and thus,

$$E[(\tilde{X}^e_{n,2}(t,y) - \hat{X}^e_{n,2}(t,y))^2] \to 0 \quad \text{as} \quad n \to \infty, \quad for \quad t, y \geq 0. \tag{4.30}$$

Therefore, it suffices to prove the tightness of the processes $\{\tilde{X}^e_{n,2} : n \geq 1\}$ in D_D.

We observe that the processes $\tilde{X}^e_{n,2}$ in (4.24) can be written as

$$\tilde{X}^e_{n,2}(t,y) = \frac{1}{\sqrt{n}} \sum_{i=A_n(t-y)}^{A_n(t)} \tilde{\gamma}_i(t - \tau^n_i). \tag{4.31}$$

We will apply Theorem 6.2 in Pang and Whitt (2010) to prove the tightness property of $\{\tilde{X}^e_{n,2} : n \geq 1\}$. First, we show the stochastic boundedness of $\tilde{X}^e_{n,2}$. It suffices to show the stochastic boundedness of

$$\check{X}^e_{n,2}(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{A_n(t)} \tilde{\gamma}_i(t - \tau^n_i) \tag{4.32}$$

since for each t and y, $\tilde{X}^e_{n,2}(t,y) \le \check{X}^e_{n,2}(t)$. We will show that for any $T > 0$,

$$\lim_{L\to\infty} \lim_{n\to\infty} P\left(\sup_{t\le T} |\check{X}^e_{n,2}(t)| > L\right) = 0. \tag{4.33}$$

For any constant $\check{L} > 0$, we can write

$$P\left(\sup_{t\le T} |\check{X}^e_{n,2}(t)| > L\right) \le P(\bar{A}_n(T+1) > \check{L}) + P\left(\sup_{t\le T} |\tilde{K}_n(\bar{A}_n(t)\wedge \check{L}, t-\tau^n_i)| > L\right) \tag{4.34}$$

where $\tilde{K}_n(t,x)$ is defined in (4.22). By Assumption 1, the sequence of processes $\{\bar{A}_n : n \ge 1\}$ is tight, and thus

$$\lim_{\check{L}\to\infty} \limsup_{n\to\infty} P(\bar{A}_n(T+1) > \check{L}) = 0. \tag{4.35}$$

Since $\{\tilde{\gamma}_k(x) : k \ge 1\}$ in an ergodic martingale difference sequence for each $x \ge 0$, by the Lenglart-Rebolledo inequality (see, e.g., p.30 in Karatzas and Shreve (1991)), for any constant $\tilde{L}$

$$P\left(\sup_{t\le T} |\tilde{K}_n(\bar{A}_n(t)\wedge \check{L}, t-\tau^n_i)| > L\right) \le \tilde{L}/L + P\left(\langle \tilde{K}_n(\bar{A}_n(T)\wedge \check{L}, T-\tau^n_i)\rangle > \tilde{L}\right), \tag{4.36}$$

where

$$\langle \tilde{K}_n(\bar{A}_n(T)\wedge \check{L}, T-\tau^n_i)\rangle = \frac{1}{n} \sum_{i=1}^{\lfloor n(\bar{A}_n(T)\wedge \check{L})\rfloor} E[\tilde{\gamma}_i(T-\tau^n_i)^2], \tag{4.37}$$

and

$$\frac{1}{n} \sum_{i=1}^{\lfloor n(\bar{A}_n(t))\rfloor} E[\tilde{\gamma}_i(t-\tau^n_i)^2] \Rightarrow \int_0^t E[\tilde{\gamma}_i(t-s)^2] d\bar{a}(s) < \infty \quad \text{as} \quad n\to\infty. \tag{4.38}$$

We can choose $\tilde{L}$ large (but fixed) so that

$$\lim_{L\to\infty} \lim_{n\to\infty} P\left(\sup_{t\le T} |\tilde{K}_n(\bar{A}_n(t)\wedge \check{L}, t-\tau^n_i)| > L\right) = 0, \tag{4.39}$$

and thus (4.33) is proved.

We next show that for any $\varsigma > 0$

$$\lim_{\vartheta\to 0} \limsup_{n\to\infty} \sup_{\kappa_n} P\left(\sup_{t\le\vartheta} d_{J_1}(\tilde{X}_{n,2}(\kappa_n+t,\cdot), \tilde{X}_{n,2}(\kappa_n,\cdot)) > \varsigma\right) = 0, \tag{4.40}$$

where $\{\kappa_n : n \ge 1\}$ is a sequence of uniformly bounded stopping times with respect to the filtration $\mathbf{H}_n \equiv \{\mathcal{H}_n(t) : t \ge 0\}$ and with upper bound κ^*, where

$$\mathcal{H}_n(t) \equiv \sigma\{\eta_i \le s-\tau^n_i : 1 \le i \le A_n(t), 0 \le s \le t\} \vee \{A_n(s) : 0 \le s \le t\} \vee \mathcal{N} \tag{4.41}$$

and $\mathbf{H}_n$ satisfies the usual conditions. It suffices to show that for any $\varsigma > 0$

$$\lim_{\vartheta\to 0} \limsup_{n\to\infty} \sup_{\kappa_n} P\left(\sup_{t\le\vartheta} \sup_{y\in[0,T\wedge(\kappa_n+t)]} \left|\tilde{X}_{n,2}(\kappa_n+t,y) - \tilde{X}_{n,2}(\kappa_n,y)\right| > \varsigma\right) = 0. \tag{4.42}$$

For each n, κ_n, $y>0$ and $t<\vartheta$ small, by (4.31)

$$\begin{aligned}
&\tilde{X}_{n,2}(\kappa_n+t,y)-\tilde{X}_{n,2}(\kappa_n,y)\\
=\;& \frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n+t-y)}^{A_n(\kappa_n+t)}\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)-\frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n-y)}^{A_n(\kappa_n)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\\
=\;& \frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n+t-y)}^{A_n(\kappa_n+t)}\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)-\frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n+t-y)}^{A_n(\kappa_n)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\\
&-\frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n-y)}^{A_n(\kappa_n+t-y)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\\
=\;& \frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n)+1}^{A_n(\kappa_n+t)}\tilde{\gamma}_i(\kappa_n+t-\tau_{i,1}^n)-\frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n+t-y)}^{A_n(\kappa_n)}\left[\tilde{\gamma}_i(\kappa_n-\tau_i^n)-\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)\right]\\
&-\frac{1}{\sqrt{n}}\sum_{i=A_n(\kappa_n-y)}^{A_n(\kappa_n+t-y)}\tilde{\gamma}_i(\kappa_n-\tau_i^n). \qquad (4.43)
\end{aligned}$$

Then, for any $L>0$, we have

$$\begin{aligned}
&P\left(\sup_{t\le\vartheta}\sup_{y\in[0,T\wedge(\kappa_n+t)]}\left|\tilde{X}_{n,2}(\kappa_n+t,y)-\tilde{X}_{n,2}(\kappa_n,y)\right|>\varsigma\right)\\
\le\;& P\left(\bar{A}_n(\kappa^*+1)>L\right)\\
&+P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)\right|>\varsigma\right)\\
&+P\left(\sup_{t\le\vartheta}\sup_{y\in[0,T\wedge(\kappa_n+t)]}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n+t-y)\wedge L)}^{n(\bar{A}_n(\kappa_n)\wedge L)}\left[\tilde{\gamma}_i(\kappa_n-\tau_i^n)-\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)\right]\right|>\varsigma\right)\\
&+P\left(\sup_{t\le\vartheta}\sup_{y\in[0,T\wedge(\kappa_n+t)]}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_{n,1}(\kappa_n-y)\wedge L)}^{n(\bar{A}_n(\kappa_n+t-y)\wedge L)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\right|>\varsigma\right). \qquad (4.44)
\end{aligned}$$

By Assumption 1, we have

$$\lim_{L\to\infty}\limsup_{n\to\infty}P\left(\bar{A}_n(\kappa^*+1)>L\right)=0. \qquad (4.45)$$

For the second term on the right hand side of (4.44),

$$\begin{aligned}
&P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)\right|>\varsigma\right)\\
\le\;& P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\right|>\varsigma\right)
\end{aligned}$$

$$+P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}[\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)-\tilde{\gamma}_i(\kappa_n-\tau_i^n)]\right|>\varsigma\right). \qquad (4.46)$$

For each n and $\tilde{\varsigma}>0$, by Lenglart-Rebolledo inequality, the first term on the right hand side of (4.46) satisfies

$$\begin{aligned} & P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\right|>\varsigma\right) \\ \le\ & \frac{\tilde{\varsigma}}{\varsigma}+P\left(\left\langle\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+\vartheta)\wedge L)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\right\rangle>\tilde{\varsigma}\right) \\ =\ & \frac{\tilde{\varsigma}}{\varsigma}+P\left(\frac{1}{n}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+\vartheta)\wedge L)}E[(\tilde{\gamma}_i(\kappa_n-\tau_i^n))^2]>\tilde{\varsigma}\right) \qquad (4.47)\end{aligned}$$

where

$$\frac{1}{n}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+\vartheta)\wedge L)}E[(\tilde{\gamma}_i(\kappa_n-\tau_i^n))^2]\le\sup_{s,t\le\Upsilon\wedge T,|s-t|<\vartheta}\frac{1}{n}\sum_{i=n(\bar{A}_n(s)\wedge L)+1}^{n(\bar{A}_n(t)\wedge L)}E[(\tilde{\gamma}_i(t-\tau_i^n))^2] \qquad (4.48)$$

and thus, by (4.38) and choosing $\tilde{\varsigma}$ arbitrarily small, we have that for any $\varsigma>0$,

$$\lim_{\vartheta\to0}\limsup_{n\to\infty}\sup_{\kappa_n}P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}\tilde{\gamma}_i(\kappa_n-\tau_i^n)\right|>\varsigma\right)=0. \qquad (4.49)$$

Since for each n and i, $\{\tilde{\gamma}_i(x):x\ge0\}$ is a square integrable martingale with respect to the filtration $\mathbf{G}\equiv\{\mathcal{G}(t):t\ge0\}$ where $\mathcal{G}(t)=\sigma\{\mathbf{1}(\eta_i\le x):0\le x\le t,i=1,2,...\}$, then by Doob's maximal inequality, for any $c>0$,

$$\begin{aligned} & P\left(\sup_{t\le\vartheta}|\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)-\tilde{\gamma}_i(\kappa_n-\tau_i^n)|>c\right] \\ \le\ & c^{-2}E\left[(\tilde{\gamma}_i(\kappa_n+\vartheta-\tau_i^n)-\tilde{\gamma}_i(\kappa_n-\tau_i^n))^2\right]\to0\quad\text{as}\quad\vartheta\to0. \qquad (4.50)\end{aligned}$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^{\lfloor n(\bar{A}_n(t))\rfloor}E[(\tilde{\gamma}_i(t+\vartheta-\tau_i^n)-\tilde{\gamma}_i(t-\tau_i^n))^2]\Rightarrow\int_0^tE[(\tilde{\gamma}_i(t+\vartheta-s)-\tilde{\gamma}_i(t-s))^2]d\bar{a}(s)<\infty. \quad (4.51)$$

as $n\to\infty$ and the limit in (4.51) goes to 0 as $\vartheta\to0$. Thus, it follows that for each $\varsigma>0$,

$$\lim_{\vartheta\to0}\limsup_{n\to\infty}\sup_{\kappa_n}P\left(\sup_{t\le\vartheta}\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(\kappa_n)\wedge L)+1}^{n(\bar{A}_n(\kappa_n+t)\wedge L)}[\tilde{\gamma}_i(\kappa_n+t-\tau_i^n)-\tilde{\gamma}_i(\kappa_n-\tau_i^n)]\right|>\varsigma\right)=0 \qquad (4.52)$$

For the third term in (4.44), a similar argument applies by observing that for any $y \in [0, T \wedge (\kappa_n + t)]$,

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=n(\bar{A}_n(\kappa_n+t-y)\wedge L)}^{n(\bar{A}_n(\kappa_n)\wedge L)} \left| \tilde{\gamma}_i(\kappa_n - \tau_i^n) - \tilde{\gamma}_i(\kappa_n + t - \tau_i^n) \right| \\
\leq \;& \frac{1}{\sqrt{n}} \sum_{i=0}^{n(\bar{A}_n(\kappa_n)\wedge L)} \left| \tilde{\gamma}_i(\kappa_n - \tau_i^n) - \tilde{\gamma}_i(\kappa_n + t - \tau_i^n) \right|. \qquad (4.53)
\end{aligned}
$$

The last term in (4.44) follows from the same argument as in the first term in (4.46). Thus, (4.42) is proven, and tightness of the processes $\{\tilde{X}^e_{n,2} : n \geq 1\}$ is proven in the space D_D, which implies the tightness of $\{\hat{X}^e_{n,2} : n \geq 1\}$.

By the tightness of $\{\hat{X}^e_{n,1} : n \geq 1\}$ and $\{\hat{X}^e_{n,2} : n \geq 1\}$, we obtain the tightness of $\{\hat{Q}^e_n : n \geq 1\}$ in D_D, which implies tightness of $\{\hat{D}^e_n : n \geq 1\}$. Finally, the joint tightness of all these processes in the product space follows from tightness of each sequence of processes in their own space (Theorem 11.6.7, Whitt (2002)). This completes the proof. ■

Lemma 4.2 (*Convergence of finite dimensional distributions*) *Under the assumptions of Theorem 3.2, the finite dimensional distributions of the processes* $(\hat{A}_n, \hat{X}^e_{n,1}, \hat{X}^e_{n,2}, \hat{D}_{n,1})$ *converge in distribution to those of the processes* $(\hat{A}, \hat{X}^e_1, \hat{X}^e_2, \hat{D})$.

Proof. As in the proof of tightness, we mainly focus on the proof for the convergence of the f.d.d.'s of the processes $\hat{X}^e_{n,2}$ to those of $\hat{X}^e_2$.

First, we write the processes $\hat{X}^e_{n,2}$ defined in (2.17) as the limit of mean square integrals, as in (3.13) for $\hat{X}^e_2$,

$$\hat{X}^e_{n,2}(t, y) \equiv \text{l.i.m.}_{k\to\infty} \hat{X}^e_{n,2,k}(t, y), \qquad (4.54)$$

where

$$
\begin{aligned}
\hat{X}^e_{n,2,k}(t, y) &\equiv \int_{t-y}^{t} \int_0^{\infty} \mathbf{1}_{k,t,y}(s, x) d\hat{R}_n(s, x) = \sum_{i=1}^{k} \Delta_{\hat{R}_n}(s^k_{i-1}, s^k_i, t - s^k_i, t) \\
&= \sum_{i=1}^{k} \Delta_{\hat{K}_n}(\bar{A}_n(s^k_{i-1}), \bar{A}_n(s^k_i), t - s^k_i, t) \qquad (4.55)
\end{aligned}
$$

with $\mathbf{1}_{k,t,y}(s, x)$ defined in (3.14) for $t - y = s^k_0 < s^k_1 < \cdots < s^k_k = t$ and $\max_{1\leq i\leq k} |s^k_i - s^k_{i-1}| \to 0$ as $k \to \infty$, and

$$\Delta_{\hat{R}_n}(s^k_{i-1}, s^k_i, t - s^k_i, t) = \hat{R}_n(s^k_i, t) - \hat{R}_n(s^k_{i-1}, t) - \hat{R}_n(s^k_i, t - s^k_i) + \hat{R}_n(s^k_{i-1}, t - s^k_i). \qquad (4.56)$$

Similarly, for $\hat{X}^e_2$, we write them as limits of mean square integrals of $\hat{X}^e_{2,k}$,

$$\hat{X}^e_{2,k}(t,y) \equiv \int_{t-y}^{t}\int_0^\infty \mathbf{1}_{k,t,y}(s,x)d\hat{K}(\bar{a}(t),x) = \sum_{i=1}^{k}\Delta_{\hat{K}}(\bar{a}(s^k_{i-1}),\bar{a}(s^k_i),t-s^k_i,t). \tag{4.57}$$

We prove the convergence of f.d.d.'s of $\hat{X}^e_{n,2}$ to those of $\hat{X}^e_2$ by using the convergence of $\hat{K}_n \Rightarrow \hat{K}$ in D^2_D in (2.11). Define the processes $\check{X}^e_{n,2,k}$ by

$$\check{X}^e_{n,2,k}(t,y) \equiv \sum_{i=1}^{k}\Delta_{\hat{K}_n}(\bar{a}(s^k_{i-1}),\bar{a}(s^k_i),t-s^k_i,t). \tag{4.58}$$

Then, by the convergence of $\hat{K}_n \Rightarrow \hat{K}$ in D^2_D and the continuity of $\bar{a}$, we can conclude that the joint convergence of f.d.d.'s of $(\hat{A}_n, \hat{X}^e_{n,1}, \check{X}^e_{n,2,k})$ converge in distribution to those of the processes $(\hat{A}, \hat{X}^e_1, \hat{X}^e_{2,k})$ as $n \to \infty$.

Now it suffices to show that the difference between $\check{X}^e_{n,2,k}$ and $\hat{X}^e_{n,2,k}$ is asymptotically negligible in probability as $n \to \infty$ for each k, and the difference between $\hat{X}^e_{n,2,k}$ and $\hat{X}^e_{n,2}$ is asymptotically negligible in probability as $n \to \infty$ and $k \to \infty$. We will next show that for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\sup_{0\le t\le T, 0\le y\le t} |\check{X}^e_{n,2,k}(t,y) - \hat{X}^e_{n,2,k}(t,y)| > \epsilon\right) = 0, \quad T > 0, \tag{4.59}$$

and

$$\lim_{k\to\infty}\limsup_{n\to\infty} P(|\hat{X}^e_{n,2,k}(t,y) - \hat{X}^e_{n,2}(t,y)| > \epsilon) = 0, \quad t \ge 0, \quad 0 \le y \le t. \tag{4.60}$$

We obtain (4.59) from the convergence of $\bar{A}_n \Rightarrow \bar{a}$ in (2.1), the continuity of $\bar{a}$, and the convergence $\hat{K}_n \Rightarrow \hat{K}$ in (2.11) and the continuity of the generalized Kiefer limit process $\hat{K}$. It remains to show (4.60). For that, we define the processes $\tilde{X}^e_{n,2,k}$ for each k and n by

$$\begin{aligned}\tilde{X}^e_{n,2,k}(t,y) &\equiv \int_{t-y}^{t}\int_0^\infty \mathbf{1}_{k,t,y}(s,x)d\tilde{R}_n(s,x) = \sum_{i=1}^{k}\Delta_{\tilde{R}_n}(s^k_{i-1},s^k_i,t-s^k_i,t)\\ &= \sum_{i=1}^{k}\Delta_{\tilde{K}_n}(\bar{A}_n(s^k_{i-1}),\bar{A}_n(s^k_i),t-s^k_i,t)\end{aligned} \tag{4.61}$$

where $\tilde{K}_n$ and $\tilde{R}_n$ are defined in (4.22) and (4.23), respectively, and the partition of interval $[t-y,t]$ and $\mathbf{1}_{k,t,y}(s,x)$ are the same as in (4.55)-(4.56). (4.28) and (4.29) imply that the processes $\tilde{X}^e_{n,2,k}$ and $\hat{X}^e_{n,2,k}$ are asymptotically negligible as $n \to \infty$ for each k, and moreover, (4.30) implies that $\tilde{X}^e_{n,2}$ in (4.24) and $\hat{X}^e_{n,2}$ are asymptotically negligible as $n \to \infty$. Thus, it suffices to show the following in order to prove (4.60),

$$\lim_{k\to\infty}\limsup_{n\to\infty} P(|\tilde{X}^e_{n,2,k}(t,y) - \tilde{X}^e_{n,2}(t,y)| > \epsilon) = 0, \quad t \ge 0, \quad 0 \le y \le t, \quad \epsilon > 0. \tag{4.62}$$

By (4.61) and (4.24), we have

$$\begin{aligned}\tilde{X}^e_{n,2,k}(t,y) - \tilde{X}^e_{n,2}(t,y) &= \int_{t-y}^{t}\int_0^\infty [\mathbf{1}_{k,t,y}(s,x) - \mathbf{1}(s+x>t)]d\tilde{R}_n(s,x) \\ &= \frac{1}{\sqrt{n}}\sum_{i=A_n(t-y)}^{A_n(t)} \tilde{\beta}^k_{i,1}(\tau^n_i,\eta_i)(t,y), \qquad (4.63)\end{aligned}$$

where $\tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y)$ is defined by

$$\tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y) = \sum_{j=1}^k \mathbf{1}(s^k_{j-1} < \tau^n_i \le s^k_j)\check{\beta}^k_i(\tau^n_i,\eta_i), \qquad (4.64)$$

$$\check{\beta}^k_i(\tau^n_i,\eta_i) = \check{\gamma}_i(\tau^n_i,\eta_i) + \hat{\check{\gamma}}_i(\tau^n_i,\eta_i) - \hat{\check{\gamma}}_{i-1}(\tau^n_{i-1},\eta_{i-1}), \qquad (4.65)$$

$$\check{\gamma}_i(\tau^n_i,\eta_i) \equiv \mathbf{1}(t-s^k_j < \eta_i \le t-\tau^n_i) - (F(t-\tau^n_i) - F(t-s^k_j)), \qquad (4.66)$$

and

$$\hat{\check{\gamma}}_i(\tau^n_i,\eta_i) \equiv \sum_{m=1}^\infty E[\check{\gamma}_{i+m}(\tau^n_{i+m},\eta_{i+m})|\mathcal{F}_i]. \qquad (4.67)$$

It is clear that by construction, for each i,n,t,y,k, the sequence $\{\check{\beta}^k_i(\tau^n_i,\eta_i) : i \ge 1\}$ is a martingale difference sequence, and so is the sequence $\{\tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y) : i \ge 1\}$. Moreover, $E[\tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y)] = E[\check{\beta}^k_i(\tau^n_i,\eta_i)] = 0$ and $E[(\tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y))^2] < \infty$, and

$$\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}(\hat{\check{\gamma}}_i(\tau^n_i,\eta_i) - \hat{\check{\gamma}}_{i-1}(\tau^n_{i-1},\eta_{i-1}))\right\|_{L^2} \to 0 \quad \text{as} \quad n\to\infty. \qquad (4.68)$$

Then, we have that for any $L>0$ and $\epsilon>0$,

$$\begin{aligned}
& P(|\tilde{X}^e_{n,2,k}(t,y) - \tilde{X}^e_{n,2}(t,y)| > \epsilon) \\
\le\;& P(\bar{A}_n(t) > L) + P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)} \tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y)\right| > \epsilon\right) \\
\le\;& P(\bar{A}_n(t) > L) + \frac{1}{\epsilon^2}E\left[\left\langle \frac{1}{\sqrt{n}}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)} \tilde{\beta}^k_i(\tau^n_i,\eta_i)(t,y)\right\rangle\right] \\
\le\;& P(\bar{A}_n(t) > L) \\
& + \frac{1}{\epsilon^2}E\left[\frac{1}{n}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)}\sum_{j=1}^k \mathbf{1}(s^k_{j-1} < \tau^n_i \le s^k_j)E\big[\big(\check{\gamma}_i(\tau^n_i,\eta_i)\big)^2\big]\right] \qquad (4.69) \\
& + \frac{1}{\epsilon^2}E\left[\left(\frac{1}{n}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)}\sum_{j=1}^k \mathbf{1}(s^k_{j-1} < \tau^n_i \le s^k_j)(\hat{\check{\gamma}}_i(\tau^n_i,\eta_i) - \hat{\check{\gamma}}_{i-1}(\tau^n_{i-1},\eta_{i-1}))\right)^2\right].
\end{aligned}$$

By Assumption 1, for the first term in (4.69), we have, for each $t \geq 0$,

$$\lim_{L\to\infty} \limsup_{n\to\infty} P(\bar{A}_n(t) > L) = 0. \tag{4.70}$$

For the second term on the right side of (4.69),

$$\begin{aligned} E\big[\big(\check{\gamma}_i(\tau_i^n, \eta_i)\big)^2\big] &= (F(t-\tau_i^n) - F(t-s_j^k))[1 - (F(t-\tau_{i,1}^n) - F(t-s_j^k))] \\ &\leq F(t-\tau_i^n) - F(t-s_j^k), \end{aligned} \tag{4.71}$$

which implies that

$$\begin{aligned} & E\left[\frac{1}{n}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)} \sum_{j=1}^{k} \mathbf{1}(s_{j-1}^k < \tau_i^n \leq s_j^k) E\big[\big(\check{\gamma}_i(\tau_i^n, \eta_i)\big)^2\big]\right] \\ \leq\ & E\left[\frac{1}{n}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)} \sum_{j=1}^{k} \mathbf{1}(s_{j-1}^k < \tau_i^n \leq s_j^k)(F(t-\tau_i^n) - F(t-s_j^k))\right] \\ \leq\ & E\left[\frac{1}{n}\sum_{j=1}^{k}(F(t-s_{j-1}^k) - F(t-s_j^k))(n(\bar{A}_n(s_j)\wedge L) - n(\bar{A}_n(s_{j-1})\wedge L))\right] \\ \leq\ & E\left[\max_{1\leq j\leq k}((\bar{A}_n(s_j)\wedge L) - (\bar{A}_n(s_{j-1})\wedge L))\right]. \end{aligned} \tag{4.72}$$

Thus, by Assumption 1, the continuity of $\bar{a}$ and (4.70), we have

$$\begin{aligned} & \lim_{k\to\infty}\limsup_{n\to\infty} E\left[\frac{1}{n}\sum_{i=n(\bar{A}_n(t-y)\wedge L)}^{n(\bar{A}_n(t)\wedge L)} \sum_{j=1}^{k} \mathbf{1}(s_{j-1}^k < \tau_i^n \leq s_j^k) E\big[\big(\check{\gamma}_{i,1}(\tau_i^n, \eta_i)\big)^2\big]\right] \\ \leq\ & \lim_{k\to\infty}\limsup_{n\to\infty} E\left[\max_{1\leq j\leq k}((\bar{A}_n(s_j)\wedge L) - (\bar{A}_n(s_{j-1})\wedge L))\right] = 0. \end{aligned} \tag{4.73}$$

For the last term on the right side of (4.69), we apply (4.68). Thus, (4.62) is proven and so is (4.60). ■

5 Conclusion

In this paper we have studied the $G_t/G^D/\infty$ infinite-server model with arrival process satisfying a FCLT (Assumption 1), allowing time-varying arrival rates and correlated interarrival times, and dependent service times that form a stationary sequence satisfying either the ϕ-mixing condition or the S-mixing condition (Assumption 2). We have established a FWLLN and a FCLT for the process $Q^e(t, y)$, which represents the number of customers in the system at time t with elapsed service time less than or equal to y, together with the departure process. We have shown that the dependence among service times does not affect the fluid limits, but does affect the limit process in the FCLT. As in Pang and Whitt (2010), for the service times a prominent role is played by the generalized Kiefer process. We have characterized the Gaussian property of these processes and shown that the variance formulas have an additional term to indicate the impact of dependence among service times. However, this additional term is quite complicated. Consequently, we have further studied the formula obtained here in order to better understand the engineering impact of dependence among the service times in Pang and Whitt (2011a, 2011b). There it is shown how the dependence might be modeled and how the performance impact of the dependence can be calculated.

There remain many open problems to investigate. It remains to study the case in which the interarrival times and service times are correlated with each other. It also remains to establish corresponding many-server heavy-traffic limits for queueing models with only finitely many servers. Finally, it remains to investigate networks of IS and many-server queues with dependence structure.

References

[1] Berkes, I., Philipp, W.: An almost sure invariance principle for empirical distribution function of mixing random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete.* **41**, 115–137. (1977)

[2] Berkes, I., Hörmann, S., Schauer, J.: Asymptotic results for the empirical process of stationary sequences. *Stochastic processes and their applications.* **119**, 1298–1324 (2009)

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

[4] Billingsley, P.: *Convergence of Probability Measures.* second ed, Wiley, New York (1999)

[5] de Véricourt, F., Zhou, Y-P: Managing response time in a call-routing problem with service failure. *Operations Research.* Vol. 53, No.6, 968–981 (2005)

[6] Green, L. V., Kolesar, P. J., Whitt, W.: Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management.* **16** 13–39 (2007)

[7] Hall, P., Heyde, C.C.: *Martingale Limit Theory and its Applications.* Academic Press, Inc. (1980)

[8] Jacobs, P. A. : A cyclic queueing network with dependent exponential service times. *J. Appl. Prob.* **15**, 573–589 (1978)

[9] Jacobs, P. A. : Heavy traffic results for single-server queues with dependent (EARMA) service and interarrival times. *Adv. Appl. Prob.* **12**, 517–529 (1980)

[10] Karatzas, I., Shreve, S.E.: *Brownian Motion and Stochastic Calculus.* second ed., Springer, Berlin (1991)

[11] Khoshnevisan D. *Multiparameter Processes: An Introduction to Random Fields*. Springer. (2002)

[12] Krichagina, E. V., Puhalskii, A. A.: A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. *Queueing Systems* 25, 235–280 (1997)

[13] Liu, L., Templeton, J.G.C.: Autocorrelations in infinite server batch arrival queues. *Queueing Systems.* **14**, 313–337 (1993)

[14] Liu, L., Whitt, W.: Many-server heavy-traffic limits for queues with time-varying parameters. Working paper. Columbia University, New York. (2011) Available at: http://www.columbia.edu/∼ww2040/allpapers.html

[15] Pang, G., Whitt, W.: Service Interruptions in Large-Scale Service Systems. *Management Science*, **55(9)**, 1499–1512 (2009)

[16] Pang, G., Whitt, W.: Two-parameter heavy-traffic limits for infinite-server queues. *Queueing Systems.* **65** 325–364 (2010)

[17] Pang, G., Whitt, W.: The impact of dependent service times on large-scale service systems. working paper. (2011a)

[18] Pang, G., Whitt, W.: Infinite-server queues with batch arrivals and dependent service times. To appear in *Probability in Engineering and Informational Sciences.* (2011b)

[19] Straf, M.L. Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 2, 187–221 (1971)

[20] van Der Vaart, A. W., Wellner, J.: *Weak Convergence and Empirical Processes*, Springer (1996)

[21] Whitt, W.: Comparing batch delays and customer delays. *The Bell System Technical Journal,* Vol. 62, No. 7, 2001–2009 (1983)

[22] Whitt, W.: *Stochastic-Process Limits.* Springer, New York (2002)