MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Gaussian Limits for A Fork-Join Network with Non-Exchangeable Synchronization in Heavy Traffic

Hongyuan Lu and Guodong Pang
The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University, University Park, PA 16802
hzl142, gup3@psu.edu

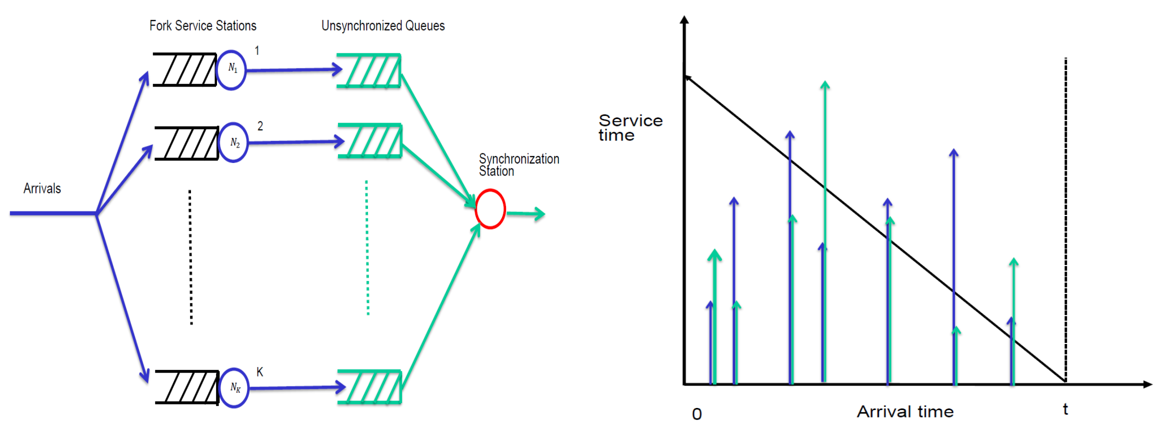
We study a fork-join network of stations with multiple servers and non-exchangeable synchronization in heavy traffic under the FCFS discipline. Tasks are only synchronized if all the tasks associated with the same job are completed. Service times of parallel tasks of each job can be correlated. We consider the number of tasks in each waiting buffer for synchronization, jointly with the number of tasks in each parallel service station and the number of synchronized jobs. We develop a new approach to show a functional central limit theorem for these processes in the quality-driven regime, under general assumptions on the arrival and service processes. Specifically, we represent these processes as functionals of a sequential empirical process driven by the sequence of service vectors for each job's parallel tasks. All the limiting processes are functionals of two independent processes - the limiting arrival process and a generalized Kiefer process driven by the service vector of each job. We characterize the transient and stationary distributions of the limiting processes.

Key words: fork-join networks; non-exchangeable synchronization; many-server heavy-traffic regime; Gaussian limits; generalized Kiefer process; multiparameter sequential empirical process; multiparameter martingale
MSC2000 subject classification: Primary: 60K25; 60F17; 90B22; 60J75; secondary: 60G44; 54A20
OR/MS subject classification: queues; networks; limit theorems; approximations

1. Introduction Fork-join networks consist of a set of service stations that serve job requests simultaneously and sequentially according to pre-designated deterministic precedence constraints. Such networks have many applications in manufacturing and telecommunications [5, 7, 8, 24, 35, 36, 56, 55, 66, 67, 49, 50, 62, 41], patient flow analysis in healthcare [31, 3, 4, 71, 72], parallel computing [60, 65, 64, 68, 42], military deployment operations [34, 70, 1], and law enforcement systems [38]. Two types of synchronization constraints are of particular interest. One is called *exchangeable synchronization* (ES) in which tasks are not tagged with a particular job and can be synchronized for a service completion once the necessary tasks are completed. This type of synchronization constraint is often used in manufacturing systems; for example, in many assembly systems, different parts of a product are processed at separate workstations or plant locations and a product will be assembled once all of its necessary parts are completed. In this case, the parts are not tagged with a particular product, since they are standardized for the same type of product. The second type is called *non-exchangeable synchronization* (NES). Tasks are tagged with a particular job and can only be synchronized when all the parallel tasks of the same job are completed.

Here we focus on the second type of synchronization constraint. We are primarily motivated to study fork-join networks with NES from patient flow analysis in hospitals [3, 4, 31, 71, 72]. For example, as a prerequisite for a doctor examination, all the test results for the same patient must be ready, and they cannot be mixed among different patients. Mixing one patient's blood test result with another patient's cardiology result can lead to severe medical consequences. Fork-join networks with NES also have applications in parallel computing. For example, a computation job can be split into several tasks processed in parallel (possibly only at some stages) and joined subsequently once these parallel tasks tagged with the same job are completed. Another example is MapReduce scheduling [16, 64, 68, 42]. In the map phase, jobs are split into parallel tasks and these tasks tagged with the same job are synchronized to be processed in the reduce phase (with additional complicated interactions between those two phases). NES also occurs in some component procurement problems that do not allow mixing between sets of component orders in certain assembly systems [24].

When there is a single server in each of the parallel service station and the service discipline is first-come-first-served (FCFS), the service completion order is preserved to be the same as the arrival order of tasks in each service station, so that the two types of synchronization constraints are equivalent. However, the arrival order of tasks in each service station can be *resequenced* at the service completion epochs when the number of servers in a service station is larger than one, or when the service discipline is not FCFS. Resequencing has been one of the most difficult obstacles in the study of fork-join networks. Some limited work has been dedicated to the study of such challenging problems. For example, substantial efforts were dedicated to the study of the max-plus recursions [30, 6, 18]. More recently, Atar et al. [4] have studied a fork-join network with single-server service stations where tasks may reenter for service at some service stations in a Bernoulli scheme so that the arrival orders of tasks at each service station are resequenced after service completion. They show that under a dynamic priority discipline, the system dynamics with NES is asymptotically equivalent to that with ES in the conventional (single-server) heavy-traffic regime. For a Markovian fork-join network with multiple servers, Zviran [72] shows that the system dynamics with NES is also asymptotically equivalent to that with ES in the conventional heavy-traffic regime. However, the two types of synchronization constraints lead to very different system dynamics when the service stations have many parallel servers in the Halfin-Whitt regime, as conjectured in [4, 72]. To the best of our knowledge, our work is the first to tackle the resequencing problem in non-Markovian fork-join networks with NES and multi-server service stations in the many-server heavy-traffic regimes.



(a) (b)

FIGURE 1. A fundamental fork-join network and a graphical representation of its system dynamics

We consider a fundamental fork-join network model with a single class of jobs and NES. As depicted in Figure 1(a), each arriving job is forked into a fixed number of parallel tasks upon arrival and each of the tasks is processed in parallel at a dedicated service station with multiple servers under the FCFS and non-idling service discipline. The parallel tasks of each job have correlated service times. Upon service completion, each task will join a buffer associated with its service station, called "unsynchronized queue", and wait for synchronization, such that each job is synchronized only if all of its tasks have been completed. In this model, in addition to the service dynamics, we are interested in the waiting buffer dynamics for synchronization. One important performance measure is the response time of a job, namely, the time from arrival to synchronization. The response time may also include the time required for the synchronization process, but we do not consider that in this work. The response time includes two delays, waiting time for service and waiting time for synchronization. Since each service station can be regarded as a separate many-server queue, the waiting time for service has been well understood [45, 63]. However, the waiting time for synchronization, which is our focus in this paper, has not been studied. Specifically, we investigate the waiting buffer dynamics for synchronization jointly with the service dynamics. In this work, we start with the situation when all the service stations operate in the quality-driven (QD) many-server regime. Asymptotically, this is equivalent to a model which has infinite numbers of servers at all service stations.

To describe the system dynamics, we can start with a graphical representation as shown in Figure 1(b) for a system of two parallel tasks. At each job's arrival epoch, we mark the arrival time on the horizontal line (x-axis) and the service times of two parallel tasks on the vertical line (y-axis). At each time t, by drawing a negative forty-five degree line, we can count the numbers of tasks in each service station and each waiting buffer for synchronization. If both parallel tasks of a job are above the line, both tasks are in service at time t; if both are below the line, the job has been synchronized and left the system; and if one task of a job is above the line and the other is below, the one above is still in service and the one below is already in the buffer waiting for synchronization. When the arrival process is Poisson, we can apply Poisson random measure theory, similarly as in the "physics" of $M/GI/\infty$ queues [22]. It can be shown that at each time t, the numbers of tasks in each service station and each waiting buffer for synchronization all have Poisson distributions and their covariances can also be obtained; see Proposition 2.1. However, when the arrival process is more general, this Poisson random measure approach does not work, and we cannot obtain the exact distributions for these performance measures. Thus, we consider heavy-traffic approximations of the system dynamics when the arrival rate gets large. For that, the graphical representation in Figure 1(b) also plays an important role; see the system's dynamic equations in §2.

It may appear that the waiting buffer dynamics for synchronization can be obtained by using previous work on infinite-server queues [37], since the number of tasks in each waiting buffer for synchronization is equal to the number of service completions in that service station minus the number of synchronized jobs. The number of synchronized jobs is equal to the number of service completions in an infinite-server queue with service times having the same distribution as the maximum of the service vector of all parallel tasks. For $G/GI/\infty$ queues, when the arrival process satisfies a functional central limit theorem (FCLT) with a Brownian motion limit, the limiting process of the number of jobs in the system (as well as the number of service completions) in the diffusion scale can be represented as a sum of two independent terms, one as an Ito's integral of the arrival Brownian motion limit and the other as a functional of a standard Kiefer process driven by service times [37]. For the fork-join networks with NES, by applying the results for $G/GI/\infty$ queues, the number of synchronized jobs also has a diffusion limit process represented as a sum of two independent terms, one as an Ito's integral of the arrival Brownian motion limit and the other as a functional of a standard Kiefer process driven by the maximum of the service vector of all parallel tasks. However, the two Kiefer processes driven by the service time at a station and the maximum of the service vector are correlated, and such correlated Kiefer processes have not been studied in the literature. It is very difficult to characterize the performance measures, in particular, the covariances between the number of tasks in service at a station and that in a waiting buffer for synchronization.

Here we develop a new approach to describe the system dynamics. The service dynamics, the waiting buffer dynamics for synchronization and the process counting synchronized jobs are all represented as functionals of the multiparameter sequential empirical process driven by the service vector of all parallel tasks. Their diffusion-scaled processes will converge weakly to limit processes that can be all represented as functionals of two independent processes - the limiting arrival process and the corresponding generalized multiparameter Kiefer process driven by the service vector (Theorem 3.3). When the limiting arrival process is Brownian motion, we show that the aforementioned limiting processes are a multidimensional continuous Gaussian process, and thus characterize the joint transient and stationary distributions of these processes (Theorem 3.4). We also study the impact of the correlation among the service vector upon these distributions; see Corollary 3.1. A numerical example is given to show the effectiveness of our approximations in §3.3.2. We characterize the difference between the mean waiting buffer sizes for synchronization in the two fork-join models with ES and NES constraints; see Proposition 3.1.

There are several advantages with this new approach. It gives a clean and elegant representation of the limiting processes, involving only two independent stochastic processes arising from the arrival and service processes. Moreover, the characterization of the limiting processes as Gaussian and their transient and stationary distributional properties can be easily obtained. We believe that this new approach launches a new framework to study more general fork-join networks, for example, the same model in Figure 1(a) when all service stations operate in the Halfin-Whitt regime and when the service vectors for parallel service times form a stationary and weakly dependent sequence satisfying strong mixing conditions, and multiclass models with multiple processing stages.

In the development of this approach, we make a fundamental contribution to the study of sequential empirical processes driven by random vectors. Sequential empirical processes driven by a sequence of random vectors and their limits as generalized Kiefer processes have been studied in the statistics literature; see e.g., [54, 11, 13, 14, 17, 19, 20], but the convergence is proved in the space $\mathbb{D}([0,T]^k,\mathbb{R})$ of real-valued càdlàg functions defined on $[0,T]^k$, $k \geq 2$ and $T \geq 0$, endowed with the generalized Skorohod $\mathcal{J}_1$ topology in [48] and [61]. In our setting, it is necessary to prove the convergence in the space $\mathbb{D}([0,T],\mathbb{D}([0,T]^k,\mathbb{R}))$ of function-valued càdlàg functions defined on $[0,T]$, endowed with the Skorohod $\mathcal{J}_1$ topology for $\mathbb{D}([0,T]^k,\mathbb{R})$-valued càdlàg functions, for $T \geq 0$.

Literature review. Most of the literature on fork-join networks is on models with single-server service stations. We only give a brief summary here on relevant work in heavy traffic. These studies are in the conventional (single-server) heavy-traffic regime. In Varma's dissertation [67], the diffusion-scaled workload processes and unsynchronized queueing processes in some fork-join network models with ES are shown to converge weakly to certain multi-dimensional reflected Brownian motions. The stationary distributions of the system response time and the processes counting the number of tasks in unsynchronized queues are specified by some partial differential equations (PDEs). Nguyen [49] shows the diffusion-scaled processes counting the queue lengths at each service station of a single-class fork-join network model with ES converge to a reflected Brownian motion in a polyhedral cone of the nonnegative orthant. Nguyen [50] discusses the difficult challenges with multiclass fork-join models with ES. As we have noted above, for a fork-join network with feedback and NES, Atar et al. [4] show that a dynamic priority discipline achieves throughput optimality asymptotically in the conventional heavy-traffic regime, as a consequence of the asymptotic equivalence between NES and ES constraints.

Very little work has been done for fork-join networks with multi-server service stations. Ko and Serfozo [35] consider a fork-join network model with a single class of Poisson arrivals and K parallel service stations with multiple servers at each station and exponential service times, and obtain an approximation for the distribution of the system response time in equilibrium under the NES constraint. Dai [15] provides an exact simulation algorithm to approximate the system response time in equilibrium for the same Markovian model in [35] by using a "coupling from the past" method. Zviran [72] studies optimal control of multi-server feedforward fork-join networks with exponential service times in the conventional heavy-traffic regime and shows that FCFS is asymptotically optimal and the resequencing disruption becomes asymptotically negligible. Zaied [71] calculates mean offered-load functions of fork-join networks with NES and multiple processing stages when the arrival process is time-inhomogeneous Poisson and service times for parallel tasks are independent, and studies staffing of time-varying emergency departments and synchronization delays under Markovian assumptions. Both dissertations of Zviran [72] and Zaied [71] are motivated from applications in patient flow analysis. Gurvich and Ward [25] study optimal matching policies for a pure join model (Markovian) with multiple classes of jobs under certain matching constraints.

This work contributes to the recent development for non-Markovian many-server queueing models. We only mention those that are most relevant to our work due to the large volume of papers on many-server models. Krichagina and Puhalskii [37] first observe that the system dynamics of an infinite-server queueing model can be represented by an integral functional of a sequential empirical process driven by service times. They show that the diffusion-scaled processes counting the number of jobs in the system can be approximated by a functional of a standard Kiefer process driven by service times. Pang and Whitt [51, 53] generalize that approach to establish two-parameter process limits for $G/G/\infty$ queues when the service times are i.i.d. and weakly dependent, respectively. Reed [58] and Puhalskii and Reed [57] have observed a relationship between finite-server and infinite-server queues and generalized the approach in [37] to obtain the diffusion limits for $G/GI/N$ queues in the Halfin-Whitt regime. Mandelbaum and Momčilović [46] generalize the approach by Reed [58] to study $G/GI/N+GI$ queues with abandonment in the Halfin-Whitt regime, and Huang et al. [26] study $G/M/N+GI$ queues in an overloaded many-server regime by applying the results in [37]. All these papers use sequential empirical processes driven by a sequence of univariate random variables. Our approach to study fork-join networks with NES uses multiparameter sequential empirical processes driven by a sequence of i.i.d. random vectors and properties of multiparameter processes and martingales. This approach is further developed to study (non-Markovian) many-server fork-join networks in the Halfin-Whitt regime in [44].

1.1. Organization of the Paper The paper is organized as follows. We finish this section with a summary of notations below. In §2, a detailed description of the model and the assumptions are given. The main results are stated in §3. A new FCLT for multiparameter sequential empirical processes is given in §3.1 and proved in §4. The functional weak law of large numbers (FWLLN) and FCLT for the service and waiting dynamics for synchronization are stated in §3.2 and proved in §6. The Gaussian characterization of the limit processes is stated in §3.3 and proved in §5. We make some concluding remarks and discuss future work in §7. Some additional proofs are given in the appendix §A and §B.

1.2. Notation The following notations are used throughout the paper. $\mathbb{R}$ and $\mathbb{R}_+$ ($\mathbb{R}^d$ and $\mathbb{R}^d_+$, respectively) denote sets of real and real non-negative numbers (d-dimensional vectors, respectively, $d \geq 2$). $\mathbb{N}$ denotes the set of natural numbers. For $a, b \in \mathbb{R}$, we denote $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. For any $x \in \mathbb{R}_+$, $\lfloor x \rfloor$ is used to denote the largest integer less than or equal to x. We use bold letter to denote a vector, e.g., $\boldsymbol{x} := (x_1, ..., x_d) \in \mathbb{R}^d$. $\mathbf{0}$ and $\boldsymbol{e}$ denote the vectors whose components are all 0 and 1, respectively. For $x \in \mathbb{R}$ and $\boldsymbol{e} \in \mathbb{R}^d$, we define $x\boldsymbol{e} := (x, ..., x) \in \mathbb{R}^d$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we denote $\boldsymbol{x} \leq \boldsymbol{y}$, $\boldsymbol{x} \geq \boldsymbol{y}$ and $\boldsymbol{x} > \boldsymbol{y}$ in the componentwise sense, and let $\boldsymbol{x} \wedge \boldsymbol{y} = (x_1 \wedge y_1, ..., x_d \wedge y_d)$ and $\boldsymbol{x} \vee \boldsymbol{y} = (x_1 \vee y_1, ..., x_d \vee y_d)$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{x} \not\leq \boldsymbol{y}$ is used to denote that at lease one component of $\boldsymbol{x}$ is strictly greater than the corresponding component of $\boldsymbol{y}$. We use $\mathbf{1}(A)$ to denote the indicator function of a set A. The abbreviation *a.s.* means almost surely. For any univariate distribution function G, we denote $G^c := 1 - G$. We denote the sets $T_{\boldsymbol{u}} := \{\boldsymbol{t} \in \mathbb{R}^d_+ : t_k > u_k, k = 1, ..., d\}$ and $L_{\boldsymbol{u}} := \{\boldsymbol{t} \in \mathbb{R}^d_+ : t_k \leq u_k, k = 1, ..., d\}$ for $\boldsymbol{u} = (u_1, ..., u_d) \in \mathbb{R}^d_+$. For any set $\mathcal{A}$, we write $\mathcal{A}^c$ as its complementary set.

All random variables and processes are defined on a common probability space $(\Omega, \mathcal{F}, P)$. For any two complete separable metric spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, we denote $\mathcal{S}_1 \times \mathcal{S}_2$ as their product space, endowed with the maximum metric, i.e., the maximum of two metrics on $\mathcal{S}_1$ and $\mathcal{S}_2$. $\mathcal{S}^k$ is used to represent k-fold product metric space of any complete and separable metric space $\mathcal{S}$ for $k \in \mathbb{N}$ with the maximum metric. For a complete separable metric space $\mathcal{S}$, $\mathbb{D}([0,\infty), \mathcal{S})$ denotes the space of all $\mathcal{S}$-valued càdlàg functions on $[0,\infty)$, and is endowed with the Skorohod $\mathcal{J}_1$ topology (see, e.g., [23]). Let $\mathbb{D} \equiv \mathbb{D}([0,\infty), \mathbb{R})$. $\mathbb{D}(\mathbb{T}, \mathbb{R})$ denotes the space of all "continuous from above with limits from below" real-valued functions on $\mathbb{T} \subseteq \mathbb{R}_+^k$ for $k \geq 2$; see [61, 9, 48] for $\mathbb{T} = [0,1]^k$ and [27, 33] for $\mathbb{T} = [0,\infty)^k$. Denote $\mathbb{D}_k \equiv \mathbb{D}([0,\infty)^k, \mathbb{R})$ for $k \geq 2$. For the space $\mathbb{D}([0,1]^k, \mathbb{R})$, we endow it with the metric d_k introduced in [61]. The spaces $\mathbb{D}([0,\infty), \mathbb{D}_k)$ and $\mathbb{D}([0,\infty), \mathbb{D}([0,1]^k, \mathbb{R}))$, $k \geq 2$, are endowed with the Skorohod $\mathcal{J}_1$ topology in [23] (see §6 in Chapter 3). For any $z \in \mathbb{D}([0,\infty), \mathbb{D}_k)$, denote $\|z\|_{T,\mathcal{A}} := \sup_{0 \leq t \leq T} \sup_{\boldsymbol{x} \in \mathcal{A}} |z(t, \boldsymbol{x})|$, where $T > 0$ and $\mathcal{A}$ is a bounded closed subset of $\mathbb{R}_+^k$. For a complete separable metric space $\mathcal{S}$, $\mathbb{C}([0,\infty)^k, \mathcal{S})$ is the space of all continuous $\mathcal{S}$-valued functions on $[0,\infty)^k$ for $k \geq 1$, and denote $\mathbb{C}_k \equiv \mathbb{C}([0,\infty)^k, \mathbb{R})$ for $k \geq 2$, and $\mathbb{C} \equiv \mathbb{C}([0,\infty), \mathbb{R})$. $\mathbb{C}([0,\infty), \mathbb{C}_k)$ is the subset of continuous functions in $\mathbb{D}([0,\infty), \mathbb{D}_k)$. Let $\mathbb{D}_\uparrow$ and $\mathbb{C}_\uparrow$ be the subset of functions in $\mathbb{D}$ and $\mathbb{C}$ which are nondecreasing, respectively. Weak convergence of probability measures μ_n to μ will be denoted as $\mu_n \Rightarrow \mu$. For any two random variables $\mathcal{X}_1$ and $\mathcal{X}_2$, we denote $\mathcal{X}_1 \stackrel{d}{=} \mathcal{X}_2$ as their equality in distribution. For a sequence of processes $\{\mathcal{X}^n : n \geq 1\}$ and a process $\mathcal{X}$, $\mathcal{X}^n \xrightarrow{d_f} \mathcal{X}$ and $\mathcal{X}^n \xrightarrow{P} \mathcal{X}$ denote the convergence in finite-dimensional distributions of $\mathcal{X}^n$ to $\mathcal{X}$ and in probability, respectively.

2. Model and Assumptions In this section, we present a detailed description of our model and the assumptions. As shown in Figure 1(a), there is a single class of jobs, and each job is forked into K parallel tasks, $K \geq 2$. Each task is processed in a service station with multiple servers under the FCFS discipline. There is an infinite number of servers at each station. After service completion, each task will join a waiting buffer for synchronization associated with each service station, and when all tasks of the same job are completed, they will be synchronized and leave the system. Here we assume that the synchronization process takes zero amount of time.

Let $A := \{A(t) : t \geq 0\}$ be the arrival process of jobs with τ_i representing the arrival time of the i^{th} job, $i \in \mathbb{N}$. Let $\{\boldsymbol{\eta}^i : i \geq 1\}$ denote the i.i.d. service time vectors of the parallel tasks. The joint distribution of the service time vector for the i^{th} job $\boldsymbol{\eta}^i$ is $F(\boldsymbol{x}) := F(x_1, ..., x_K)$ for $x_k \geq 0$, $k = 1, ..., K$. Their marginal distributions are $F_k(x)$, for $x \geq 0$, $k = 1, ..., K$. The joint distribution of any two service times η_j^i and η_k^i is $F_{j,k}(x_j, x_k) := P(\eta_j^i \leq x_j, \eta_k^i \leq x_k)$ for $x_j, x_k \geq 0$, $j, k = 1, ..., K$. Note $F_{j,k}(\cdot,\cdot) = F_k(\cdot)$ when $j = k$ for $j, k = 1, ..., K$. We denote $F_{j,k}^c(x_j, x_k) := P(\eta_j^i > x_j, \eta_k^i > x_k) = 1 - F_j(x_j) - F_k(x_k) + F_{j,k}(x_j, x_k)$ for $x_j, x_k \geq 0$, $j, k = 1, ..., K$. Note $F_{j,k}^c(\cdot,\cdot) = F_k^c(\cdot)$ when $j = k$ for $j, k = 1, ..., K$. Let $\eta_{\mathrm{m}}^i := \max\{\eta_1^i, ..., \eta_K^i\}$ be the maximum of the components in the service vector $\boldsymbol{\eta}^i$, and $F_{\mathrm{m}}(x) := P(\eta_{\mathrm{m}}^i \leq x) = F(x, ..., x)$ for $x \geq 0$. (Throughout the paper, we use subscript "m" to index quantities and processes associated with the maximum.) The service process is assumed to be independent of the arrivals. We exclude the case of perfectly positively correlated parallel services since that will lead to empty waiting buffers for synchronization.

Let $X_k := \{X_k(t) : t \geq 0\}$ be the process counting the number of tasks in service at the service station k, and $Y_k := \{Y_k(t) : t \geq 0\}$ be the process counting the number of tasks in the waiting buffer for synchronization (unsynchronized queue) after service completion at service station k, $k = 1, ..., K$. Let $S := \{S(t) : t \geq 0\}$ be the process counting the number of synchronized jobs and $D_k := \{D_k(t) : t \geq 0\}$ be the process counting the number of tasks that have completed service at station k, $k = 1, ..., K$. Denote $\boldsymbol{X} := (X_1, ..., X_K)$, $\boldsymbol{Y} := (Y_1, ..., Y_K)$ and $\boldsymbol{D} := (D_1, ..., D_K)$. We assume that the system starts empty.

We first obtain the following properties on the processes $\boldsymbol{X}(t)$, $\boldsymbol{Y}(t)$ and $S(t)$ at each time $t \geq 0$ when the arrival process A is Poisson.

PROPOSITION 2.1. *If the arrival process A is Poisson with rate λ, then at each time $t \geq 0$, for $k = 1, ..., K$, $X_k(t)$ has a Poisson distribution with rate $\lambda \int_0^t F_k^c(s)ds$, $Y_k(t)$ has a Poisson distribution with rate $\lambda \int_0^t (F_{\mathrm{m}}^c(s) - F_k^c(s))ds$, and $S(t)$ has a Poisson distribution with rate $\lambda \int_0^t F_{\mathrm{m}}(s)ds$. For each time $t \geq 0$ and $j, k = 1, ..., K$,*

$$Cov(X_j(t), X_k(t)) = \lambda \int_0^t F_{j,k}^c(s,s)ds, \tag{2.1}$$

$$Cov(Y_j(t), Y_k(t)) = \lambda \int_0^t (F_{j,k}(s,s) - F_{\mathrm{m}}(s))ds, \tag{2.2}$$

$$Cov(X_j(t), Y_k(t)) = \lambda \int_0^t (F_k(s) - F_{j,k}(s,s))ds. \tag{2.3}$$

For each time $t \geq 0$ and $k = 1, ..., K$, $S(t)$ is independent of $X_k(t)$ and $Y_k(t)$. When $K = 2$, $Y_1(t)$ and $Y_2(t)$ are independent for each $t \geq 0$.

Proof. The results follow from applying Poisson random measure theory and direct calculations, together with an illustrative figure generalizing Figure 1(b) for $K \geq 2$. □

When the arrival process A is general, we will obtain heavy-traffic limits for the fluid and diffusion scaled processes of $(\boldsymbol{X}, \boldsymbol{Y}, S)$ jointly. We will let the arrival rate grow large for the system to be in heavy traffic. For that, we consider a sequence of such systems indexed by n and use superscript n for the processes $A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{D}, S$, and the arrival times $\{\tau_i : i \geq 1\}$, but we let the service vectors $\{\boldsymbol{\eta}^i : i \geq 1\}$ and their distribution functions be independent of n. We make the following assumption on the arrival process A^n.

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 $\hat{A}$ with continuous sample paths, such that

$$\hat{A}^n := n^{-\frac{1}{2}}(A^n - n\bar{a}) \Rightarrow \hat{A} \quad \text{in} \quad \mathbb{D} \quad \text{as} \quad n \to \infty. \tag{2.4}$$

□

It follows from (2.4) that we have the associated FWLLN

$$\bar{A}^n := \frac{A^n}{n} \Rightarrow \bar{a} \quad \text{in} \quad \mathbb{D} \quad \text{as} \quad n \to \infty. \tag{2.5}$$

When the arrival process is renewal, the limit in (2.5) is $\bar{a}(t) = \lambda t$, for $t \geq 0$ and some positive constant λ, and the limit in (2.4) is $\hat{A} = \sqrt{\lambda c_a^2} B_a$, where c_a^2 is the squared coefficient of variation (SCV) of an interarrival time, and B_a is a standard Brownian motion (BM).

We also make a regularity assumption on the joint service-time distribution function $F(\boldsymbol{x})$.

Assumption 2: Service time distributions. The joint distribution function $F(\boldsymbol{x})$ of the service time vectors $\{\boldsymbol{\eta}^i : i \in \mathbb{N}\}$ is continuous. □

From the graphical representation of the system dynamics in Figure 1(b), we can write, for each $t \geq 0$ and $k = 1, ..., K$,

$$X_k^n(t) = \sum_{i=1}^{A^n(t)} \mathbf{1}(\tau_i^n + \eta_k^i > t), \tag{2.6}$$

$$\begin{aligned} Y_k^n(t) &= \sum_{i=1}^{A^n(t)} \mathbf{1}(\tau_i^n + \eta_k^i \leq t \quad \text{and} \quad \tau_i^n + \eta_{k'}^i > t \quad \text{for some } k' \neq k) \\ &= \sum_{i=1}^{A^n(t)} \left(\mathbf{1}(\tau_i^n + \eta_k^i \leq t) - \mathbf{1}(\tau_i^n + \eta_{\mathrm{m}}^i \leq t) \right) \\ &= \sum_{i=1}^{A^n(t)} \left(\mathbf{1}(\tau_i^n + \eta_{\mathrm{m}}^i > t) - \mathbf{1}(\tau_i^n + \eta_k^i > t) \right), \end{aligned} \tag{2.7}$$

$$S^n(t) = \sum_{i=1}^{A^n(t)} \mathbf{1}(\tau_i^n + \eta_{\rm m}^i \le t) = \sum_{i=1}^{A^n(t)} \mathbf{1}(\tau_i^n + \eta_k^i \le t, \ \forall k), \tag{2.8}$$

$$D_k^n(t) = \sum_{i=1}^{A^n(t)} \mathbf{1}(\tau_i^n + \eta_k^i \le t). \tag{2.9}$$

The following balanced equations hold for each $t \ge 0$ and $k = 1, ..., K$,

$$D_k^n(t) = A^n(t) - X_k^n(t), \tag{2.10}$$

$$Y_k^n(t) = D_k^n(t) - S^n(t). \tag{2.11}$$

As we have remarked in the introduction, by previous work on $G/GI/\infty$ queues [37], each individual process X_k^n and D_k^n (*resp.*, S^n) can be represented by an integral of a sequential empirical process driven by a sequence of i.i.d. random variables $\{\eta_k^i : i \ge 1\}$ (*resp.*, $\{\eta_{\rm m}^i : i \ge 1\}$) for each $k = 1, ..., K$. Thus, Gaussian limits for the diffusion-scaled processes X_k^n, D_k^n and S^n in heavy traffic for each k can be established, and as a consequence, a Gaussian limit for the diffusion-scaled process Y_k^n can be obtained from those of D_k^n and S^n, $k = 1, ..., K$. However, that approach does not give a characterization of the joint Gaussian distribution of the limiting processes of the diffusion-scaled processes $(\boldsymbol{X}^n, \boldsymbol{Y}^n, S^n)$.

We will represent all the processes $\boldsymbol{X}^n, \boldsymbol{Y}^n, S^n$ as integrals of a multiparameter sequential empirical process $\bar{K}^n := \{\bar{K}^n(t, \boldsymbol{x}) : t \ge 0, \boldsymbol{x} \in \mathbb{R}_+^K\}$ driven by the sequence of service vectors $\{\boldsymbol{\eta}^i : i \ge 1\}$:

$$\bar{K}^n(t, \boldsymbol{x}) := \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \mathbf{1}(\boldsymbol{\eta}^i \le \boldsymbol{x}), \quad t \ge 0, \quad \boldsymbol{x} \in \mathbb{R}_+^K. \tag{2.12}$$

That is, we write, for $t \ge 0$ and $k = 1, ..., K$,

$$X_k^n(t) = n \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}(s + x_k > t) d\bar{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right), \tag{2.13}$$

$$Y_k^n(t) = n \int_0^t \int_{\mathbb{R}_+^K} \left(\mathbf{1}(s + x_k \le t) - \mathbf{1}(s + x_j \le t, \ \forall j)\right) d\bar{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right), \tag{2.14}$$

and

$$S^n(t) = n \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}(s + x_j \le t, \ \forall j) d\bar{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right). \tag{2.15}$$

The integrals in (2.13), (2.14) and (2.15) are well-defined as a Stieltjes integral for functions of bounded variation as integrators.

2.1. Comparison with a fork-join network with ES We make a comparison with an associated fork-join network with ES. We use superscript "ES" in the corresponding processes for this model. Let the arrival and service processes be the same as the model described above. The only difference is the synchronization constraint. Here tasks are not tagged with a particular job, so that whenever there are tasks completed at all parallel service stations, the oldest completed task at each waiting buffer for synchronization will be synchronized. It is evident that when the arrival process $A(t)$ is Poisson, the processes $Y_k^{ES}(t)$ and $S^{ES}(t)$ do not have a Poisson distribution at each time $t \ge 0$, $k = 1, ..., K$. In this case, for each $k = 1, ..., K$, $X_k^{n,ES}$ and $D_k^{n,ES}$ will have the same representations as in (2.6) and (2.9), but the processes $S^{n,ES}$ and $Y_k^{n,ES}$ become

$$S^{n,ES}(t) = \min_{1 \le j \le K} \{D_j^{n,ES}(t)\}, \quad t \ge 0, \tag{2.16}$$

and

$$Y_k^{n,ES}(t) = D_k^{n,ES}(t) - S^{n,ES}(t) = D_k^{n,ES}(t) - \min_{1\le j\le K}\{D_j^{n,ES}(t)\}, \quad t\ge 0. \tag{2.17}$$

Thus, at any time, one of the waiting buffers for synchronization should be empty. It is evident that the processes $S^{n,ES}$ and $Y_k^{n,ES}$, $k=1,...,K$, cannot be represented as a single integral of the multiparameter sequential empirical process $\bar{K}^n$ as in equations (2.15) and (2.14), respectively. We will discuss more on the comparison in §3.3 for the steady-state mean values of the fluid limits of these processes when the arrival rate is constant.

3. Main Results In this section, we present the main results of the paper. In §3.1, we state an FCLT for multiparameter sequential empirical processes driven by a sequence of i.i.d. random vectors. In §3.2, we present the FWLLN and FCLT for the fluid and diffusion scaled processes $(\boldsymbol{X}^n,\boldsymbol{Y}^n)$ and S^n. In §3.3, we give the Gaussian characterizations of the limit processes. We provide the proofs in §§4-6.

3.1. An FCLT for Multiparameter Sequential Empirical Processes We present an FCLT for multiparameter sequential empirical processes $\hat{U}^n := \{\hat{U}^n(t,\boldsymbol{x}): t\ge 0,\ \boldsymbol{x}\in[0,1]^K\}$ driven by a sequence of i.i.d. random vectors with uniform marginals:

$$\hat{U}^n(t,\boldsymbol{x}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbf{1}(\boldsymbol{\xi}^i\le \boldsymbol{x}) - H(\boldsymbol{x})\right), \quad t\ge 0, \quad \boldsymbol{x}\in[0,1]^K, \tag{3.1}$$

where for each $i\in\mathbb{N}$, $\boldsymbol{\xi}^i := (\xi_1^i,...,\xi_K^i)$ is a vector of nonnegative random variables with continuous joint distribution function $H(\cdot)$ and uniform marginals over $[0,1]$.

The convergence for the processes $\hat{U}^n(t,\boldsymbol{x})$ is established in the space $\mathbb{D}([0,\infty),\mathbb{D}([0,1]^K,\mathbb{R}))$. We remark that this theorem is in the same spirit as Lemma 3.1 in [37], where an FCLT is proved for the two-parameter process $\hat{U}^n(t,x)$ in the univariate case in the space $\mathbb{D}([0,\infty),\mathbb{D}([0,1],\mathbb{R}))$ when $K=1$. We generalize that result to the multivariate setting. The proof of the theorem is given in §4.

THEOREM 3.1. *The multiparameter sequential empirical processes $\hat{U}^n(t,\boldsymbol{x})$ defined in (3.1) converge weakly to a continuous Gaussian limit,*

$$\hat{U}^n(t,\boldsymbol{x}) \Rightarrow U(t,\boldsymbol{x}) \quad in \quad \mathbb{D}([0,\infty),\mathbb{D}([0,1]^K,\mathbb{R})) \quad as \quad n\to\infty, \tag{3.2}$$

where $U(t,\boldsymbol{x})$ is a continuous Gaussian random field with mean function $E[U(t,\boldsymbol{x})]=0$ and covariance function

$$Cov(U(t,\boldsymbol{x}),U(s,\boldsymbol{y})) = (t\wedge s)(H(\boldsymbol{x}\wedge\boldsymbol{y}) - H(\boldsymbol{x})H(\boldsymbol{y})), \quad t,s\ge 0, \quad \boldsymbol{x},\boldsymbol{y}\in[0,1]^K.$$

To show the FCLT for the processes $(\boldsymbol{X}^n,\boldsymbol{Y}^n,S^n)$, we define the diffusion-scaled multiparameter sequential empirical processes $\hat{K}^n := \{\hat{K}^n(t,\boldsymbol{x}): t\ge 0, \boldsymbol{x}\in\mathbb{R}_+^K\}$ by

$$\hat{K}^n(t,\boldsymbol{x}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbf{1}(\boldsymbol{\eta}^i\le \boldsymbol{x}) - F(\boldsymbol{x})\right), \quad t\ge 0, \quad \boldsymbol{x}\in\mathbb{R}_+^K. \tag{3.3}$$

Theorem 3.1 can be applied to show an FCLT for the processes $\hat{K}^n$. Define $\boldsymbol{F}:\mathbb{R}^K\to[0,1]^K$ with $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1),...,F_K(x_K))$. By Sklar's theorem [59], a multidimensional version of probability integral transformation, for any multivariate distribution function F, there exists a unique multivariate distribution function H_F (called "copula", depending on F) with uniform marginals on $[0,1]$ such that $F(\boldsymbol{x}) = H_F(\boldsymbol{F}(\boldsymbol{x}))$ when the marginal distribution functions F_k, $k=1,...,K$, are continuous. Then, $\hat{K}^n(\cdot,\cdot)$ can be represented as a composition of $\hat{U}^n(\cdot,\cdot)$ with $\boldsymbol{F}(\cdot)$ in the second component, i.e.,

$$\hat{K}^n(t,\boldsymbol{x}) = \hat{U}^n(t,\boldsymbol{F}(\boldsymbol{x})), \quad t \ge 0, \quad \boldsymbol{x} \in \mathbb{R}_+^K.$$

Thus, it follows from Theorem 3.1 that the processes $\hat{K}^n(t,\boldsymbol{x})$ converge in distribution:

$$\hat{K}^n(t,\boldsymbol{x}) = \hat{U}^n(t,\boldsymbol{F}(\boldsymbol{x})) \Rightarrow \hat{K}(t,\boldsymbol{x}) := U(t,\boldsymbol{F}(\boldsymbol{x})) \quad \text{in} \quad \mathbb{D}([0,\infty),\mathbb{D}_K) \quad \text{as} \quad n\to\infty, \tag{3.4}$$

which implies that

$$\bar{K}^n(t,\boldsymbol{x}) \Rightarrow \bar{k}(t,\boldsymbol{x}) := tF(\boldsymbol{x}) \quad \text{in} \quad \mathbb{D}([0,\infty),\mathbb{D}_K) \quad \text{as} \quad n\to\infty. \tag{3.5}$$

3.2. FWLLN and FCLT for the processes $(\boldsymbol{X}^n, \boldsymbol{Y}^n, S^n)$ We define fluid-scaled processes $\bar{\boldsymbol{X}}^n$, $\bar{\boldsymbol{Y}}^n$ and $\bar{S}^n$ by

$$\bar{\boldsymbol{X}}^n := \frac{1}{n}\boldsymbol{X}^n, \quad \bar{\boldsymbol{Y}}^n := \frac{1}{n}\boldsymbol{Y}^n, \quad \bar{S}^n := \frac{1}{n}S^n. \tag{3.6}$$

The FWLLN for $(\bar{\boldsymbol{X}}^n, \bar{\boldsymbol{Y}}^n, \bar{S}^n)$ is stated in the following theorem.

THEOREM 3.2 **(FWLLN).** *Under Assumptions* 1 *and* 2*, the fluid-scaled processes converge to deterministic fluid functions,*

$$(\bar{A}^n, \bar{\boldsymbol{X}}^n, \bar{\boldsymbol{Y}}^n, \bar{S}^n) \Rightarrow (\bar{a}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}, \bar{S}) \tag{3.7}$$

in $\mathbb{D}^{2K+2}$ *as* $n\to\infty$*, where the limits are all deterministic functions:* $\bar{a}$ *is the limit in* (2.5)*, for each* $t\ge 0$*,*

$$\bar{\boldsymbol{X}}(t) := (\bar{X}_1(t),\dots,\bar{X}_K(t)), \quad \bar{X}_k(t) := \int_0^t F_k^c(t-s)d\bar{a}(s), \quad \textit{for} \quad k=1,\dots,K, \tag{3.8}$$

$$\bar{\boldsymbol{Y}}(t) := (\bar{Y}_1(t),\dots,\bar{Y}_K(t)), \quad \bar{Y}_k(t) := \int_0^t (F_{\mathrm{m}}^c(t-s) - F_k^c(t-s))d\bar{a}(s), \quad \textit{for} \quad k=1,\dots,K, \tag{3.9}$$

$$\bar{S}(t) := \int_0^t F_{\mathrm{m}}(t-s)d\bar{a}(s). \tag{3.10}$$

When $\bar{a}(t)=\lambda t$ *for a constant arrival rate* $\lambda>0$ *and* $E[\eta_k^1]<\infty$ *for* $k=1,\dots,K$*,*

$$\bar{X}_k(\infty) := \lim_{t\to\infty}\bar{X}_k(t) = \lambda E[\eta_k^1], \quad k=1,\dots,K, \tag{3.11}$$

$$\bar{Y}_k(\infty) := \lim_{t\to\infty}\bar{Y}_k(t) = \lambda(E[\eta_{\mathrm{m}}^1] - E[\eta_k^1]), \quad k=1,\dots,K, \tag{3.12}$$

$$\lim_{t\to\infty}\frac{\bar{S}(t)}{t} = \lambda. \tag{3.13}$$

We define the diffusion scaling of $\boldsymbol{X}^n$, $\boldsymbol{Y}^n$ and S^n by

$$\hat{\boldsymbol{X}}^n := \sqrt{n}(\bar{\boldsymbol{X}}^n - \bar{\boldsymbol{X}}), \quad \hat{\boldsymbol{Y}}^n := \sqrt{n}(\bar{\boldsymbol{Y}}^n - \bar{\boldsymbol{Y}}), \quad \hat{S}^n := \sqrt{n}(\bar{S}^n - \bar{S}). \tag{3.14}$$

We will show the following FCLT for these diffusion-scaled processes. The proof is given in §6. Theorem 3.2 follows immediately from this FCLT and thus its proof is omitted.

THEOREM 3.3 **(FCLT).** *Under Assumptions* 1 *and* 2*, the diffusion-scaled processes converge in distribution,*

$$(\hat{A}^n, \hat{K}^n, \hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n) \Rightarrow (\hat{A}, \hat{K}, \hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}, \hat{S}) \tag{3.15}$$

in $\mathbb{D}\times\mathbb{D}([0,\infty),\mathbb{D}_K)\times\mathbb{D}^{2K+1}$ *as* $n\to\infty$*, where* $\hat{A}$ *is the limit in* (2.4)*,* $\hat{K}$ *is the limit in* (3.4)*, which is independent of* $\hat{A}$*, and for* $t\ge 0$ *and* $k=1,...,K$*,*

$$\hat{\boldsymbol{X}}(t) := \hat{\boldsymbol{M}}_1(t)+\hat{\boldsymbol{M}}_2(t), \quad \hat{\boldsymbol{M}}_i(t) := (\hat{M}_{1,i}(t),...,\hat{M}_{K,i}(t)),\ i=1,2, \tag{3.16}$$

$$\hat{M}_{k,1}(t) := \int_0^t F_k^c(t-s)d\hat{A}(s) = \hat{A}(t) - \int_0^t \hat{A}(s)dF_k^c(t-s), \tag{3.17}$$

$$\hat{M}_{k,2}(t) := \int_0^t\int_{\mathbb{R}_+^K} \mathbf{1}(s+x_k>t)d\hat{K}(\bar{a}(s),\boldsymbol{x}) = -\int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_k\le t)d\hat{K}(\bar{a}(s),\boldsymbol{x}), \tag{3.18}$$

$$\hat{S}(t) := \hat{V}_1(t)+\hat{V}_2(t), \tag{3.19}$$

$$\hat{V}_1(t) := \int_0^t F_{\mathrm{m}}(t-s)d\hat{A}(s) = -\int_0^t \hat{A}(s)dF_{\mathrm{m}}(t-s), \tag{3.20}$$

$$\hat{V}_2(t) := \int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_j\le t,\ \forall j)d\hat{K}(\bar{a}(s),\boldsymbol{x}), \tag{3.21}$$

$$\hat{\boldsymbol{Y}}(t) := \hat{\boldsymbol{Z}}_1(t)+\hat{\boldsymbol{Z}}_2(t), \quad \hat{\boldsymbol{Z}}_i(t) := (\hat{Z}_{1,i}(t),...,\hat{Z}_{K,i}(t)),\ i=1,2, \tag{3.22}$$

$$\hat{Z}_{k,1}(t) := \int_0^t (F_k(t-s)-F_{\mathrm{m}}(t-s))d\hat{A}(s) = \int_0^t \hat{A}(s)d(F_{\mathrm{m}}(t-s)-F_k(t-s)), \tag{3.23}$$

$$\hat{Z}_{k,2}(t) := \int_0^t\int_{\mathbb{R}_+^K}(\mathbf{1}(s+x_k\le t)-\mathbf{1}(s+x_j\le t,\ \forall j))d\hat{K}(\bar{a}(s),\boldsymbol{x}) = -\hat{M}_{k,2}(t)-\hat{V}_2(t). \tag{3.24}$$

The processes $\hat{\boldsymbol{M}}_2$, $\hat{\boldsymbol{Z}}_2$ and $\hat{V}_2$ are defined in the mean-square sense; see the precise definitions in Definition 5.1. This is in the same way as the limit process with respect to a standard Kiefer process for the $G/GI/\infty$ queue is defined in [37, 51]. The limit processes are characterized in the next subsection.

3.3. Characterization of the Limit Processes In this section, we show the Gaussian property of the limiting processes $(\hat{\boldsymbol{X}},\hat{\boldsymbol{Y}})$ and $\hat{S}$ when the arrival limit process is a Brownian motion. Its proof is given in §5.

THEOREM 3.4 **(Gaussian Property).** *Under Assumptions* 1 *and* 2*, when the arrival limit process* $\hat{A}$ *is a Brownian motion, i.e.,* $\hat{A}(t)=c_aB_a(\bar{a}(t))$ *for a standard Brownian motion* B_a*, a positive constant* $c_a>0$ *and* $t\ge 0$*, the limiting processes* $(\hat{\boldsymbol{X}},\hat{\boldsymbol{Y}})$ *and* $\hat{S}$ *in Theorem* 3.3 *are well-defined continuous Gaussian processes. For each* $t\ge 0$*,*

$$(\hat{\boldsymbol{X}}(t),\hat{\boldsymbol{Y}}(t)) \stackrel{d}{=} N(\mathbf{0},\Sigma(t)), \quad \textit{and} \quad \hat{S}(t)\stackrel{d}{=}N(0,\sigma^S(t)),$$

where for $j,k=1,...,K$*,*

$$\sigma_{jk}^X(t) := Cov(\hat{X}_j(t),\hat{X}_k(t)) = \int_0^t \Big[F_{j,k}^c(t-s,t-s)+(c_a^2-1)F_j^c(t-s)F_k^c(t-s)\Big]d\bar{a}(s), \tag{3.25}$$

$$\begin{aligned}\sigma_{jk}^Y(t) := Cov(\hat{Y}_j(t),\hat{Y}_k(t)) = \int_0^t &\Big[(F_{j,k}(t-s,t-s)-F_{\mathrm{m}}(t-s)) \\ &+(c_a^2-1)(F_j(t-s)-F_{\mathrm{m}}(t-s))(F_k(t-s)-F_{\mathrm{m}}(t-s))\Big]d\bar{a}(s),\end{aligned} \tag{3.26}$$

$$\begin{aligned}\sigma_{jk}^{XY}(t) := Cov(\hat{X}_j(t),\hat{Y}_k(t)) = \int_0^t &\Big[\,(F_k(t-s)-F_{j,k}(t-s,t-s)) \\ &+(c_a^2-1)\left(F_j^c(t-s)(F_k(t-s)-F_{\mathrm{m}}(t-s))\right)\Big]d\bar{a}(s),\end{aligned} \tag{3.27}$$

and

$$\sigma^S(t) := Var(\hat{S}(t)) = \int_0^t F_{\mathrm{m}}(t-s)d\bar{a}(s)+(c_a^2-1)\int_0^t(F_{\mathrm{m}}(t-s))^2d\bar{a}(s). \tag{3.28}$$

When the arrival rate function $\bar{a}(t) = \lambda t$ *for a positive constant* $\lambda > 0$*,*

$$(\hat{\boldsymbol{X}}(t), \hat{\boldsymbol{Y}}(t)) \Rightarrow (\hat{\boldsymbol{X}}(\infty), \hat{\boldsymbol{Y}}(\infty)) \stackrel{d}{=} N(\mathbf{0}, \Sigma(\infty)) \quad as \quad t \to \infty,$$

$$\lim_{t\to\infty} t^{-1} Var(\hat{S}(t)) = \lambda c_a^2, \tag{3.29}$$

where for $j,k = 1,...,K$*,*

$$\sigma^X_{jk}(\infty) := \lambda \int_0^\infty F^c_{j,k}(s,s)ds + \lambda(c_a^2 - 1)\int_0^\infty F^c_j(s) F^c_k(s) ds, \tag{3.30}$$

$$\sigma^Y_{jk}(\infty) := \lambda \int_0^\infty \Big[(F_{j,k}(s,s) - F_{\mathrm{m}}(s)) + (c_a^2 - 1)(F_j(s) - F_{\mathrm{m}}(s))(F_k(s) - F_{\mathrm{m}}(s)) \Big] ds, \tag{3.31}$$

$$\sigma^{XY}_{jk}(\infty) := \lambda \int_0^\infty \Big[(F_k(s) - F_{j,k}(s,s)) + (c_a^2 - 1)\left(F^c_j(s)(F_k(s) - F_{\mathrm{m}}(s))\right) \Big] ds. \tag{3.32}$$

We make the following remarks on the Gaussian property of the limiting processes.

(i) When we set $c_a^2 = 1$, the variance and covariance formulas coincide with those in the Poisson arrival case in Proposition 2.1.

(ii) When $K = 2$ and $c_a^2 = 1$, $Cov(\hat{Y}_1(t), \hat{Y}_2(t)) = 0$ for $t \geq 0$, even if the service times of parallel tasks are correlated, since both terms inside the integral in (3.26) vanish. This can also be explained using Figure 1(b). In fact, when the arrival process is Poisson and $K = 2$, using Poisson random measure theory, it is shown in Proposition 2.1 that the unscaled processes $Y_1(t)$ and $Y_2(t)$ are independent for each $t \geq 0$.

(iii) We emphasize the interesting structure of the variances of $\hat{X}_k$ and $\hat{Y}_k$ and their covariances, $k = 1,...,K$. Recall that for $G/GI/\infty$ queues [37], the steady-state variance formula of the number of jobs in the system is given as the sum of two terms, the mean and the coefficient $(c_a^2 - 1)$ multiplying an integral associated with the service time distribution; for example, when $E[\eta^1_k] < \infty$, the variance of the steady-state number of tasks in the k^{th} service station is

$$Var(\hat{X}_k(\infty)) = \lambda E[\eta^1_k] + \lambda(c_a^2 - 1)\int_0^\infty (F^c_k(s))^2 ds, \quad k = 1,...,K.$$

It turns out that the steady-state variance formula for the number of tasks in the waiting buffer for synchronization has the same structure; for instance, when $E[\eta^1_k] < \infty$ for $k = 1,...,K$, the variance of the steady-state waiting buffer size at the k^{th} service station is

$$Var(\hat{Y}_k(\infty)) = \lambda(E[\eta^1_{\mathrm{m}}] - E[\eta^1_k]) + \lambda(c_a^2 - 1)\int_0^\infty (F^c_{\mathrm{m}}(s) - F^c_k(s))^2 ds, \quad k = 1,...,K.$$

The same structure also exists for the covariances between $\hat{X}_j$ and $\hat{Y}_k$, as shown in (3.27), for $j,k = 1,...,K$.

(iv) The synchronized process does not have a Brownian motion limit, but its limiting process is Gaussian, and has the same variability as the arrival process when the arrival rate is constant, as shown in (3.29). This can be also explained by regarding the synchronized process as the departure process of a $G/GI/\infty$ queue with the same arrival process and service times as the maximum of the service vectors (see [37, 51]).

□

To explore the impact of the correlation among the service times of each job's parallel tasks on the system dynamics, we consider the case when the service vector $\boldsymbol{\eta}^i$ has the joint continuous distribution function

$$F(\boldsymbol{x}) = (1-\rho)\prod_{k=1}^K G(x_k) + \rho G\left(\min_{k=1,\dots,K}\{x_k\}\right) \tag{3.33}$$

with a marginal continuous distribution function $G(\cdot)$, for $0 \le \rho < 1$, $x_k \ge 0$ and $k = 1, ..., K$. Namely, the service times at the parallel stations have the same distribution, and are symmetrically correlated with a correlation parameter $\rho \in [0,1)$. We state the mean and covariance functions of the performance measures studied above as functions of the parameter ρ in the following corollary. Its proof follows from a direct calculation and is thus omitted.

Corollary 3.1. *Under the same assumptions in Theorem 3.4, when the service vector $\boldsymbol{\eta}^i$ has the joint distribution function F in (3.33), for each $t \ge 0$ and $k = 1, ..., K$, $\bar{X}_k(t)$ and $Var(\hat{X}_k(t))$ are the same as in (3.8) and (3.25), respectively,*

$$\begin{aligned}
\bar{Y}_k(t) &= (1-\rho)\int_0^t \left[G(t-s)(1-(G(t-s))^{K-1})\right] d\bar{a}(s),\\
Var(\hat{Y}_k(t)) &= \int_0^t \Big[(1-\rho)G(t-s)(1-(G(t-s))^{K-1})\\
&\qquad + (1-\rho)^2(c_a^2-1)(G(t-s))^2\left(1-(G(t-s))^{K-1}\right)^2\Big] d\bar{a}(s),\\
Cov(\hat{X}_k(t),\hat{Y}_k(t)) &= (c_a^2-1)(1-\rho)\int_0^t \left[G^c(t-s)G(t-s)(1-(G(t-s))^{K-1})\right] d\bar{a}(s),
\end{aligned}$$

for $j, k = 1, ..., K$ and $j \ne k$,

$$\begin{aligned}
Cov(\hat{X}_j(t),\hat{X}_k(t)) &= \int_0^t \Big[(1-\rho)(G^c(t-s))^2 + \rho G^c(t-s) + (c_a^2-1)(G^c(t-s))^2\Big] d\bar{a}(s),\\
Cov(\hat{Y}_j(t),\hat{Y}_k(t)) &= \int_0^t \Big[(1-\rho)(G(t-s))^2(1-(G(t-s))^{K-2})\\
&\qquad + (1-\rho)^2(c_a^2-1)(G(t-s))^2\left(1-(G(t-s))^{K-1}\right)^2\Big] d\bar{a}(s),\\
Cov(\hat{X}_j(t),\hat{Y}_k(t)) &= (1-\rho)\int_0^t \Big[G(t-s)G^c(t-s)\\
&\qquad + (c_a^2-1)G^c(t-s)G(t-s)\left(1-(G(t-s))^{K-1}\right)\Big] d\bar{a}(s),
\end{aligned}$$

and

$$\begin{aligned}
\bar{S}(t) &= \int_0^t \Big[(1-\rho)(G(t-s))^K + \rho G(t-s)\Big] d\bar{a}(s),\\
Var(\hat{S}(t)) &= \int_0^t \left[(1-\rho)(G(t-s))^K + \rho G(t-s)\right] d\bar{a}(s)\\
&\qquad + (c_a^2-1)\int_0^t \left[(1-\rho)(G(t-s))^K + \rho G(t-s)\right]^2 d\bar{a}(s).
\end{aligned}$$

We make several remarks on the impact of the correlation ρ among the service vector with distribution (3.33). The mean and the variance of $X_k(t)$ are not affected by the correlation, but the covariances of $X_j(t)$ and $X_k(t)$ increase linearly in ρ for $t \ge 0$ and $j, k = 1, ...K$ with $j \ne k$, as $\rho \uparrow 1$. The mean of $Y_k(t)$ decreases linearly in ρ and the mean of $S(t)$ increases linearly in ρ for $t \ge 0$ and $k = 1, ..., K$, as $\rho \uparrow 1$. The covariances of $Y_j(t)$ and $Y_k(t)$ decrease nonlinearly in ρ, but the covariances of $X_j(t)$ and $Y_k(t)$ decrease linearly in ρ for $t \ge 0$ and $j, k = 1, ...K$, as $\rho \uparrow 1$. The variance of $S(t)$ increases nonlinearly in ρ, for $t \ge 0$, as $\rho \uparrow 1$. The intuitive interpretation for these observations is that positive correlation makes the parallel tasks more likely to finish close to each other so that the waiting time for synchronization becomes less and more jobs are synchronized. It is also important to emphasize that the covariances of $Y_j(t)$ and $Y_k(t)$ and the covariances of $X_j(t)$ and $Y_k(t)$ decrease in different orders in the correlation parameter ρ for $t \ge 0$ and $j, k = 1, ...K$, as $\rho \uparrow 1$. The same observations hold for the associated steady-state performance measures.

3.3.1. Comparing the steady-state mean values with a fork-join network with ES In §2.1, we have discussed the difference between our model and a fork-join network with ES. In the ES model, the synchronization process S^{ES} can be represented as the minimum of the departure processes from all parallel stations, and these departure processes are dependent due to the correlation of service vector of each job. Thus, we are unable to obtain a distributional approximation of the processes S^{ES} and Y_k^{ES}, $k=1,...,K$. However, for each $t\geq 0$, by applying the previous results on $G/GI/\infty$ queues [37], we can obtain the mean values of the fluid limit $\bar{Y}_k^{ES}(t)$, $k=1,...,K$, and $\bar{S}^{ES}(t)$:

$$\bar{Y}_k^{ES}(t) := \lambda\left[\int_0^t F_k(s)ds - \min_{1\leq j\leq K}\left\{\int_0^t F_j(s)ds\right\}\right] \tag{3.34}$$

$$\longrightarrow \bar{Y}_k^{ES}(\infty) := \lambda\left(\max_{1\leq j\leq K}\{E[\eta_j^1]\} - E[\eta_k^1]\right) \quad \text{as} \quad t\to\infty,$$

$$\bar{S}^{ES}(t) := \lambda \min_{1\leq j\leq K}\left\{\int_0^t F_j(s)ds\right\} = \lambda t - \lambda\max_{1\leq j\leq K}\left\{\int_0^t F_j^c(s)ds\right\}, \quad \text{and} \quad \lim_{t\to\infty}\frac{\bar{S}^{ES}(t)}{t} = \lambda. \tag{3.35}$$

Recall that the steady-state mean value of the waiting buffer for synchronization in our model $\bar{Y}_k(\infty) = \lambda\left(E[\eta_{\rm m}^1] - E[\eta_k^1]\right)$ in (3.12), $k=1,...,K$, denoted as $\bar{Y}_k^{NES}(\infty)$ for the comparison purpose. It is evident that the average waiting buffer sizes for synchronization under NES constraint are larger than those under ES constraint, even though the total synchronization throughput rates are the same, $\lim_{t\to\infty}\bar{S}^{ES}(t)/t = \lim_{t\to\infty}\bar{S}^{NES}(t)/t = \lambda$. We also observe that when the parallel service times are perfectly positively correlated, the difference $\bar{Y}_k^{NES}(\infty) - \bar{Y}_k^{ES}(\infty)$ becomes zero for $k=1,...,K$. We summarize this comparison result in the following proposition.

PROPOSITION 3.1. *Under Assumptions* 1 *and* 2*, when* $\bar{a}(t) = \lambda t$ *for a positive arrival rate* $\lambda > 0$ *and* $E[\eta_k^1] < \infty$ *for* $k=1,...,K$*,*

$$\bar{Y}_k^{NES}(\infty) - \bar{Y}_k^{ES}(\infty) = \lambda\big(E[\eta_{\rm m}^1] - \max_{1\leq j\leq K}\{E[\eta_j^1]\}\big) \geq 0, \quad \textit{for} \quad k=1,...,K. \tag{3.36}$$

By the extreme value theory, if the service vector has i.i.d. components such that the service time distribution lies in the domain of attraction for Gumbel extremal distribution, then we have $a_K(\eta_{\rm m}^1 - b_K) \Rightarrow Z$ as $K\to\infty$, where Z has a Gumbel distribution, and a_K and b_K are constants depending on K; see Chapter 1 in [40]. The Gumbel distribution has cdf $P(Z\leq z) = e^{-e^{-z}}$, $z\geq 0$, with mean $E[Z] = \gamma \approx 0.5772$, the Euler-Mascheroni constant, and variance $Var(Z) = \pi/\sqrt{6} \approx 1.2825$. For one example, if the service vector has i.i.d. components of an exponential distribution with rate 1, then $a_K = 1$ and $b_K = \ln(K)$ (see Example 1.7.2 of [40]), for $k=1,...,K$,

$$\bar{Y}_k^{NES}(\infty) - \bar{Y}_k^{ES}(\infty) \approx \lambda(\ln(K) + \gamma - 1) \quad \text{as} \quad K\to\infty. \tag{3.37}$$

For another example, if the service vector has i.i.d. components of a lognormal distribution $LN(0,1)$, we have, for $k=1,...,K$,

$$\bar{Y}_k^{NES}(\infty) - \bar{Y}_k^{ES}(\infty) \approx \lambda(\gamma/a_K + b_K - e^{1/2}) \quad \text{as} \quad K\to\infty, \tag{3.38}$$

where a_K and b_K are (see Example 1.7.4 of [40]):

$$a_K = (2\ln K)^{1/2}\exp\left\{-(2\ln K)^{1/2} + 0.5(2\ln K)^{-1/2}(\ln\ln K + \ln(4\pi))\right\},$$

and

$$b_K = \exp\left\{(2\ln K)^{1/2} - 0.5(2\ln K)^{-1/2}(\ln\ln K + \ln(4\pi))\right\}.$$

3.3.2. Numerical Example In this section, we provide a numerical example with two parallel tasks ($K = 2$), comparing our approximations with simulations. We let the arrival process be renewal with arrival rate $\lambda = 100$ and the SCV $c_a^2 = 5$. The service times of the two parallel tasks are assumed to be a bivariate Marshall-Olkin hyperexponential distribution, which is a mixture of two independent bivariate Marshall-Olkin exponential distributions [47]. A bivariate Marshall-Olkin exponential distribution function $F_{MO}(x,y)$ for the random vector (X,Y) can be written as $F^c_{MO}(x,y) := P(\mathrm{X} > x, \mathrm{Y} > y) = \exp(-\mu_1 x - \mu_2 y - \mu_{12}(x \vee y))$, $x, y \geq 0$, where three parameters μ_1, μ_2, μ_{12} are such that the two marginals are exponential with rates $\mu_1 + \mu_{12}$ and $\mu_2 + \mu_{12}$ and their correlation $\rho_o = \mu_{12}/(\mu_1 + \mu_2 + \mu_{12}) \in [0,1]$. We denote $MO(\lambda_1, \lambda_2, \rho_o)$ for a bivariate Marshall-Olkin exponential distribution, where λ_1 and λ_2 are the rates for the marginals, and ρ_o is the correlation parameter, for which the parameters $\mu_1 = (\lambda_1 - \rho_o\lambda_2)/(1+\rho_o)$, $\mu_2 = (\lambda_2 - \rho_o\lambda_1)/(1+\rho_o)$ and $\mu_{12} = (\rho_o(\lambda_1 + \lambda_2))/(1+\rho_o)$.

In Table 1, we show the approximation values for the mean, variance and covariance of X_k and Y_k, for $k = 1, 2$, and compare them with the corresponding simulated values, in two cases - independent parallel service times and correlated parallel service times (with correlation coefficient ρ equal to 0.5). In the numerical examples, we take a mixture of $MO(4/5, 1, \rho_{o,1})$ with probability 0.4 and $MO(6/5, 6/5, \rho_{o,2})$ with probability 0.6, such that the means of the two hyperexponential marginals are $m_{s,1} = 1$ and $m_{s,2} = 0.9$. By setting $\rho_{o,1} = \rho_{o,2} = 0$, we have two independent parallel service times ($\rho = 0$ in Table 1), and by setting $\rho_{o,1} = 0.7$ and $\rho_{o,2} = 172/679$, we obtain that the correlation coefficient (see the correlation formula in §5.2 [52]) between the two parallel service times is equal to 0.5 ($\rho = 0.5$ in Table 1).

To estimate the simulated values, we simulated the system up to time 40 with 4000 independent replications starting with an empty system, which we call one experiment. In each replication, we collected data over the time interval $[20, 40]$ and formed the time average (the system tends to be in steady state in less than 5 time units). We conducted 5 independent experiments and took sample averages as estimations for simulated values. To construct the 95% confidence interval (CI), we used Student t-distribution with four degrees of freedom. The halfwidth of the 95% CI is $2.776 s_5/\sqrt{5}$, where s_5 is the sample deviation.

Table 1. Comparing approximations with simulations in a stationary model

(X_1, X_2)		$(E[X_1], E[X_2])$	$(Var(X_1), Var(X_2))$	$Cov(X_1, X_2)$
$\rho = 0$	Sim. (95% CI.)	(99.99 ± 0.17 , 89.98 ± 0.12)	(296.26 ± 0.66, 269.46 ± 0.70)	234.14 ± 0.66
	Approx.	(100.00, 90.00)	(296.00, 269.27)	233.99
$\rho = 0.5$	Sim. (95% CI.)	(99.98 ± 0.04, 89.99 ± 0.04)	(296.08 ± 0.57, 269.23 ± 0.80)	256.34 ± 0.43
	Approx.	(100.00, 90.00)	(296.00, 269.27)	256.30

(Y_1, Y_2)		$(E[Y_1], E[Y_2])$	$(Var(Y_1), Var(Y_2))$	$Cov(Y_1, Y_2)$
$\rho = 0$	Sim. (95% CI.)	(43.18 ± 0.05 , 53.20 ± 0.10)	(70.12 ± 0.20, 89.85 ± 0.40)	31.53 ± 0.30
	Approx.	(43.20, 53.20)	(70.31, 90.08)	31.55
$\rho = 0.5$	Sim. (95% CI.)	(20.89 ± 0.01, 30.88 ± 0.02)	(27.14 ± 0.15, 42.23 ± 0.35)	8.36 ± 0.07
	Approx.	(20.89, 30.89)	(27.05, 42.23)	8.31

(X, Y)		$Cov(X_1, Y_1)$	$Cov(X_1, Y_2)$	$Cov(X_2, Y_1)$	$Cov(X_2, Y_2)$
$\rho = 0$	Sim. (95% CI.)	60.80 (± 0.59)	122.87 (± 0.61)	99.21 (± 0.42)	64.56 (± 0.54)
	Approx.	61.09	123.10	99.85	64.57
$\rho = 0.5$	Sim. (95% CI.)	28.72 (± 0.33)	68.37 (± 0.73)	47.51 (± 0.42)	34.49 (± 0.44)
	Approx.	28.67	68.37	47.41	34.44

We make several remarks for the numerical example. First, our approximations match very well with the simulated values. Second, the size of waiting buffers for synchronization is quite large, of the same order as the number of tasks in the service stations. Third, we find that when the two parallel tasks are positively correlated, the mean and the variance of X_k's are the same as those in the independent case, while the covariance between X_1 and X_2 gets larger, the mean and the variance and covariances of Y_k's and the covariances between X_k and Y_j become smaller than those in the independent case, $j,k=1,2$. These are also consistent with the observations in Corollary 3.1. Note that this numerical example is more general than that considered in Corollary 3.1.

4. Proof of Theorem 3.1 In this section, we prove Theorem 3.1. We will use properties of multiparameter martingales and multiparameter point processes, which we first review briefly for completeness in §§4.1 and 4.2.

4.1. Preliminaries We will first review the definitions associated with multiparameter martingales; see, e.g., [33] for a more detailed introduction. We will then state the Cairoli's Strong (p,p) inequality, which is an extension of Doob's maximal inequality in the multiparameter setting (see, e.g., Theorem 2.3.2 of §7 in [33]). For completeness, we will state the result without proof.

Definition 4.1 (Multiparameter Filtration). In a probability space $(\Omega,\mathcal{G},P)$, a collection $\mathcal{F}:=\{\mathcal{F}_{\boldsymbol{t}}:\boldsymbol{t}\in\mathbb{R}_+^K\}$ is said to be a (K-*parameter*) *filtration* if $\mathcal{F}$ is a collection of sub-σ-fields of $\mathcal{G}$ such that $\mathcal{F}_{\boldsymbol{s}}\subseteq\mathcal{F}_{\boldsymbol{t}}$ for $\boldsymbol{s}\le\boldsymbol{t}$ and $\boldsymbol{s},\boldsymbol{t}\in\mathbb{R}_+^K$.

Definition 4.2 (Multiparameter Submartingale). Suppose $\mathcal{F}:=\{\mathcal{F}_{\boldsymbol{t}}:\boldsymbol{t}\in\mathbb{R}_+^K\}$ is a filtration of sub-σ-fields of $\mathcal{G}$ in a probability space $(\Omega,\mathcal{G},P)$. A real-valued stochastic process $M:=\{M(\boldsymbol{t}):\boldsymbol{t}\in\mathbb{R}_+^K\}$ is a *multiparameter submartingale* with respect to $\mathcal{F}$ if

(i) M is adapted to $\mathcal{F}$. That is, for all $\boldsymbol{t}\in\mathbb{R}_+^K$, $M(\boldsymbol{t})$ is $\mathcal{F}_{\boldsymbol{t}}$-measurable.
(ii) M is integrable.
(iii) For all $\boldsymbol{s}\le\boldsymbol{t}$ both in $\mathbb{R}_+^K$, $E[M(\boldsymbol{t})|\mathcal{F}_{\boldsymbol{s}}]\ge M(\boldsymbol{s})$, *a.s.*

A stochastic process M is a *multiparameter supermartingale* if $-M$ is a multiparameter submartingale. It is a *multiparameter martingale* if it is both a multiparameter submartingale and supermartingale.

Definition 4.3 (Separability). A stochastic process $X:=\{X(\boldsymbol{t}):\boldsymbol{t}\in\mathbb{R}_+^K\}$ is said to be *separable* if there exists an at most countable collection $T\subset\mathbb{R}_+^K$ and a null set $\mathcal{N}$ such that for all closed sets $C\subset\mathbb{R}$ and all open sets $I\subset\mathbb{R}_+^K$ of the form $I=(\boldsymbol{\alpha},\boldsymbol{\beta}):=\{\boldsymbol{t}\in\mathbb{R}_+^K:\alpha_k<t_k<\beta_k,\ k=1,...,K\}$, where $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_K)$, $\boldsymbol{\beta}=(\beta_1,...,\beta_K)$ and α_k and β_k are rational or infinite for $k=1,...,K$, then

$$\{X(\boldsymbol{s})\in C,\ \forall\ \boldsymbol{s}\in I\cap T\}\backslash\{X(\boldsymbol{s})\in C,\ \forall\ \boldsymbol{s}\in I\}\subset\mathcal{N}.$$

We remark that by Doob's separability theorem (see Theorem 2.2.1 of §5 in [33]), any stochastic process $X:=\{X(\boldsymbol{t}):\boldsymbol{t}\in\mathbb{R}_+^K\}$ has a separable modification. Thus, we assume the processes defined in the paper are all separable.

Definition 4.4 (Commuting Filtration). The K-parameter filtration $\mathcal{F}$ is *commuting* if for all $\boldsymbol{s},\boldsymbol{t}\in\mathbb{R}_+^K$ and all bounded $\mathcal{F}_{\boldsymbol{t}}$-measurable random variables Y,

$$E[Y|\mathcal{F}_{\boldsymbol{s}}]=E[Y|\mathcal{F}_{\boldsymbol{s}\wedge\boldsymbol{t}}],\quad a.s.$$

Equivalently, for all $\boldsymbol{s},\boldsymbol{t}\in\mathbb{R}_+^K$, given $\mathcal{F}_{\boldsymbol{s}\wedge\boldsymbol{t}}$, $\mathcal{F}_{\boldsymbol{s}}$ and $\mathcal{F}_{\boldsymbol{t}}$ are conditionally independent (see, e.g., Theorem 2.1.1 of §7 in [33]).

Recall that a function $f:\mathbb{R}_+^K\to\mathbb{R}$ is right-continuous (with respect to the partial order $\le$) if for all $\boldsymbol{t}\in\mathbb{R}_+^K$, $\lim_{\boldsymbol{t}\le\boldsymbol{s}:\boldsymbol{s}\to\boldsymbol{t}}f(\boldsymbol{s})=f(\boldsymbol{t})$. Now we are ready to state the Cairoli's Strong (p,p) inequality.

Lemma 4.1 **(Cairoli's Strong (p,p) Inequality)**. *If $M:=\{M(\boldsymbol{t}):\boldsymbol{t}\in\mathbb{R}_+^K\}$ is a separable, nonnegative multiparameter submartingale with respect to a commuting (K-parameter) filtration, and $\boldsymbol{t}\to E[(M(\boldsymbol{t}))^p]$ is right-continuous, then for any $\boldsymbol{t}\in\mathbb{R}_+^K$ and all $p>1$,*

$$E\left[\sup_{\boldsymbol{s}\le\boldsymbol{t}}\{(M(\boldsymbol{s}))^p\}\right]\le\left(\frac{p}{p-1}\right)^{Kp}E[(M(\boldsymbol{t}))^p].$$

4.2. Compensator of Multiparameter Point Processes We introduce the compensator of multiparameter point processes. For one-parameter point processes, we refer to Jacod and Shiryaev [29] (§II.3c), but for multiparameter point processes, we refer to Ivanoff and Merzbach [28] for the introduction of general set-valued point processes and their compensators (pp. 94-95, Theorem 4.5.3). Here we review multiparameter point processes and their compensators, and generalize the proofs in [29] to the multiparameter setting. The proofs are included in the appendix for completeness. We start with multiparameter one-point process $N^i := \{N^i(\boldsymbol{t}) : t \in \mathbb{R}_+^K\}$ defined by

$$N^i(\boldsymbol{t}) := \mathbf{1}(\boldsymbol{\zeta}^i \le \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}_+^K, \quad i \in \mathbb{N},$$

where $\{\boldsymbol{\zeta}^i : i \in \mathbb{N}\}$ is a sequence of i.i.d. random vectors of $\mathbb{R}_+^K$ with joint distribution function $H_\zeta(\cdot)$. We assume that $H_\zeta(\cdot)$ is continuous. Before introducing the compensator of N^i, we construct the filtration $\mathcal{F}^i := \{\mathcal{F}^i_{\boldsymbol{t}} : \boldsymbol{t} \in \mathbb{R}_+^K\}$ shown in the following lemma.

Lemma 4.2. *The filtration $\mathcal{F}^i = \{\mathcal{F}^i_{\boldsymbol{t}} : \boldsymbol{t} \in \mathbb{R}_+^K\}$ generated by N^i is the following: for all $\boldsymbol{t} \in \mathbb{R}_+^K$, $\mathcal{F}^i_{\boldsymbol{t}}$ is the class of all sets of the form $C = (\boldsymbol{\zeta}^i)^{-1}(B)$, where B is a Borel subset of $\mathbb{R}_+^K$ such that either $L^c_{\boldsymbol{t}} \subseteq B$ or $L^c_{\boldsymbol{t}} \cap B = \emptyset$.*

With the filtration constructed above, the compensator of N^i is given in the following lemma.

Lemma 4.3. *Let $\mathcal{F}^i$ be the filtration defined in Lemma 4.2. Then the compensator of N^i is*

$$\Lambda^i(\boldsymbol{t}) := \int_{L_{\boldsymbol{t}}} \mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u}) \frac{H_\zeta(d\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}, \quad \boldsymbol{t} \in \mathbb{R}_+^K,$$

and $N^i(\boldsymbol{t}) - \Lambda^i(\boldsymbol{t})$ is a multiparameter martingale with respect to the filtration $\mathcal{F}^i$. Moreover,

$$E\left[\left(\int_{\mathbb{R}_+^K} \mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u}) \frac{H_\zeta(d\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}\right)^2\right] < \infty. \tag{4.1}$$

We next consider the compensator of multiparameter point processes $X^\ell(\boldsymbol{t})$, $\ell \in \mathbb{N}$, defined by

$$X^\ell(\boldsymbol{t}) := \sum_{i=1}^{\ell} N^i(\boldsymbol{t}) = \sum_{i=1}^{\ell} \mathbf{1}(\boldsymbol{\zeta}^i \le \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}_+^K.$$

The following result then follows from Lemma 4.3.

Lemma 4.4. *The compensator of the multiparameter point process X^ℓ, with respect to the filtration $\mathcal{F}^{X^\ell} := \{\mathcal{F}^{X^\ell}_{\boldsymbol{t}} : \boldsymbol{t} \in \mathbb{R}_+^K\}$, is given by*

$$Z^\ell(\boldsymbol{t}) := \int_{L_{\boldsymbol{t}}} \sum_{i=1}^{\ell} \frac{\mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(d\boldsymbol{u}), \quad \boldsymbol{t} \in \mathbb{R}_+^K,$$

where $\mathcal{F}^{X^\ell}_{\boldsymbol{t}} = \bigvee_{1 \le i \le \ell} \mathcal{F}^i_{\boldsymbol{t}}$. Moreover, for each $\boldsymbol{t} \in \mathbb{R}_+^K$, $E[(Z^\ell(\boldsymbol{t}))^2] < \infty$.

4.3. Proof of Theorem 3.1 In this section, we prove Theorem 3.1. We first introduce a decomposition of $\hat{U}^n$ in (3.1), which has two terms, one as a multiparameter martingale and the other as a process of finite variation. For an integrable function f defined on $\mathbb{R}_+^K$, we denote

$$\int_{L_{\boldsymbol{x}}} f(\boldsymbol{s}) d\boldsymbol{s} := \int_0^{x_1} \left(\dots \int_0^{x_K} f(s_1, \dots, s_K) ds_K \dots \right) ds_1, \quad \boldsymbol{x} \in \mathbb{R}_+^K.$$

Lemma 4.5 **(Decomposition of $\hat{U}^n$).** *For $\boldsymbol{x}:=(x_1,...,x_K)\in[0,1]^K$ and $t\geq 0$, we can decompose $\hat{U}^n$ in (3.1) as*

$$\hat{U}^n(t,\boldsymbol{x})=\hat{U}^n_0(t,\boldsymbol{x})+\hat{U}^n_1(t,\boldsymbol{x}), \tag{4.2}$$

where

$$\hat{U}^n_0(t,\boldsymbol{x}):=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\mathbf{1}(\boldsymbol{\xi}^i\leq\boldsymbol{x})-\int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})\right), \tag{4.3}$$

and

$$\hat{U}^n_1(t,\boldsymbol{x}):=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})-H(T_{\boldsymbol{u}})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u}). \tag{4.4}$$

For fixed $t\geq 0$, $\hat{U}^n_0(t,\boldsymbol{x})$ is a multiparameter martingale with respect to the smallest adapted filtration, and $\hat{U}^n_1(t,\boldsymbol{x})$ has finite variation as a multiparameter process of $\boldsymbol{x}$.

Proof. For $t\geq 0$ and $\boldsymbol{x}\in[0,1]^K$, by the definition of $\hat{U}^n(t,\boldsymbol{x})$ in (3.1),

$$\begin{aligned}\hat{U}^n(t,\boldsymbol{x})=&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\big(\mathbf{1}(\boldsymbol{\xi}^i\leq\boldsymbol{x})-H(\boldsymbol{x})\big)\\ =&\hat{U}^n_0(t,\boldsymbol{x})+\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\left(\int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})-H(\boldsymbol{x})\right)\\ =&\hat{U}^n_0(t,\boldsymbol{x})+\hat{U}^n_1(t,\boldsymbol{x}).\end{aligned}$$

Thus, (4.2) holds. The fact that $\hat{U}^n_0(t,\boldsymbol{x})$ is a multiparameter martingale follows from Lemma 4.4 for fixed $t\geq 0$. □

We remark that in the one-parameter case $K=1$, this decomposition is called a semimartingale decomposition for sequential empirical processes (Proposition II.3.36, [29]), and the second term $\hat{U}^n_1(t,x)$ is equal to

$$\hat{U}^n_1(t,x)=-\int_0^x\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}(\mathbf{1}(\xi^i\leq u)-H(u))\right)\frac{1}{H^c(u)}H(du)=-\int_0^x\frac{\hat{U}^n(t,u)}{H^c(u)}H(du), \tag{4.5}$$

for $t,x\geq 0$. That holds because for $K=1$, $\mathbf{1}(\xi^i>u)-H^c(u)=-\mathbf{1}(\xi^i\leq u)+H(u)$ for each $u\geq 0$. However, in the multiparameter setting, that equality does not hold, and thus, the second term $\hat{U}^n_1(t,\boldsymbol{x})$ in (4.4) cannot be written as such a simple form in (4.5). This decomposition for multiparameter sequential empirical processes is new, and it requires new methods to handle the convergence associated with the second term in the proof of Theorem 3.1 below.

Proof of Theorem 3.1. We now start proving the weak convergence of the multiparameter sequential empirical processes $\hat{U}^n(\cdot,\cdot)$. We take the standard approach by proving that the finite dimensional distributions of $\hat{U}^n(\cdot,\cdot)$ converge weakly to those of $U(\cdot,\cdot)$, and that the sequence of processes $\{\hat{U}^n:n\geq 1\}$ is tight in $\mathbb{D}([0,\infty),\mathbb{D}([0,1]^K,\mathbb{R}))$. Denote $\hat{U}^n(t):=\hat{U}^n(t,\cdot)$ and $U(t):=U(t,\cdot)$ for $t\in[0,\infty)$. As shown in Section 3.9 of [23], in order to prove the convergence of the finite dimensional distributions of $\hat{U}^n(\cdot,\cdot)$, it suffices to show that for any $\ell\in\mathbb{N}$ and $0\leq t_1<t_2<...<t_\ell$,

$$(\hat{U}^n(t_1),...,\hat{U}^n(t_\ell))\Rightarrow(U(t_1),...,U(t_\ell))\quad\text{in}\quad(\mathbb{D}([0,1]^K,\mathbb{R}))^\ell\quad\text{as}\quad n\to\infty. \tag{4.6}$$

This follows directly from Theorem 1 in [54]. Thus, it only remains to prove that the sequence of processes $\{\hat{U}^n:n\geq 1\}$ is tight in $\mathbb{D}([0,\infty),\mathbb{D}([0,1]^K,\mathbb{R}))$.

We first introduce the σ-fields

$$\begin{aligned}
&\mathcal{F}^i_{\boldsymbol{x}} := \sigma(\mathbf{1}(\boldsymbol{\xi}^i \le \boldsymbol{y}),\ 0 \le \boldsymbol{y} \le \boldsymbol{x}) \vee \mathcal{N}, \quad \boldsymbol{x} \in [0,1]^K,\\
&\mathcal{F}_{\boldsymbol{x}} := \bigvee_{i \ge 1} \mathcal{F}^i_{\boldsymbol{x}}, \quad \boldsymbol{x} \in [0,1]^K,\\
&\mathcal{G}^n_t := \bigvee_{i \le \lfloor nt \rfloor} \mathcal{F}^i_{\boldsymbol{e}},\ t \ge 0,
\end{aligned}$$

where $\mathcal{N}$ is the family of P-null sets and we recall $\boldsymbol{e} \in \mathbb{R}^K_+$ with all components 1. Let $\mathcal{G}^n := \{\mathcal{G}^n_t : t \ge 0\}$ and $\mathcal{F} := \{\mathcal{F}_{\boldsymbol{x}} : \boldsymbol{x} \in [0,1]^K\}$. The definitions of $\mathcal{G}^n$ and $\mathcal{F}$ with Lemma 4.2 imply that they are complete filtrations. Moreover, $\mathcal{G}^n$ satisfies the usual conditions (increasing and right-continuous families of complete σ-fields, see, e.g., Definition 2.25 of Chapter 1 in [32]). By (3.1), the sequence $\{\hat{U}^n(t) : t \ge 0\}$ is adapted to the filtration $\mathcal{G}^n$. From (4.6), for each $t \ge 0$, the sequence $\{\hat{U}^n(t) : n \ge 1\}$ is tight in $\mathbb{D}([0,1]^K, \mathbb{R})$.

Let $\tilde{d}_K := d_K \wedge 1$, where we recall d_K is the metric for the space $\mathbb{D}([0,1]^K, \mathbb{R})$. By the relative compactness criteria in [23] (see Theorem 8.6(a) and Remark 8.7(a), (b) of Chapter 3), it suffices to show that, for each $T > 0$, there exists a family of $\{\gamma^n(\delta) :\ 0 < \delta < 1, n \ge 1\}$ of nonnegative random variables satisfying

$$E[\tilde{d}_K(\hat{U}^n(t+u), \hat{U}^n(t)) | \mathcal{G}^n_t] \le E[\gamma^n(\delta) | \mathcal{G}^n_t], \quad a.s., \tag{4.7}$$

where $\gamma^n(\delta)$ satisfies

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E[\gamma^n(\delta)] = 0, \tag{4.8}$$

for $0 \le t \le T$ and $0 \le u \le \delta$.

By the definition of $\tilde{d}_K$, $\tilde{d}_K = (d_K \wedge 1) \le d_K$. Note that, for $v_1(\cdot), v_2(\cdot) \in \mathbb{D}([0,1]^K, \mathbb{R})$, $d_K(v_1(\cdot), v_2(\cdot)) \le \sup_{\boldsymbol{x} \in [0,1]^K} |v_1(\boldsymbol{x}) - v_2(\boldsymbol{x})|$. Instead of showing (4.7), it suffices to show

$$E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n(t+u, \boldsymbol{x}) - \hat{U}^n(t, \boldsymbol{x})| | \mathcal{G}^n_t\right] \le E[\gamma^n(\delta) | \mathcal{G}^n_t], \quad a.s., \tag{4.9}$$

for $0 \le t \le T$ and $0 \le u \le \delta$. By (3.1), $\hat{U}^n(t+u) - \hat{U}^n(t)$ is independent of $\mathcal{G}^n_t$. Thus, (4.9) is equivalent to

$$E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n(t+u, \boldsymbol{x}) - \hat{U}^n(t, \boldsymbol{x})|\right] \le E[\gamma^n(\delta) | \mathcal{G}^n_t], \quad a.s., \tag{4.10}$$

for $0 \le t \le T$ and $0 \le u \le \delta$.

We will first give a bound to the left hand side (LHS) of (4.10). By the decomposition property in Lemma 4.5,

$$\begin{aligned}
E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n(t+u, \boldsymbol{x}) - \hat{U}^n(t, \boldsymbol{x})|\right] &\le E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n_0(t+u, \boldsymbol{x}) - \hat{U}^n_0(t, \boldsymbol{x})|\right]\\
&\quad + E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n_1(t+u, \boldsymbol{x}) - \hat{U}^n_1(t, \boldsymbol{x})|\right],
\end{aligned} \tag{4.11}$$

where $\hat{U}^n_0(t, \boldsymbol{x})$ and $\hat{U}^n_1(t, \boldsymbol{x})$ are defined in (4.3) and (4.4), respectively. For the first term on the right hand side (RHS) of (4.11), by Jensen's inequality,

$$\begin{aligned}
E\left[\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n_0(t+u, \boldsymbol{x}) - \hat{U}^n_0(t, \boldsymbol{x})|\right] &\le \left(E\left[\left(\sup_{\boldsymbol{x} \in [0,1]^K} |\hat{U}^n_0(t+u, \boldsymbol{x}) - \hat{U}^n_0(t, \boldsymbol{x})|\right)^2\right]\right)^{\frac{1}{2}}\\
&\le \left(E\left[\sup_{\boldsymbol{x} \in [0,1]^K} (\hat{U}^n_0(t+u, \boldsymbol{x}) - \hat{U}^n_0(t, \boldsymbol{x}))^2\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

By Lemma 4.5, we note that, for the fixed t and u, $\hat{U}_0^n(t+u,\boldsymbol{x}) - \hat{U}_0^n(t,\boldsymbol{x})$ is a martingale with respect to the filtration $\mathcal{F}$. For the fixed t, since $\hat{U}^n(t,\boldsymbol{x})$ is right continuous in $\boldsymbol{x}$, $\hat{U}_0^n(t,\boldsymbol{x})$ is right continuous in $\boldsymbol{x}$. Thus, $E[(\hat{U}_0^n(t+u,\boldsymbol{x}) - \hat{U}_0^n(t,\boldsymbol{x}))^2]$ is right continuous in $\boldsymbol{x}$ for the fixed t and u. By the construction of the filtration and Definition 4.4, it is easy to check that $\mathcal{F}$ is commuting. Thus, by Cairoli's Strong (p,p) inequality in Lemma 4.1 and setting $p=2$,

$$\begin{aligned}
E\left[\sup_{\boldsymbol{x}\in[0,1]^K} (\hat{U}_0^n(t+u,\boldsymbol{x}) - \hat{U}_0^n(t,\boldsymbol{x}))^2\right] &\le 4^K E\left[\left(\hat{U}_0^n(t+u,\boldsymbol{e}) - \hat{U}_0^n(t,\boldsymbol{e})\right)^2\right] \\
&= 4^K E\left[\frac{1}{n}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+u)\rfloor}\left(1 - \int_{L_{\boldsymbol{e}}} \frac{\mathbf{1}(\boldsymbol{\xi}^i > \boldsymbol{u})}{H(T_{\boldsymbol{u}})} H(d\boldsymbol{u})\right)^2\right] \\
&= 4^K C_1\left(\frac{\lfloor n(t+u)\rfloor - \lfloor nt\rfloor}{n}\right) \\
&\le 4^K C_1\left(\delta + \frac{1}{n}\right),
\end{aligned} \tag{4.12}$$

where

$$C_1 := E\left[\left(1 - \int_{L_{\boldsymbol{e}}} \frac{\mathbf{1}(\boldsymbol{\xi}^i > \boldsymbol{u})}{H(T_{\boldsymbol{u}})} H(d\boldsymbol{u})\right)^2\right]$$

is a constant. By Lemma 4.3, we see that C_1 is finite. By (4.12), we have

$$E\left[\sup_{\boldsymbol{x}\in[0,1]^K} |\hat{U}_0^n(t+u,\boldsymbol{x}) - \hat{U}_0^n(t,\boldsymbol{x})|\right] \le \gamma_1^n(\delta) := 2^K\sqrt{C_1}\left(\delta + \frac{1}{n}\right)^{\frac{1}{2}}. \tag{4.13}$$

We now consider the second term on the RHS of (4.11). Denote

$$\sigma_{\boldsymbol{x}}^2 := Var\left(\int_{L_{\boldsymbol{x}}} \frac{\mathbf{1}(\boldsymbol{\xi}^i > \boldsymbol{u}) - H(T_{\boldsymbol{u}})}{H(T_{\boldsymbol{u}})} H(d\boldsymbol{u})\right).$$

Note that, for $\boldsymbol{x}\in[0,1]^K$,

$$\sigma_{\boldsymbol{x}}^2 \le 2\left(E\left[\left(\int_{L_{\boldsymbol{x}}} \frac{\mathbf{1}(\boldsymbol{\xi}^i > \boldsymbol{u})}{H(T_{\boldsymbol{u}})} H(d\boldsymbol{u})\right)^2\right] + H(\boldsymbol{x})\right) \le 2\left(E\left[\left(\int_{L_{\boldsymbol{e}}} \frac{\mathbf{1}(\boldsymbol{\xi}^i > \boldsymbol{u})}{H(T_{\boldsymbol{u}})} H(d\boldsymbol{u})\right)^2\right] + 1\right). \tag{4.14}$$

Recall $\boldsymbol{e}$ is the vector in $\mathbb{R}^K$ with all components 1. By Lemma 4.3, we see that the RHS of (4.14) is finite. Thus, $\sup_{\boldsymbol{x}\in[0,1]^K}\sigma_{\boldsymbol{x}}^2$ is finite, which implies $\sup_{\boldsymbol{x}\in[0,1]^K}\sigma_{\boldsymbol{x}}$ is finite. By strong approximation of random walks by Brownian motion (see section 3.5 in [39]), $\exists\ n_0$ and c_1, when $n\ge n_0$, we have

$$\sup_{\boldsymbol{x}\in[0,1]^K}\ \sup_{0\le t\le T+\delta}\left|\hat{U}_1^n(t,\boldsymbol{x}) - \sigma_{\boldsymbol{x}}B(t)\right| \le c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n, \quad a.s., \tag{4.15}$$

where $B(t)$ is a standard Brownian motion. Denote $c_2 := \sup_{\boldsymbol{x}\in[0,1]^K}\sigma_{\boldsymbol{x}}$. Then, for $0\le t\le T$, $0\le u\le\delta$, and $n\ge n_0$,

$$\begin{aligned}
&E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\left|\hat{U}_1^n(t+u,\boldsymbol{x}) - \hat{U}_1^n(t,\boldsymbol{x})\right|\right] \\
&\le E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\left|\hat{U}_1^n(t+u,\boldsymbol{x}) - \sigma_{\boldsymbol{x}}B(t+u)\right|\right] + E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\sigma_{\boldsymbol{x}}\left|B(t+u) - B(t)\right|\right]
\end{aligned}$$

$$
\begin{aligned}
&+E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\left|\sigma_{\boldsymbol{x}}B(t)-\hat{U}_1^n(t,\boldsymbol{x})\right|\right]\\
&\le 2c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n + c_2 E\left[|B(u)|\right]\\
&\le 2c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n + c_2 \left(E\left[(B(u))^2\right]\right)^{\frac{1}{2}}\\
&\le 2c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n + c_2 u^{\frac{1}{2}}\\
&\le 2c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n + c_2 \delta^{\frac{1}{2}}.
\end{aligned}
\tag{4.16}
$$

For $n < n_0$, $0 \le t \le T$ and $0 \le u \le \delta$,

$$
\begin{aligned}
&E\left[\sup_{\boldsymbol{x}\in[0,1]^K}|\hat{U}_1^n(t+u,\boldsymbol{x})-\hat{U}_1^n(t,\boldsymbol{x})|\right]\\
&\le E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\left|\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+u)\rfloor}\int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})-H(T_{\boldsymbol{u}})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})\right|\right]\\
&\le (\lfloor n(t+\delta)\rfloor-\lfloor nt\rfloor)E\left[\sup_{\boldsymbol{x}\in[0,1]^K}\left|\int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})\right|+\sup_{\boldsymbol{x}\in[0,1]^K}H(\boldsymbol{x})\right]\\
&=(\lfloor n(t+\delta)\rfloor-\lfloor nt\rfloor)\left(E\left[\int_{L_{\boldsymbol{e}}}\frac{\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})\right]+1\right)\\
&=(\lfloor n(t+\delta)\rfloor-\lfloor nt\rfloor)\left(\int_{L_{\boldsymbol{e}}}\frac{E[\mathbf{1}(\boldsymbol{\xi}^i>\boldsymbol{u})]}{H(T_{\boldsymbol{u}})}H(d\boldsymbol{u})+1\right) \qquad \text{(By Fubini's theorem)}\\
&=2(\lfloor n(t+\delta)\rfloor-\lfloor nt\rfloor)\\
&\le 2(n\delta+1).
\end{aligned}
\tag{4.17}
$$

Define $\gamma_2^n(\delta)$ by

$$
\gamma_2^n(\delta)=\begin{cases}2c_1 n^{-\frac{1}{4}}\log^{\frac{3}{2}} n + c_2\delta^{\frac{1}{2}}, & \text{if } n\ge n_0,\\ 2(n\delta+1), & \text{if } n<n_0.\end{cases}
\tag{4.18}
$$

Thus, by (4.16) and (4.17), we have obtained an upper bound for the second term on the RHS of (4.11):

$$
E\left[\sup_{\boldsymbol{x}\in[0,1]^K}|\hat{U}_1^n(t+u,\boldsymbol{x})-\hat{U}_1^n(t,\boldsymbol{x})|\right]\le\gamma_2^n(\delta).
\tag{4.19}
$$

Let $\gamma^n(\delta):=\gamma_1^n(\delta)+\gamma_2^n(\delta)$. From (4.11), (4.13) and (4.19), we see that

$$
E\left[\sup_{\boldsymbol{x}\in[0,1]^K}|\hat{U}^n(t+u,\boldsymbol{x})-\hat{U}^n(t,\boldsymbol{x})|\right]\le\gamma^n(\delta)=E[\gamma^n(\delta)|\mathcal{G}_t],
$$

and

$$
\lim_{\delta\to 0}\limsup_{n\to\infty}E[\gamma^n(\delta)|\mathcal{G}_t^n]=\lim_{\delta\to 0}\limsup_{n\to\infty}\gamma^n(\delta)=0.
$$

Thus, we have verified (4.7) and (4.8), and shown the weak convergence of $\{\hat{U}^n : n\ge 1\}$. □

5. Proofs for the Characterization of the Limit Processes In this section, we prove the Gaussian characterizations of the limiting processes, Theorem 3.4. We first introduce some notations. For a set $\mathcal{J}$, let $|\mathcal{J}|$ be the cardinality of $\mathcal{J}$. Let $\mathcal{J}_k^1$ and $\mathcal{J}_{N-k}^2$ be the partition of $\mathcal{A}:=\{1,...,N\}$, where N is a positive integer, $\mathcal{J}_k^1\cap\mathcal{J}_{N-k}^2=\emptyset$, $|\mathcal{J}_k^1|=k$ and $|\mathcal{J}_{N-k}^2|=N-k$. Note that $\mathcal{J}_0^1 = \mathcal{J}_0^2 = \emptyset$. Let $\Phi : \mathbb{R}^N \to \mathbb{R}$. For $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^N$ and $\boldsymbol{x} \le \boldsymbol{y}$, define $\Phi^{\mathcal{J}_k^1, \mathcal{J}_{N-k}^2}(\boldsymbol{x}; \boldsymbol{y}) := \Phi(\boldsymbol{z})$, where $z_j = x_j$ for $j \in \mathcal{J}_k^1$ and $z_j = y_j$ for $j \in \mathcal{J}_{N-k}^2$. Then, we define

$$\Delta\Phi(\boldsymbol{x};\boldsymbol{y}) := \sum_{k=0}^{N} (-1)^k \sum_{\mathcal{J}_k^1, \mathcal{J}_{N-k}^2 \text{ partitions of } \mathcal{A}} \Phi^{\mathcal{J}_k^1, \mathcal{J}_{N-k}^2}(\boldsymbol{x};\boldsymbol{y}). \tag{5.1}$$

This notion "Δ" can be interpreted as the following: for a real-valued function Φ defined on $\mathbb{R}^N$, $\Delta\Phi(\boldsymbol{x};\boldsymbol{y})$ represents the *increment* of Φ between $\boldsymbol{x}$ and $\boldsymbol{y}$ for each $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ satisfying $\boldsymbol{x} \le \boldsymbol{y}$. For $N = 1$, $\Delta\Phi(x;y) = \Phi(y) - \Phi(x)$ for $x \le y$. For $N = 2$, $\Delta\Phi(\boldsymbol{x};\boldsymbol{y}) = \Phi(y_1, y_2) - \Phi(x_1, y_2) - \Phi(x_2, y_1) + \Phi(x_1, x_2)$ for $\boldsymbol{x} = (x_1, x_2) \le \boldsymbol{y} = (y_1, y_2)$. In the following proofs, we will use $\Delta\hat{K}(\boldsymbol{x};\boldsymbol{y})$ as defined in (5.1) for $N = K + 1$, and $\Delta F(\boldsymbol{x};\boldsymbol{y})$ as defined in (5.1) for $N = K$.

For $k = 1, ..., K$, define

$$\hat{K}_k(\bar{a}(t), y) := \hat{K}(\bar{a}(t), \boldsymbol{x}),$$

where $\boldsymbol{x} \in \mathbb{R}_+^K$ satisfies $x_k = y$ and $x_j = \infty$ for $j \ne k$, $j = 1, ..., K$. Define

$$\hat{K}_{\mathrm{m}}(\bar{a}(t), y) := \hat{K}(\bar{a}(t), \boldsymbol{x}),$$

where $\boldsymbol{x} \in \mathbb{R}_+^K$ satisfies $x_k = y$ for $k = 1, ..., K$.

We first provide the definitions of the processes $\hat{\boldsymbol{M}}_2$ in (3.18), $\hat{\boldsymbol{Z}}_2$ in (3.24) and $\hat{V}_2$ in (3.21).

Definition 5.1. For $k = 1, ..., K$, the processes $\hat{M}_{k,2}$ in (3.18), $\hat{V}_2$ in (3.21) and $\hat{Z}_{k,2}$ in (3.24) are defined as mean-square integrals, i.e., for each $t \ge 0$,

$$\lim_{\ell\to\infty} E[(\hat{M}_{k,2}(t) - \hat{M}_{k,2,\ell}(t))^2] = 0, \tag{5.2}$$
$$\lim_{\ell\to\infty} E[(\hat{V}_2(t) - \hat{V}_{2,\ell}(t))^2] = 0, \tag{5.3}$$
$$\lim_{\ell\to\infty} E[(\hat{Z}_{k,2}(t) - \hat{Z}_{k,2,\ell}(t))^2] = 0, \tag{5.4}$$

where

$$\hat{M}_{k,2,\ell}(t) := -\int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}_{k,\ell,t}(s, \boldsymbol{x}) d\hat{K}(\bar{a}(s), \boldsymbol{x}), \tag{5.5}$$
$$\hat{V}_{2,\ell}(t) := \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}_{\mathrm{m},\ell,t}(s, \boldsymbol{x}) d\hat{K}(\bar{a}(s), \boldsymbol{x}), \tag{5.6}$$
$$\hat{Z}_{k,2,\ell}(t) := -\hat{M}_{k,2,\ell}(t) - \hat{V}_{2,\ell}(t), \tag{5.7}$$

and

$$\mathbf{1}_{k,\ell,t}(s, \boldsymbol{x}) := \sum_{i=1}^{\ell} \mathbf{1}(s_{i-1}^\ell < s \le s_i^\ell) \mathbf{1}(x_k \le t - s_i^\ell), \tag{5.8}$$
$$\mathbf{1}_{\mathrm{m},\ell,t}(s, \boldsymbol{x}) := \sum_{i=1}^{\ell} \mathbf{1}(s_{i-1}^\ell < s \le s_i^\ell) \mathbf{1}(x_j \le t - s_i^\ell, \ \forall j = 1, ..., K), \tag{5.9}$$

with $0 = s_0^\ell < s_1^\ell < ... < s_\ell^\ell = t$ and $\max_{1 \le i \le \ell} |s_i^\ell - s_{i-1}^\ell| \to 0$ as $\ell \to \infty$. We call $\{s_i^\ell : 0 \le i \le \ell\}$ is a partition of $[0, t]$.

We next show the well-definedness and Gaussian property for the process $(\hat{\boldsymbol{M}}_2, \hat{\boldsymbol{Z}}_2, \hat{V}_2)$ in (3.18), (3.24) and (3.21).

LEMMA 5.1. *The process* $(\hat{\boldsymbol{M}}_2, \hat{\boldsymbol{Z}}_2, \hat{V}_2)$ *in* (3.18), (3.24) *and* (3.21) *is a well-defined continuous Gaussian process with mean* $\mathbf{0}$*, and covariance functions: for each* $t \ge 0$ *and* $j,k = 1,...,K$,

$$Cov(\hat{M}_{j,2}(t), \hat{M}_{k,2}(t)) = \int_0^t [F_{j,k}(t-s,t-s) - F_j(t-s)F_k(t-s)]d\bar{a}(s), \tag{5.10}$$

$$\begin{aligned} Cov(\hat{Z}_{j,2}(t), \hat{Z}_{k,2}(t)) = \int_0^t \Big[&F_{j,k}(t-s,t-s) - F_j(t-s)F_k(t-s) - F_{\mathrm{m}}(t-s) \\ &+ F_k(t-s)F_{\mathrm{m}}(t-s) + F_{\mathrm{m}}(t-s)F_j(t-s) - (F_{\mathrm{m}}(t-s))^2\Big] d\bar{a}(s), \end{aligned} \tag{5.11}$$

$$\begin{aligned} Cov(\hat{M}_{j,2}(t), \hat{Z}_{k,2}(t)) = \int_0^t &[F_{\mathrm{m}}(t-s) - F_{j,k}(t-s,t-s) - F_j(t-s)F_{\mathrm{m}}(t-s) \\ &+F_j(t-s)F_k(t-s)]\, d\bar{a}(s), \end{aligned} \tag{5.12}$$

$$Cov(\hat{M}_{k,2}(t), \hat{V}_2(t)) = -\int_0^t [F_k^c(t-s)F_{\mathrm{m}}(t-s)]\, d\bar{a}(s), \tag{5.13}$$

$$Cov(\hat{Z}_{k,2}(t), \hat{V}_2(t)) = \int_0^t [F_{\mathrm{m}}(t-s)(F_{\mathrm{m}}(t-s) - F_k(t-s))]\, d\bar{a}(s), \tag{5.14}$$

$$Var(\hat{V}_2(t)) = \int_0^t [F_{\mathrm{m}}(t-s)F_{\mathrm{m}}^c(t-s)]\, d\bar{a}(s). \tag{5.15}$$

Proof. We first show that $(\hat{\boldsymbol{M}}_2, \hat{\boldsymbol{Z}}_2, \hat{V}_2)$ is a well-defined continuous Gaussian process. By (3.18), (3.21) and (3.24), it suffices to show $\hat{M}_{k,2}(\cdot)$ and $\hat{V}_2(\cdot)$ are well-defined continuous Gaussian processes, $k = 1,...,K$.

Recall Definition 5.1. The processes $\hat{M}_{k,2,\ell}(t)$ in (5.5) and $\hat{V}_{2,\ell}(t)$ in (5.6) can be written as

$$\begin{aligned} \hat{M}_{k,2,\ell}(t) &= -\sum_{i=1}^{\ell} \Delta\hat{K}\left((\bar{a}(s_{i-1}^{\ell}), \mathbf{0}); (\bar{a}(s_i^{\ell}), \boldsymbol{x}^i)\right) \\ &= -\sum_{i=1}^{\ell} \left[\hat{K}_k(\bar{a}(s_i^{\ell}), t-s_i^{\ell}) - \hat{K}_k(\bar{a}(s_{i-1}^{\ell}), t-s_i^{\ell})\right], \quad t \ge 0, \end{aligned} \tag{5.16}$$

$$\begin{aligned} \hat{V}_{2,\ell}(t) &= \sum_{i=1}^{\ell} \Delta\hat{K}\left((\bar{a}(s_{i-1}^{\ell}), \mathbf{0}); (\bar{a}(s_i^{\ell}), \boldsymbol{y}^i)\right) \\ &= \sum_{i=1}^{\ell} \left[\hat{K}_{\mathrm{m}}(\bar{a}(s_i^{\ell}), t-s_i^{\ell}) - \hat{K}_{\mathrm{m}}(\bar{a}(s_{i-1}^{\ell}), t-s_i^{\ell})\right], \quad t \ge 0, \end{aligned} \tag{5.17}$$

where $\boldsymbol{x}^i \in \mathbb{R}^K$ with $x_j^i = \infty$ for $j \ne k$ and $x_k^i = t - s_i^{\ell}$, and $\boldsymbol{y}^i \in \mathbb{R}^K$ with $y_j^i = t - s_i^{\ell}$ for $1 \le j \le K$.

To show $\hat{M}_{k,2}(\cdot)$ and $\hat{V}_2(\cdot)$ are well-defined, $k = 1,...,K$, we need to prove, for each $t \ge 0$,

$$\lim_{l,\ell\to\infty} E\left[(\hat{M}_{k,2,l}(t) - \hat{M}_{k,2,\ell}(t))^2\right] = 0, \tag{5.18}$$

$$\lim_{l,\ell\to\infty} E\left[(\hat{V}_{2,l}(t) - \hat{V}_{2,\ell}(t))^2\right] = 0, \tag{5.19}$$

where we define $\hat{M}_{k,2,l}(t)$, $\hat{V}_{2,l}(t)$ and their associated partition $\{s_i^l : 0 \le i \le l\}$ of $[0,t]$ similarly as $\hat{M}_{k,2,\ell}(t)$ in (5.5), $\hat{V}_{2,\ell}(t)$ in (5.6) and the partition $\{s_i^{\ell} : 0 \le i \le \ell\}$ of $[0,t]$ in Definition 5.1 for each $t \ge 0$, respectively. We focus on the details of showing (5.19), as a similar argument can be applied to the proof of (5.18). Recall $\boldsymbol{e} \in \mathbb{R}^K$ with all entries 1. Without loss of generality, we assume that the partition $\{s_i^{\ell} : 0 \le i \le \ell\}$ of $[0,t]$ is finer than the partition $\{s_i^l : 0 \le i \le l\}$ of $[0,t]$. By (5.17), we have

$$\hat{V}_{2,l}(t) - \hat{V}_{2,\ell}(t) = \sum_{i=1}^{l} \sum_{j: s_{i-1}^l < s_j^{\ell} \le s_i^l} \Delta\hat{K}\left((\bar{a}(s_{j-1}^{\ell}), (t-s_i^l)\boldsymbol{e}); (\bar{a}(s_j^{\ell}), (t-s_j^{\ell})\boldsymbol{e})\right), \quad t \ge 0. \tag{5.20}$$

We first claim that, for $0 \le t_1 \le t_2 \le t_1' \le t_2'$, $0 \le x \le y$ and $0 \le x' \le y'$,

$$E\left[\left(\Delta \hat{K}\left((\bar{a}(t_1), x\boldsymbol{e}); (\bar{a}(t_2), y\boldsymbol{e})\right)\right)^2\right] = (\bar{a}(t_2) - \bar{a}(t_1))(\Delta F(x\boldsymbol{e}; y\boldsymbol{e}))(1 - \Delta F(x\boldsymbol{e}; y\boldsymbol{e})), \tag{5.21}$$

$$E\left[\Delta \hat{K}((\bar{a}(t_1), x\boldsymbol{e}); (\bar{a}(t_2), y\boldsymbol{e}))\Delta \hat{K}((\bar{a}(t_1'), x'\boldsymbol{e}); (\bar{a}(t_2'), y'\boldsymbol{e}))\right] = 0. \tag{5.22}$$

For conciseness, we only show (5.21) and (5.22) with $K = 2$. The case $K > 2$ can be easily generalized. Note that for $0 \le t_1 \le t_2$ and $0 \le x \le y$,

$$\begin{aligned}
&E\left[\left(\Delta \hat{K}\left((\bar{a}(t_1), x, x); (\bar{a}(t_2), y, y)\right)\right)^2\right] \\
&= E\Big[\Big(\hat{K}(\bar{a}(t_2), y, y) - \hat{K}(\bar{a}(t_1), y, y) - \hat{K}(\bar{a}(t_2), x, y) - \hat{K}(\bar{a}(t_2), y, x) \\
&\qquad + \hat{K}(\bar{a}(t_1), x, y) + \hat{K}(\bar{a}(t_1), y, x) + \hat{K}(\bar{a}(t_2), x, x) - \hat{K}(\bar{a}(t_1), x, x)\Big)^2\Big] \\
&= (\bar{a}(t_2) - \bar{a}(t_1))(F(y, y) - F(x, y) - F(y, x) + F(x, x)) \\
&\qquad \times (1 - (F(y, y) - F(x, y) - F(y, x) + F(x, x))) \\
&= (\bar{a}(t_2) - \bar{a}(t_1))(\Delta F((x, x); (y, y)))(1 - \Delta F((x, x); (y, y))).
\end{aligned}$$

Similarly, if $0 \le t_1 \le t_2 \le t_1' \le t_2'$, $0 \le x \le y$ and $0 \le x' \le y'$, then

$$E\left[\Delta \hat{K}((\bar{a}(t_1), x, x); (\bar{a}(t_2), y, y))\Delta \hat{K}((\bar{a}(t_1'), x', x'); (\bar{a}(t_2'), y', y'))\right] = 0.$$

Thus, we have shown (5.21) and (5.22). By these two equations and (5.20), we have

$$\begin{aligned}
&E[(\hat{V}_{2,l}(t) - \hat{V}_{2,\ell}(t))^2] \\
&= \sum_{i=1}^{l} \sum_{j: s_{i-1}^l < s_j^\ell \le s_i^l} (\bar{a}(s_j^\ell) - \bar{a}(s_{j-1}^\ell))(\Delta F((t - s_i^l)\boldsymbol{e}; (t - s_j^\ell)\boldsymbol{e}))(1 - \Delta F((t - s_i^l)\boldsymbol{e}; (t - s_j^\ell)\boldsymbol{e})) \\
&\le \sum_{i=1}^{l} \sum_{j: s_{i-1}^l < s_j^\ell \le s_i^l} (\bar{a}(s_j^\ell) - \bar{a}(s_{j-1}^\ell))(\Delta F((t - s_i^l)\boldsymbol{e}; (t - s_j^\ell)\boldsymbol{e})) \\
&\le \sum_{i=1}^{l} (\bar{a}(s_i^l) - \bar{a}(s_{i-1}^l))(\Delta F((t - s_i^l)\boldsymbol{e}; (t - s_{i-1}^l)\boldsymbol{e})) \\
&\le \max_{1 \le i \le l} (\bar{a}(s_i^l) - \bar{a}(s_{i-1}^l)).
\end{aligned}$$

Since $\bar{a}(\cdot)$ is continuous and $\max_{1 \le i \le l}(\bar{a}(s_i^l) - \bar{a}(s_{i-1}^l)) \to 0$ as $l \to \infty$, we have proved (5.19).

Since $\hat{K}(\cdot, \cdot)$ is Gaussian with mean 0, for a fixed $t \ge 0$, $\hat{M}_{k,2,\ell}(t)$ and $\hat{V}_{2,\ell}(t)$ are normally distributed with mean 0, $k = 1, ..., K$. By the definitions of $\hat{M}_{k,2}(t)$ in (5.2) (respectively, $\hat{V}_2(t)$ in (5.3)), $\hat{M}_{k,2,\ell}(t)$ (respectively, $\hat{V}_{2,\ell}(t)$) converges to $\hat{M}_{k,2}(t)$ (respectively, $\hat{V}_2(t)$) in probability as $\ell \to \infty$, for each $t \ge 0$ and $k = 1, ..., K$. Recall the fact that if a sequence of normally distributed random variables in probability converges to a random variable, the limit is also normally distributed (see Lemma 4.9.4 in [43]). Thus, $\hat{M}_{k,2}(t)$ and $\hat{V}_2(t)$ are normally distributed and their means are 0, for each $t \ge 0$ and $k = 1, ..., K$.

Next, we show that $\hat{M}_{k,2}(t)$ and $\hat{V}_2(t)$ are continuous in t, $k = 1, ..., K$. Again, we focus on the proof of the continuity of $\hat{V}_2(t)$. We assume that the same partition $\{s_i^\ell : 0 \le i \le \ell\}$ of $[0, t]$ is applied for $\hat{V}_{2,\ell}(t)$ and $\hat{V}_{2,\ell}(s)$ for $0 \le s \le t$. By (5.17), we have

$$\hat{V}_{2,\ell}(t) - \hat{V}_{2,\ell}(s) = \sum_{i=1}^{\ell} \Delta \hat{K}\left((\bar{a}(s_{i-1}^\ell), (s - s_i^\ell)\boldsymbol{e}); (\bar{a}(s_i^\ell), (t - s_i^\ell)\boldsymbol{e})\right),$$

where we note that $\hat{K}(t,x)=0$ for $t\geq 0$ and $x\leq 0$. With (5.21) and (5.22), we obtain that

$$\begin{aligned} &E[(\hat{V}_{2,\ell}(t)-\hat{V}_{2,\ell}(s))^2] \\ &=\sum_{i=1}^{\ell}(\bar{a}(s_i^\ell)-\bar{a}(s_{i-1}^\ell))(\Delta F((s-s_i^\ell)\boldsymbol{e};(t-s_i^\ell)\boldsymbol{e}))(1-\Delta F((s-s_i^\ell)\boldsymbol{e};(t-s_i^\ell)\boldsymbol{e})). \end{aligned} \tag{5.23}$$

By Lebesgue's theorem, we then have

$$\begin{aligned} \lim_{\ell\to\infty} E[(\hat{V}_{2,\ell}(t)-\hat{V}_{2,\ell}(s))^2] = \int_0^t \Big[&(\Delta F((s-u)\boldsymbol{e};(t-u)\boldsymbol{e})) \\ &\times (1-\Delta F((s-u)\boldsymbol{e};(t-u)\boldsymbol{e}))\Big] d\bar{a}(u). \end{aligned} \tag{5.24}$$

By the definition of $\hat{V}_2(t)$ in (5.3), and the fact that $\hat{V}_{2,\ell}(t)$ is normally distributed for each $t\geq 0$, Lemma 4.9.4 of [43] implies

$$E[(\hat{V}_2(t)-\hat{V}_2(s))^2] = \lim_{\ell\to\infty} E[(\hat{V}_{2,\ell}(t)-\hat{V}_{2,\ell}(s))^2]. \tag{5.25}$$

Thus, with (5.24) and (5.25), we see that

$$E[(\hat{V}_2(t)-\hat{V}_2(s))^2] = \int_0^t \Big[(\Delta F((s-u)\boldsymbol{e};(t-u)\boldsymbol{e}))(1-\Delta F((s-u)\boldsymbol{e};(t-u)\boldsymbol{e}))\Big] d\bar{a}(u). \tag{5.26}$$

From (5.26), we obtain that $\hat{V}_2(\cdot)$ is continuous in probability. By Lemma 4.9.6 in [43], to show $\hat{V}_2(\cdot)$ has continuous sample paths almost surely, it is sufficient to show, for any partition $\{s_i^\ell : 0\leq i\leq \ell\}$ of $[0,t]$,

$$\lim_{L\to\infty}\limsup_{\ell\to\infty} P\left(\sum_{i=1}^{\ell}(\hat{V}_2(s_i^\ell)-\hat{V}_2(s_{i-1}^\ell))^2\geq L\right)=0. \tag{5.27}$$

By Markov inequality and (5.26), we note that

$$\begin{aligned} &P\left(\sum_{i=1}^{\ell}(\hat{V}_2(s_i^\ell)-\hat{V}_2(s_{i-1}^\ell))^2\geq L\right) \\ &\leq \frac{1}{L}\sum_{i=1}^{\ell} E[(\hat{V}_2(s_i^\ell)-\hat{V}_2(s_{i-1}^\ell))^2] \\ &= \frac{1}{L}\sum_{i=1}^{\ell}\int_0^t \Big[(\Delta F((s_{i-1}^\ell-u)\boldsymbol{e};(s_i^\ell-u)\boldsymbol{e}))(1-\Delta F((s_{i-1}^\ell-u)\boldsymbol{e};(s_i^\ell-u)\boldsymbol{e}))\Big] d\bar{a}(u) \\ &\leq \frac{1}{L}\sum_{i=1}^{\ell}\int_0^t \Big[\Delta F((s_{i-1}^\ell-u)\boldsymbol{e};(s_i^\ell-u)\boldsymbol{e})\Big] d\bar{a}(u) \\ &\leq \frac{1}{L}\bar{a}(t). \end{aligned}$$

The last inequality above follows from the fact that $\sum_{i=1}^{\ell}\Delta F((s_{i-1}^\ell-u)\boldsymbol{e};(s_i^\ell-u)\boldsymbol{e})\leq F_{\mathrm{m}}(t)\leq 1$. Thus, (5.27) holds, which implies that $\hat{V}_2(\cdot)$ is a continuous process. A similar argument shows that $\hat{M}_{k,2}(\cdot)$ is also a continuous process, $k=1,...,K$. In summary, we have shown that $\hat{M}_{k,2}(\cdot)$ and $\hat{V}_2(\cdot)$ are well-defined continuous Gaussian processes, and thus, so are $\hat{Z}_{k,2}(\cdot)$, $k=1,...,K$.

Now we show the covariance of $(\hat{\boldsymbol{M}}_2,\hat{\boldsymbol{Z}}_2,\hat{V}_2)$. We here only focus on the covariance of $\hat{Z}_{j,2}(t)$ and $\hat{Z}_{k,2}(t)$ in (5.11), as other covariance functions in (5.10) and (5.12)-(5.14) and the variance of $\hat{V}_2(t)$ in (5.15) follow from a similar argument, for each $t\geq 0$ and $j,k=1,...,K$. For a fixed $t\geq 0$

and $j,k=1,...,K$, without loss of generality, we assume that $\hat{Z}_{j,2,\ell}(t)$ and $\hat{Z}_{k,2,\ell}(t)$ have the same partition $\{s_i^\ell : 1\le i\le \ell\}$ for the interval $[0,t]$. First, by Cauchy-Schwarz inequality, we have, for $t\ge 0$ and $j,k=1,...,K$,

$$\begin{aligned}
&\left|Cov(\hat{Z}_{j,2,\ell}(t),\hat{Z}_{k,2,\ell}(t)) - Cov(\hat{Z}_{j,2}(t),\hat{Z}_{k,2}(t))\right| \\
&= \left|E[\hat{Z}_{j,2,\ell}(t)\hat{Z}_{k,2,\ell}(t)] - E[\hat{Z}_{j,2}(t)\hat{Z}_{k,2,\ell}(t)] + E[\hat{Z}_{j,2}(t)\hat{Z}_{k,2,\ell}(t)] - E[\hat{Z}_{j,2}(t)\hat{Z}_{k,2}(t)]\right| \\
&\le \left|E[(\hat{Z}_{j,2,\ell}(t)-\hat{Z}_{j,2}(t))\hat{Z}_{k,2,\ell}(t)]\right| + \left|E[\hat{Z}_{j,2}(t)(\hat{Z}_{k,2,\ell}(t)-\hat{Z}_{k,2}(t))]\right| \\
&\le \Big(E[(\hat{Z}_{j,2,\ell}(t)-\hat{Z}_{j,2}(t))^2]\Big)^{\frac{1}{2}}\Big(E[(\hat{Z}_{k,2,\ell}(t))^2]\Big)^{\frac{1}{2}} + \Big(E[(\hat{Z}_{j,2}(t))^2]\Big)^{\frac{1}{2}}\Big(E[(\hat{Z}_{k,2,\ell}(t)-\hat{Z}_{k,2}(t))^2]\Big)^{\frac{1}{2}}.
\end{aligned} \tag{5.28}$$

For each $t\ge 0$ and $j,k=1,...,K$, since $E[(\hat{Z}_{j,2,\ell}(t)-\hat{Z}_{j,2}(t))^2]\to 0$ and $E[(\hat{Z}_{k,2,\ell}(t)-\hat{Z}_{k,2}(t))^2]\to 0$ as $\ell\to\infty$, by (5.28), we have

$$\lim_{\ell\to\infty} Cov(\hat{Z}_{j,2,\ell}(t),\hat{Z}_{k,2,\ell}(t)) = Cov(\hat{Z}_{j,2}(t),\hat{Z}_{k,2}(t)). \tag{5.29}$$

Note that

$$\begin{aligned}
&Cov(\hat{Z}_{j,2,\ell}(t),\hat{Z}_{k,2,\ell}(t)) \\
&= Cov\Big(\sum_{i=1}^{\ell}\big[\hat{K}_j(\bar{a}(s_i^\ell),t-s_i^\ell) - \hat{K}_j(\bar{a}(s_{i-1}^\ell),t-s_i^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_i^\ell),t-s_i^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{i-1}^\ell),t-s_i^\ell)\big], \\
&\qquad \sum_{l=1}^{\ell}\big[\hat{K}_k(\bar{a}(s_l^\ell),t-s_l^\ell) - \hat{K}_k(\bar{a}(s_{l-1}^\ell),t-s_l^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_l^\ell),t-s_l^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{l-1}^\ell),t-s_l^\ell)\big]\Big) \\
&= \sum_{i=1}^{\ell}\sum_{l=1}^{\ell} Cov\Big(\big[\hat{K}_j(\bar{a}(s_i^\ell),t-s_i^\ell) - \hat{K}_j(\bar{a}(s_{i-1}^\ell),t-s_i^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_i^\ell),t-s_i^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{i-1}^\ell),t-s_i^\ell)\big], \\
&\qquad \big[\hat{K}_k(\bar{a}(s_l^\ell),t-s_l^\ell) - \hat{K}_k(\bar{a}(s_{l-1}^\ell),t-s_l^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_l^\ell),t-s_l^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{l-1}^\ell),t-s_l^\ell)\big]\Big).
\end{aligned}$$

By some simple calculations, if $s_i^\ell \le s_{l-1}^\ell$ and $s_l^\ell\le s_{i-1}^\ell$,

$$\begin{aligned}
&Cov\Big(\big[\hat{K}_j(\bar{a}(s_i^\ell),t-s_i^\ell) - \hat{K}_j(\bar{a}(s_{i-1}^\ell),t-s_i^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_i^\ell),t-s_i^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{i-1}^\ell),t-s_i^\ell)\big], \\
&\qquad \big[\hat{K}_k(\bar{a}(s_l^\ell),t-s_l^\ell) - \hat{K}_k(\bar{a}(s_{l-1}^\ell),t-s_l^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_l^\ell),t-s_l^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{l-1}^\ell),t-s_l^\ell)\big]\Big) = 0.
\end{aligned}$$

If $s_i^\ell = s_l^\ell$,

$$\begin{aligned}
&Cov\Big(\big[\hat{K}_j(\bar{a}(s_i^\ell),t-s_i^\ell) - \hat{K}_j(\bar{a}(s_{i-1}^\ell),t-s_i^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_i^\ell),t-s_i^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{i-1}^\ell),t-s_i^\ell)\big], \\
&\qquad \big[\hat{K}_k(\bar{a}(s_l^\ell),t-s_l^\ell) - \hat{K}_k(\bar{a}(s_{l-1}^\ell),t-s_l^\ell) - \hat{K}_{\mathrm{m}}(\bar{a}(s_l^\ell),t-s_l^\ell) + \hat{K}_{\mathrm{m}}(\bar{a}(s_{l-1}^\ell),t-s_l^\ell)\big]\Big) \\
&= (\bar{a}(s_i^\ell)-\bar{a}(s_{i-1}^\ell))\Big[F_{j,k}(t-s_i^\ell,t-s_i^\ell) - F_j(t-s_i^\ell)F_k(t-s_i^\ell) + F_j(t-s_i^\ell)F_{\mathrm{m}}(t-s_i^\ell) \\
&\qquad + F_k(t-s_i^\ell)F_{\mathrm{m}}(t-s_i^\ell) - (F_{\mathrm{m}}(t-s_{i-1}^\ell))^2\Big].
\end{aligned}$$

Thus,

$$\begin{aligned}
Cov(\hat{Z}_{j,2,\ell}(t),\hat{Z}_{k,2,\ell}(t)) &= \sum_{i=1}^{\ell}(\bar{a}(s_i^\ell)-\bar{a}(s_{i-1}^\ell))\Big[F_{j,k}(t-s_i^\ell,t-s_i^\ell) - F_j(t-s_i^\ell)F_k(t-s_i^\ell) \\
&\qquad + F_j(t-s_i^\ell)F_{\mathrm{m}}(t-s_i^\ell) + F_k(t-s_i^\ell)F_{\mathrm{m}}(t-s_i^\ell) - (F_{\mathrm{m}}(t-s_i^\ell))^2\Big].
\end{aligned}$$

By Lebesgue's theorem, we have

$$\lim_{\ell\to\infty} Cov(\hat{Z}_{j,2,\ell}(t), \hat{Z}_{k,2,\ell}(t)) = \int_0^t \Big[F_{j,k}(t-s,t-s) - F_j(t-s)F_k(t-s) - F_{\mathrm{m}}(t-s) + F_k(t-s)F_{\mathrm{m}}(t-s) + F_{\mathrm{m}}(t-s)F_j(t-s) - (F_{\mathrm{m}}(t-s))^2\Big] d\bar{a}(s). \quad (5.30)$$

Thus, by (5.29) and (5.30), we obtain (5.11). This completes the proof of the lemma. □

Proof of Theorem 3.4. By the fact that $\hat{\boldsymbol{Z}}_1$ and $\hat{\boldsymbol{M}}_1$ are well-defined continuous Gaussian processes, Lemma 5.1, (3.16) and (3.22), we see that $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}})$ is a Gaussian process. It is evident that $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}})$ has mean $\mathbf{0}$. For the covariance functions of $\hat{X}_j$ and $\hat{Y}_k$, $j,k=1,...,K$, here we only show (3.27), since the other covariance formulas follow from the similar argument. By the independence of $\hat{A}(\cdot)$ and $\hat{K}(\cdot,\cdot)$, Lemma 5.1, the Ito isometry, (3.16) and (3.22), we have, for each $t \geq 0$ and $j,k=1,...,K$,

$$\begin{aligned}
&Cov(\hat{X}_j(t), \hat{Y}_k(t)) \\
&= Cov\big(\hat{M}_{j,1}(t) + \hat{M}_{j,2}(t), \hat{Z}_{k,1}(t) + \hat{Z}_{k,2}(t)\big) \\
&= c_a^2 \int_0^t F_j^c(t-s)(F_k(t-s) - F_{\mathrm{m}}(t-s))d\bar{a}(s) + Cov\big(\hat{M}_{j,2}(t), \hat{Z}_{k,2}(t)\big) \\
&= c_a^2 \int_0^t F_j^c(t-s)(F_k(t-s) - F_{\mathrm{m}}(t-s))d\bar{a}(s) + \int_0^t \Big[F_{\mathrm{m}}(t-s) - F_{j,k}(t-s,t-s) \\
&\qquad - F_j(t-s)F_{\mathrm{m}}(t-s) + F_j(t-s)F_k(t-s)\Big] d\bar{a}(s) \\
&= \int_0^t \Big[(c_a^2 - 1)\Big(F_j^c(t-s)(F_k(t-s) - F_{\mathrm{m}}(t-s))\Big) \\
&\qquad + \Big(F_j^c(t-s)F_k(t-s) + F_j(t-s)F_k(t-s) - F_{j,k}(t-s,t-s)\Big)\Big] d\bar{a}(s).
\end{aligned}$$

Thus, (3.27) holds.

The Gaussian property of $\hat{S}$, and its mean and variance can be obtained similarly. The claim that $\lim_{t\to\infty} t^{-1} Var(\hat{S}(t)) = \lambda c_a^2$ follows from a direct calculation when $\bar{a}(t) = \lambda t$, $t \geq 0$. The proof of this theorem is completed. □

6. Proof of the FCLT

In this section, we prove the FCLT for the processes $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$, Theorem 3.3. We first give representations for the processes $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$ by the multiparameter sequential empirical processes $\hat{K}^n$. In §6.1, we show the tightness of the diffusion-scaled processes $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$. We prove their convergence of the finite dimensional distributions in §6.2.

Lemma 6.1 **(Representations of $\hat{\boldsymbol{X}}^n$, $\hat{\boldsymbol{Y}}^n$ and $\hat{S}^n$).** *The processes $\hat{\boldsymbol{X}}^n$, $\hat{\boldsymbol{Y}}^n$ and $\hat{S}^n$ in (3.14) can be represented as: for each $t \geq 0$ and $k=1,...,K$,*

$$\hat{\boldsymbol{X}}^n(t) = \hat{\boldsymbol{M}}_1^n(t) + \hat{\boldsymbol{M}}_2^n(t), \quad \hat{\boldsymbol{M}}_i^n(t) := (\hat{M}_{1,i}^n(t), ..., \hat{M}_{K,i}^n(t)), \quad i=1,2, \quad (6.1)$$

$$\hat{\boldsymbol{Y}}^n(t) = \hat{\boldsymbol{Z}}_1^n(t) + \hat{\boldsymbol{Z}}_2^n(t), \quad \hat{\boldsymbol{Z}}_i^n(t) := (\hat{Z}_{1,i}^n(t), ..., \hat{Z}_{K,i}^n(t)), \quad i=1,2, \quad (6.2)$$

$$\hat{S}^n(t) = \hat{V}_1^n(t) + \hat{V}_2^n(t), \quad (6.3)$$

where

$$\hat{M}_{k,1}^n(t) := \int_0^t F_k^c(t-s)d\hat{A}^n(s) = \hat{A}^n(t) - \int_0^t \hat{A}^n(s)dF_k^c(t-s), \quad (6.4)$$

$$\hat{M}_{k,2}^n(t) := \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}(s + x_k > t) d\hat{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right) = -\int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}(s + x_k \leq t) d\hat{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right), \quad (6.5)$$

$$\hat{V}_1^n(t) := \int_0^t F_{\mathrm{m}}(t-s)d\hat{A}^n(s) = -\int_0^t \hat{A}^n(s)dF_m(t-s), \tag{6.6}$$
$$\hat{V}_2^n(t) := \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}(s+x_j \le t,\ \forall\ j)d\hat{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right), \tag{6.7}$$
$$\hat{Z}_{k,1}^n(t) := \int_0^t (F_k(t-s) - F_{\mathrm{m}}(t-s))d\hat{A}^n(s) = \int_0^t \hat{A}^n(s)d(F_{\mathrm{m}}(t-s) - F_k(t-s)), \tag{6.8}$$
$$\hat{Z}_{k,2}^n(t) := \int_0^t \int_{\mathbb{R}_+^K} \left(\mathbf{1}(s+x_k \le t) - \mathbf{1}(s+x_j \le t,\ \forall\ j)\right)d\hat{K}^n\left(\bar{A}^n(s), \boldsymbol{x}\right) = -\hat{M}_{k,2}^n(t) - \hat{V}_2^n(t), \tag{6.9}$$

and the integrals in (6.4)-(6.9) are defined as Stieltjes integrals for functions of bounded variation as integrators.

Proof. The representations of the processes $\hat{\boldsymbol{X}}^n$, $\hat{\boldsymbol{Y}}^n$ and $\hat{S}^n$ follow from equations (2.13), (2.14), (2.15), (3.14) and direct calculations. □

6.1. Proof of Tightness In this section, we will show the tightness of the sequence of diffusion-scaled processes in (3.15).

Lemma 6.2. *The sequence of the processes $\{(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n):\ n \ge 1\}$ is tight in $\mathbb{D}^{2K+1}$.*

We provide two proofs for the lemma. The first proof is based on the representation of the processes $\hat{X}_k^n$ and $\hat{S}^n$ with the two-parameter sequential empirical process for each k separately, as for $G/GI/\infty$ queues in [37]. The second approach is based on the representation of the processes $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$ with the multiparameter sequential empirical processes $\hat{K}^n$. We state the proof of the first approach below, and the proof of the second approach in the Appendix B. We think that the techniques developed in the new approach will turn out to be useful to study fork-join networks with NES in the many-server heavy-traffic regimes (see, e.g., [44]).

Proof of Lemma 6.2. Since tightness on product spaces is equivalent to tightness on each of the component spaces (Theorem 11.6.7 of [69]), we only need to show $\{\hat{X}_k^n : n \ge 1\}$, $\{\hat{Y}_k^n : n \ge 1\}$ and $\{\hat{S}^n : n \ge 1\}$ are tight separately for each $k = 1, ..., K$. As in [37], we have the representations for $\hat{X}_k^n$ and $\hat{S}^n$, $k = 1, ..., K$, as follows:

$$\hat{X}_k^n(t) = \hat{M}_{k,1}^n(t) + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}(s+x > t)d\hat{K}_k^n(\bar{A}^n(s), x), \quad t \ge 0,$$
$$\hat{S}^n(t) = \hat{V}_1^n(t) + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}(s+x \le t)d\hat{K}_{\mathrm{m}}^n(\bar{A}^n(s), x), \quad t \ge 0,$$

where $\hat{M}_{k,1}^n$ and $\hat{V}_1^n$ are defined in (6.4) and (6.6), respectively, and $\hat{K}_k^n(t,x)$ and $\hat{K}_{\mathrm{m}}^n(t,x)$ are denoted as

$$\hat{K}_k^n(t,x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \left(\mathbf{1}(\eta_k^i \le x) - F_k(x)\right),\ t \ge 0,\ x \ge 0, \tag{6.10}$$
$$\hat{K}_{\mathrm{m}}^n(t,x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \left(\mathbf{1}(\eta_{\mathrm{m}}^i \le x) - F_{\mathrm{m}}(x)\right),\ t \ge 0,\ x \ge 0. \tag{6.11}$$

Note that $\hat{K}_k^n(\cdot,\cdot)$ and $\hat{K}_{\mathrm{m}}^n(\cdot,\cdot)$ are all two-parameter sequential empirical processes in space $\mathbb{D}([0,\infty),\mathbb{D})$, for $k = 1, ..., K$. Following the proof of [37] and [51], we see that $\hat{X}_k^n$ and $\hat{S}^n$ weakly converge, as $n \to \infty$, which implies that $\{\hat{X}_k^n : n \ge 1\}$ and $\{\hat{S}^n : n \ge 1\}$ are tight, $k = 1, ..., K$. Since $\hat{Y}_k^n = \hat{A}^n - \hat{X}_k^n - \hat{S}^n$, we also obtain the tightness of $\{\hat{Y}_k^n : n \ge 1\}$ for $k = 1, ..., K$. This complete the proof of the lemma. □

6.2. Proof of Convergence of the Finite-Dimensional Distributions In this subsection, we will show that the finite-dimensional distributions of $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$ converge to those of $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}, \hat{S})$ as $n \to \infty$. Since we have already proved the tightness of $\{(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n) : n \geq 1\}$, the results of this subsection complete the proof of the convergence $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n) \Rightarrow (\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}, \hat{S})$ in $\mathbb{D}^{2K+1}$ as $n \to \infty$.

We first show the convergence of $(\hat{\boldsymbol{M}}_1^n, \hat{\boldsymbol{Z}}_1^n, \hat{V}_1^n)$ in finite-dimensional distributions. As a prerequisite, we define the mapping $\chi : \mathbb{D} \to \mathbb{D}^{2K+1}$ by

$$\chi(x) = (\phi_1(x), ..., \phi_K(x), \varphi_1(x), ..., \varphi_K(x), \psi(x)), \quad \text{for} \quad x \in \mathbb{D}, \tag{6.12}$$

where the mappings $\phi_k : \mathbb{D} \to \mathbb{D}$, $\varphi_k : \mathbb{D} \to \mathbb{D}$ and $\psi : \mathbb{D} \to \mathbb{D}$ are defined by

$$\phi_k(x)(t) := x(t) - \int_0^t x(s) dF_k^c(t-s), \quad t \geq 0, \tag{6.13}$$

$$\varphi_k(x)(t) := \int_0^t x(s) d(F_{\mathrm{m}}(t-s) - F_k(t-s)), \quad t \geq 0, \tag{6.14}$$

$$\psi(x)(t) := -\int_0^t x(s) dF_{\mathrm{m}}(t-s), \quad t \geq 0. \tag{6.15}$$

for $x \in \mathbb{D}$ and $k = 1, ..., K$. We now state the continuity property of the above mappings in the following lemma.

LEMMA 6.3. *The mappings ϕ_k defined in (6.13), φ_k defined in (6.14), $k = 1, ..., K$, and ψ defined in (6.15) are all continuous in $\mathbb{D}$, and the mapping $\chi(\cdot)$ defined in (6.12) is continuous in $\mathbb{D}^{2K+1}$.*

Proof. The proof for the continuity of ϕ_k, φ_k, $k = 1, ..., K$, and ψ is analogous to the proof of Lemma 9.1 in [51], and we omit it here for brevity. By noting that we endow the maximum metric on the product space $\mathbb{D}^{2K+1}$, it is easy to see that the mapping χ is also continuous. □

From (6.4), (3.17), (6.8), (3.23), (6.6) and (3.20), we see that $\hat{M}_{k,1}^n(t) = \phi_k(\hat{A}^n)(t)$, $\hat{M}_{k,1}(t) = \phi_k(\hat{A})(t)$, $\hat{Z}_{k,1}^n(t) = \varphi_k(\hat{A}^n)(t)$, $\hat{Z}_{k,1}(t) = \varphi_k(\hat{A})(t)$, $\hat{V}_1^n(t) = \psi(\hat{A}^n)(t)$ and $\hat{V}_1(t) = \psi(\hat{A})(t)$, for $t \geq 0$ and $k = 1, ..., K$. By Assumption 1 and the continuity of the mapping χ (Lemma 6.3) as well as the continuous mapping theorem, we immediately obtain

$$(\hat{\boldsymbol{M}}_1^n, \hat{\boldsymbol{Z}}_1^n, \hat{V}_1^n) \Rightarrow (\hat{\boldsymbol{M}}_1, \hat{\boldsymbol{Z}}_1, \hat{V}_1) \quad \text{in} \quad \mathbb{D}^{2K+1} \quad \text{as} \quad n \to \infty, \tag{6.16}$$

which implies

$$(\hat{\boldsymbol{M}}_1^n, \hat{\boldsymbol{Z}}_1^n, \hat{V}_1^n) \xrightarrow{d_f} (\hat{\boldsymbol{M}}_1, \hat{\boldsymbol{Z}}_1, \hat{V}_1) \quad \text{as} \quad n \to \infty. \tag{6.17}$$

Next, we will show the convergence of $(\hat{\boldsymbol{M}}_2^n, \hat{\boldsymbol{Z}}_2^n, \hat{V}_2^n)$ in finite-dimensional distributions, jointly with $(\hat{\boldsymbol{M}}_1^n, \hat{\boldsymbol{Z}}_1^n, \hat{V}_1^n)$. In order to achieve that, we first introduce some additional processes. For $t \geq 0$, we divide the interval $[0, t]$ by the sequence $\{s_i^l : 0 \leq i \leq l\}$: $0 = s_0^l < s_1^l < ... < s_l^l = t$ satisfying $\max_{1 \leq i \leq l} |s_i^l - s_{i-1}^l| \to 0$, as $l \to \infty$. We define, for $t \geq 0$ and $k = 1, ..., K$,

$$\tilde{M}_{k,2,l}^n(t) := -\int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}_{k,l,t}(s, \boldsymbol{x}) d\hat{K}^n(\bar{a}(s), \boldsymbol{x}), \tag{6.18}$$

$$\tilde{Z}_{k,2,l}^n(t) := \int_0^t \int_{\mathbb{R}_+^K} [\mathbf{1}_{k,l,t}(s, \boldsymbol{x}) - \mathbf{1}_{\mathrm{m},l,t}(s, \boldsymbol{x})] d\hat{K}^n(\bar{a}(s), \boldsymbol{x}), \tag{6.19}$$

$$\tilde{V}_{2,l}^n(t) := \int_0^t \int_{\mathbb{R}_+^K} \mathbf{1}_{\mathrm{m},l,t}(s, \boldsymbol{x}) d\hat{K}^n(\bar{a}(s), \boldsymbol{x}), \tag{6.20}$$

where $\mathbf{1}_{k,l,t}(\cdot,\cdot)$ and $\mathbf{1}_{m,l,t}(\cdot,\cdot)$ are defined in (5.8) and (5.9), respectively. $\hat{M}^n_{k,2,l}(t)$, $\hat{Z}^n_{k,2,l}(t)$ and $\hat{V}^n_{2,l}(t)$ are defined analogously as $\tilde{M}^n_{k,2,l}(t)$, $\tilde{Z}^n_{k,2,l}(t)$ and $\tilde{V}^n_{2,l}(t)$ with $\bar{a}(\cdot)$ replaced by $\bar{A}^n(\cdot)$ in the above integrals respectively, for $t\geq 0$ and $k=1,...,K$. We set $\tilde{M}^n_{k,2,l} := \{\tilde{M}^n_{k,2,l}(t) : t\geq 0\}$, $\tilde{Z}^n_{k,2,l} := \{\tilde{Z}^n_{k,2,l}(t) : t\geq 0\}$, $k=1,...,K$, and $\tilde{V}^n_{2,l} := \{\tilde{V}^n_{2,l}(t) : t\geq 0\}$. Note that, for $k=1,...,K$ and $t\geq 0$, $\tilde{M}^n_{k,2,l}(t)$, $\tilde{Z}^n_{k,2,l}(t)$ and $\tilde{V}^n_{2,l}(t)$ can be rewritten as

$$\tilde{M}^n_{k,2,l}(t) = -\sum_{i=1}^{l}\Delta\hat{K}^n\left((\bar{a}(s^l_{i-1}),\mathbf{0});(\bar{a}(s^l_i),\boldsymbol{x}^i)\right), \quad \tilde{V}^n_{2,l}(t) = \sum_{i=1}^{l}\Delta\hat{K}^n\left((\bar{a}(s^l_{i-1}),\mathbf{0});(\bar{a}(s^l_i),\boldsymbol{y}^i)\right),$$
$$\tilde{Z}^n_{k,2,l}(t) = \sum_{i=1}^{l}\left\{\Delta\hat{K}^n\left((\bar{a}(s^l_{i-1}),\mathbf{0});(\bar{a}(s^l_i),\boldsymbol{x}^i)\right) - \Delta\hat{K}^n\left((\bar{a}(s^l_{i-1}),\mathbf{0});(\bar{a}(s^l_i),\boldsymbol{y}^i)\right)\right\},$$

where $\boldsymbol{x}^i\in\mathbb{R}^K$ with $x^i_j=\infty$ for $j\neq k$ and $x^i_k = t-s^l_i$, and $\boldsymbol{y}^i\in\mathbb{R}^K$ with $y^i_j = t-s^l_i$ for $1\leq j\leq K$. Set $\tilde{\boldsymbol{M}}^n_{2,l} := (\tilde{M}^n_{1,2,l},...,\tilde{M}^n_{K,2,l})$ and $\tilde{\boldsymbol{Z}}^n_{2,l} := (\tilde{Z}^n_{1,2,l},...,\tilde{Z}^n_{K,2,l})$. For $t^k_{i,1}, t^k_{i',2}, t_j\geq 0$, $c^k_{i,1}, c^k_{i',2}, c_j\in\mathbb{R}$, and positive integers $I_{k,1}, I_{k,2}$ and I_3, where $i=1,...,I_{k,1}$, $i'=1,...,I_{k,2}$, $j=1,...,I_3$ and $k=1,...,K$, with the weak convergence of $\hat{K}^n$ in (3.4), we see that, as $n\to\infty$,

$$\sum_{k=1}^{K}\left[\sum_{i=1}^{I_{k,1}}c^k_{i,1}\tilde{M}^n_{k,2,l}(t^k_{i,1}) + \sum_{i'=1}^{I_{k,2}}c^k_{i',2}\tilde{Z}^n_{k,2,l}(t^k_{i',2})\right] + \sum_{j=1}^{I_3}c_j\tilde{V}^n_{2,l}(t_j)$$
$$\Rightarrow \sum_{k=1}^{K}\left[\sum_{i=1}^{I_{k,1}}c^k_{i,1}\hat{M}_{k,2,l}(t^k_{i,1}) + \sum_{i'=1}^{I_{k,2}}c^k_{i',2}\hat{Z}_{k,2,l}(t^k_{i',2})\right] + \sum_{j=1}^{I_3}c_j\hat{V}_{2,l}(t_j),$$

where we recall $\hat{M}_{k,2,l}$, $\hat{Z}_{k,2,l}$ and $\hat{V}_{2,l}$ are defined in (5.5), (5.7) and (5.6), respectively, for $k=1,...,K$. By the Cramer-Wold theorem (see Theorem 3.9.5 in [21]), we have

$$(\tilde{\boldsymbol{M}}^n_{2,l}, \tilde{\boldsymbol{Z}}^n_{2,l}, \tilde{V}^n_{2,l}) \xrightarrow{d_f} (\hat{\boldsymbol{M}}_{2,l}, \hat{\boldsymbol{Z}}_{2,l}, \hat{V}_{2,l}) \quad \text{as} \quad n\to\infty,$$

where $\hat{\boldsymbol{M}}_{2,l} := (\hat{M}_{1,2,l},...,\hat{M}_{K,2,l})$ and $\hat{\boldsymbol{Z}}_{2,l} := (\hat{Z}_{1,2,l},...,\hat{Z}_{K,2,l})$. Since $\hat{A}^n(\cdot)$ and $\hat{K}^n(\cdot,\cdot)$ are two independent processes, $\hat{\boldsymbol{M}}^n_1, \hat{\boldsymbol{Z}}^n_1$ and $\hat{V}^n_1$ are independent of $\tilde{\boldsymbol{M}}^n_{2,l}, \tilde{\boldsymbol{Z}}^n_{2,l}$ and $\tilde{V}^n_{2,l}$. Thus,

$$(\hat{\boldsymbol{M}}^n_1, \hat{\boldsymbol{Z}}^n_1, \hat{V}^n_1, \tilde{\boldsymbol{M}}^n_{2,l}, \tilde{\boldsymbol{Z}}^n_{2,l}, \tilde{V}^n_{2,l}) \xrightarrow{d_f} (\hat{\boldsymbol{M}}_1, \hat{\boldsymbol{Z}}_1, \hat{V}_1, \hat{\boldsymbol{M}}_{2,l}, \hat{\boldsymbol{Z}}_{2,l}, \hat{V}_{2,l}) \quad \text{as} \quad n\to\infty.$$

As $l\to\infty$, by the definitions of $\hat{\boldsymbol{M}}_{2,l}$, $\hat{\boldsymbol{Z}}_{2,l}$ and $\hat{V}_{2,l}$, $(\hat{\boldsymbol{M}}_{2,l}, \hat{\boldsymbol{Z}}_{2,l}, \hat{V}_{2,l})$ converges to $(\hat{\boldsymbol{M}}_2, \hat{\boldsymbol{Z}}_2, \hat{V}_2)$ in L^2; see Definition 5.1. Note that L^2 convergence implies convergence in probability. Thus,

$$(\hat{\boldsymbol{M}}_{2,l}, \hat{\boldsymbol{Z}}_{2,l}, \hat{V}_{2,l}) \xrightarrow{P} (\hat{\boldsymbol{M}}_2, \hat{\boldsymbol{Z}}_2, \hat{V}_2) \quad \text{as} \quad l\to\infty.$$

Therefore, it suffices to show the following, for $T>0$, $\epsilon>0$,

$$\lim_{n\to\infty}P\left(\sup_{0\leq t\leq T}\left\{\sum_{k=1}^{K}\left[|\hat{M}^n_{k,2,l}(t)-\tilde{M}^n_{k,2,l}(t)| + |\hat{Z}^n_{k,2,l}(t)-\tilde{Z}^n_{k,2,l}(t)|\right] + |\hat{V}^n_{2,l}(t)-\tilde{V}^n_{2,l}(t)|\right\} > \epsilon\right) = 0, \tag{6.21}$$

and, for $t>0$ and $\epsilon>0$,

$$\lim_{l\to\infty}\limsup_{n\to\infty}P\left(\sum_{k=1}^{K}\left[|\hat{M}^n_{k,2}(t)-\hat{M}^n_{k,2,l}(t)| + |\hat{Z}^n_{k,2}(t)-\hat{Z}^n_{k,2,l}(t)|\right] + |\hat{V}^n_2(t)-\hat{V}^n_{2,l}(t)| > \epsilon\right) = 0. \tag{6.22}$$

First, we focus on the proof of (6.21). Note that

$$\begin{aligned}
&P\left(\sup_{0\le t\le T}\left\{\sum_{k=1}^K\left[|\hat M^n_{k,2,l}(t)-\tilde M^n_{k,2,l}(t)|+|\hat Z^n_{k,2,l}(t)-\tilde Z^n_{k,2,l}(t)|\right]+|\hat V^n_{2,l}(t)-\tilde V^n_{2,l}(t)|\right\}>\epsilon\right)\\
&\le P\left(\sup_{0\le t\le T}\sum_{k=1}^K|\hat M^n_{k,2,l}(t)-\tilde M^n_{k,2,l}(t)|+\sup_{0\le t\le T}\sum_{k=1}^K|\hat Z^n_{k,2,l}(t)-\tilde Z^n_{k,2,l}(t)|+\sup_{0\le t\le T}|\hat V^n_{2,l}(t)-\tilde V^n_{2,l}(t)|>\epsilon\right)\\
&\le P\left(\sup_{0\le t\le T}\sum_{k=1}^K|\hat M^n_{k,2,l}(t)-\tilde M^n_{k,2,l}(t)|>\frac{\epsilon}{3}\right)+P\left(\sup_{0\le t\le T}\sum_{k=1}^K|\hat Z^n_{k,2,l}(t)-\tilde Z^n_{k,2,l}(t)|>\frac{\epsilon}{3}\right)\\
&\qquad+P\left(\sup_{0\le t\le T}|\hat V^n_{2,l}(t)-\tilde V^n_{2,l}(t)|>\frac{\epsilon}{3}\right). \qquad (6.23)
\end{aligned}$$

Since $\bar a(\cdot)$ and $\hat K(\cdot,\cdot)$ are continuous, with Assumption 1 and (3.4), we easily see that the three terms on (6.23) all converge to 0 when $n\to\infty$. Thus, (6.21) holds.

Next, we will show (6.22). We can represent $\hat M^n_{k,2,l}(\cdot)$, $\hat M^n_{k,2}(\cdot)$, $\hat V^n_{2,l}(\cdot)$ and $\hat V^n_2(\cdot)$ in the form of the two-parameter sequential empirical process, and represent $\hat Z^n_{k,2,l}(\cdot)$ and $\hat Z^n_{k,2}(\cdot)$ as the difference of two two-parameter sequential empirical processes, for $k=1,\dots,K$. These representations are shown in the following:

$$\begin{aligned}
\hat M^n_{k,2}(t)&=-\int_0^t\int_0^\infty\mathbf{1}(s+x\le t)d\hat K^n_k(\bar A^n(s),x),\quad t\ge0,\\
\hat Z^n_{k,2}(t)&=\int_0^t\int_0^\infty\mathbf{1}(s+x\le t)d\hat K^n_k(\bar A^n(s),x)-\int_0^t\int_0^\infty\mathbf{1}(s+x\le t)d\hat K^n_{\mathrm m}(\bar A^n(s),x),\quad t\ge0,\\
\hat V^n_2(t)&=\int_0^t\int_0^\infty\mathbf{1}(s+x\le t)d\hat K^n_{\mathrm m}(\bar A^n(s),x),\quad t\ge0,
\end{aligned}$$

where $\hat K^n_k(\cdot,\cdot)$ and $\hat K^n_{\mathrm m}(\cdot,\cdot)$ are defined in (6.10) and (6.11), respectively, $k=1,\dots,K$, and the integrals above are defined as Stieltjes integrals for functions of bounded variation. Following the proof of Lemma 11.1 in [51], we obtain that (6.22) also holds. Thus, we have shown

$$(\hat{\boldsymbol M}^n_1,\hat{\boldsymbol Z}^n_1,\hat V^n_1,\hat{\boldsymbol M}^n_2,\hat{\boldsymbol Z}^n_2,\hat V^n_2)\xrightarrow{d_f}(\hat{\boldsymbol M}_1,\hat{\boldsymbol Z}_1,\hat V_1,\hat{\boldsymbol M}_2,\hat{\boldsymbol Z}_2,\hat V_2)\quad\text{as}\quad n\to\infty.$$

By the continuous mapping theorem, we have

$$(\hat{\boldsymbol M}^n_1+\hat{\boldsymbol M}^n_2,\hat{\boldsymbol Z}^n_1+\hat{\boldsymbol Z}^n_2,\hat V^n_1+\hat V^n_2)\xrightarrow{d_f}(\hat{\boldsymbol M}_1+\hat{\boldsymbol M}_2,\hat{\boldsymbol Z}_1+\hat{\boldsymbol Z}_2,\hat V_1+\hat V_2)\quad\text{as}\quad n\to\infty.$$

Thus,

$$(\hat{\boldsymbol X}^n,\hat{\boldsymbol Y}^n,\hat S^n)\xrightarrow{d_f}(\hat{\boldsymbol X},\hat{\boldsymbol Y},\hat S)\quad\text{as}\quad n\to\infty.$$

This completes the proof of this lemma. □

7. Concluding Remarks We have developed a new approach to study fork-join networks with NES when all service stations have infinitely many servers and the parallel tasks of each job have correlated service times. By representing the service processes, the waiting buffer dynamics for synchronization and the synchronized processes via a common sequential empirical process driven by the service vectors of parallel tasks of each job, we characterize the joint transient and stationary distributions of these processes as a multidimensional Gaussian distribution asymptotically when the arrival process satisfies an FCLT with a Brownian motion limit. We have assumed that the system starts from empty. It remains to study general initial conditions as in [37, 51, 2]. That requires tracking the status (in service or in the waiting buffer for synchronization) of each task of jobs initially in the system.

This approach can be further developed to study fork-join networks with NES in the many-server regimes, for example, the recent development in the Halfin-Whitt regime in [44], where non-empty initial condition is considered for $K=2$. It remains open to investigate if our approach can be used to study scheduling and routing control problems in single class or multi-class models with potentially multiple processing stages. In our model, we have assumed that each job is forked into a fixed number of tasks. It may be interesting to study models in which each job is forked into a random number of tasks and/or the number of service stations is large.

Our results may be already useful to study many-server fork-join network models with NES. For example, the mean and covariance approximations in Theorem 3.4 can be regarded as approximations for the offered load processes in the corresponding many-server models, and thus, can be potentially used for staffing decisions to achieve certain service level constraints and/or stabilize delay probabilities for service and synchronization. It will be also interesting to investigate how these results can be applied in practical applications by conducting data-driven research, for example, hospital patient flows and MapReduce scheduling.

Appendix A: Proofs of the Results in §4.2 In the appendix, we provide the proofs for the results in §4.2 for multiparameter point processes.

Proof of Lemma 4.2. By the definition of $\mathcal{F}^i_{\boldsymbol{t}}$, $\mathcal{F}^i_{\boldsymbol{t}}$ is a σ-algebra for $\boldsymbol{t}\in\mathbb{R}^K_+$. Note that since $L^c_{\boldsymbol{t}}\subseteq L^c_{\boldsymbol{s}}$ for $\boldsymbol{s}\le\boldsymbol{t}$, $\mathcal{F}^i_{\boldsymbol{s}}\subseteq\mathcal{F}^i_{\boldsymbol{t}}$.

We next prove $\mathcal{F}^i=\{\mathcal{F}^i_{\boldsymbol{t}}:\boldsymbol{t}\in\mathbb{R}^K_+\}$ is the smallest filtration to which N^i is adapted. Let $\mathcal{G}:=\{\mathcal{G}_{\boldsymbol{t}}:\boldsymbol{t}\in\mathbb{R}^K_+\}$ be any other filtration where $(\boldsymbol{\zeta}^i)^{-1}(L_{\boldsymbol{t}})\in\mathcal{G}_{\boldsymbol{t}}$ for any $\boldsymbol{t}\in\mathbb{R}^K_+$. Let $C=(\boldsymbol{\zeta}^i)^{-1}(B)\in\mathcal{F}^i_{\boldsymbol{t}}$. If $L^c_{\boldsymbol{t}}\cap B=\emptyset$, $B\subseteq L_{\boldsymbol{t}}$. Then, $C=(\boldsymbol{\zeta}^i)^{-1}(B)\cap(\boldsymbol{\zeta}^i)^{-1}(L_{\boldsymbol{t}})\in\mathcal{G}_{\boldsymbol{t}}$. If $L^c_{\boldsymbol{t}}\subseteq B$, then $C=\{(\boldsymbol{\zeta}^i)^{-1}(B)\cap(\boldsymbol{\zeta}^i)^{-1}(L_{\boldsymbol{t}})\}\cup(\zeta^i)^{-1}(L^c_{\boldsymbol{t}})\in\mathcal{G}_{\boldsymbol{t}}$. Hence, $\mathcal{F}^i_{\boldsymbol{t}}\subseteq\mathcal{G}_{\boldsymbol{t}}$. □

Proof of Lemma 4.3. For $i\in\mathbb{N}$, by the definition of Λ^i, Λ^i is increasing, predictable and càdlàg with $\Lambda^i(0)=0$. It is sufficient to show that $N^i-\Lambda^i$ is a multiparameter martingale, which is equivalent to show that $\forall\boldsymbol{s}\le\boldsymbol{t}$, and $\forall B\in\mathcal{F}^i_{\boldsymbol{s}}$,

$$E[\mathbf{1}(B)(N^i(\boldsymbol{t})-N^i(\boldsymbol{s})-\Lambda^i(\boldsymbol{t})+\Lambda^i(\boldsymbol{s})]=0. \tag{A.1}$$

Without loss of generality, we assume that $K=2$. When $K>2$, the proof can be easily generalized. If $\boldsymbol{\zeta}^i\le\boldsymbol{s}$, it is easily seen that $N^i(\boldsymbol{t})-N^i(\boldsymbol{s})-\Lambda^i(\boldsymbol{t})+\Lambda^i(\boldsymbol{s})=0,\ a.s.$ Thus, it is sufficient to show

$$E[\mathbf{1}(B)\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})(N^i(\boldsymbol{t})-N^i(\boldsymbol{s})-\Lambda^i(\boldsymbol{t})+\Lambda^i(\boldsymbol{s})]=0. \tag{A.2}$$

By the construction of the filtration $\mathcal{F}^i$, we have either $B\cap\{\boldsymbol{\zeta}^i\not\le\boldsymbol{s}\}=\emptyset$ or $B\cap\{\boldsymbol{\zeta}^i\not\le\boldsymbol{s}\}=\{\boldsymbol{\zeta}^i\not\le\boldsymbol{s}\}$. In the former case, (A.2) holds evidently. We now consider the latter case. In this case, the LHS of (A.2) can be written as

$$\begin{aligned}
&E[\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})(N^i(\boldsymbol{t})-N^i(\boldsymbol{s})-\Lambda^i(\boldsymbol{t})+\Lambda^i(\boldsymbol{s})]\\
&=E[\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})\mathbf{1}(\boldsymbol{\zeta}^i\le\boldsymbol{t})-\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})(\Lambda^i(\boldsymbol{t})-\Lambda^i(\boldsymbol{s})]\\
&=E[\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})\mathbf{1}(\boldsymbol{\zeta}^i\le\boldsymbol{t})]-E\left[\mathbf{1}(\boldsymbol{\zeta}^i\not\le\boldsymbol{s})\Big(\int_{L_{\boldsymbol{t}}}\frac{\mathbf{1}(\boldsymbol{\zeta}^i>\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(d\boldsymbol{u})\Big)-\int_{L_{\boldsymbol{s}}}\frac{\mathbf{1}(\boldsymbol{\zeta}^i>\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(d\boldsymbol{u})\Big)\right].
\end{aligned}$$

Observe that

$$\begin{aligned}
&\int_{L_{\boldsymbol{t}}}\frac{\mathbf{1}(\boldsymbol{\zeta}^i>\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(d\boldsymbol{u})-\int_{L_{\boldsymbol{s}}}\frac{\mathbf{1}(\boldsymbol{\zeta}^i>\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(d\boldsymbol{u})\\
&=\int_{s_1}^{t_1}\int_{s_2}^{t_2}\frac{\mathbf{1}(\zeta^i_1>u_1)\mathbf{1}(\zeta^i_2>u_2)}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(du_1,du_2)+\int_0^{s_1}\int_{s_2}^{t_2}\frac{\mathbf{1}(\zeta^i_1>u_1)\mathbf{1}(\zeta^i_2>u_2)}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(du_1,du_2)\\
&\quad+\int_{s_1}^{t_1}\int_0^{s_2}\frac{\mathbf{1}(\zeta^i_1>u_1)\mathbf{1}(\zeta^i_2>u_2)}{H_\zeta(T_{\boldsymbol{u}})}H_\zeta(du_1,du_2),
\end{aligned}$$

and

$$\mathbf{1}(\boldsymbol{\zeta}^i \nleq \boldsymbol{s}) = \mathbf{1}(\zeta_1^i \le s_1)\mathbf{1}(\zeta_2^i > s_2) + \mathbf{1}(\zeta_1^i > s_1)\mathbf{1}(\zeta_2^i > s_2) + \mathbf{1}(\zeta_1^i > s_1)\mathbf{1}(\zeta_2^i \le s_2).$$

Thus, we obtain

$$\begin{aligned}
&E\left[\mathbf{1}(\boldsymbol{\zeta}^i \nleq \boldsymbol{s})\Big(\int_{L_{\boldsymbol{t}}} \frac{\mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(d\boldsymbol{u})\Big) - \int_{L_{\boldsymbol{s}}} \frac{\mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(d\boldsymbol{u})\Big)\right] \\
&= E\left[\int_{s_1}^{t_1}\int_{s_2}^{t_2} \frac{\mathbf{1}(\zeta_1^i > u_1)\mathbf{1}(\zeta_2^i > u_2)}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(du_1, du_2) + \int_0^{s_1\wedge\zeta_1^i}\int_{s_2}^{t_2} \frac{\mathbf{1}(\zeta_1^i > u_1)\mathbf{1}(\zeta_2^i > u_2)}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(du_1, du_2)\right. \\
&\quad \left. + \int_{s_1}^{t_1}\int_0^{\zeta_2^i\wedge s_2} \frac{\mathbf{1}(\zeta_1^i > u_1)\mathbf{1}(\zeta_2^i > u_2)}{H_\zeta(T_{\boldsymbol{u}})} H_\zeta(du_1, du_2)\right] \qquad \text{(By Fubini's Theorem)} \\
&= \int_{s_1}^{t_1}\int_{s_2}^{t_2} H_\zeta(du_1, du_2) + \int_0^{s_1}\int_{s_2}^{t_2} H_\zeta(du_1, du_2) + \int_{s_1}^{t_1}\int_0^{s_2} H_\zeta(du_1, du_2).
\end{aligned}$$

Then we can see that equation (A.2) holds, and thus, $N^i - \Lambda^i$ is a multiparameter martingale with the filtration $\mathcal{F}^i$. Therefore, Λ^i is a compensator of N^i.

The final step is to show (4.1). Following the definitions of the multivariate integral and the increment in $\mathbb{R}_+^K$ (see (5.1)), by Lemma B.5, we obtain that

$$\int_{\mathbb{R}_+^K} \mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u}) \frac{H_\zeta(d\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})} = (-1)^K \int_{\mathbb{R}_+^K} \mathbf{1}(\boldsymbol{\zeta}^i > \boldsymbol{u}) \frac{H_\zeta^c(d\boldsymbol{u})}{H_\zeta(T_{\boldsymbol{u}})},$$

where $H_\zeta^c(\boldsymbol{x}) := P(\boldsymbol{\zeta}^i > \boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}_+^K$. Then using the expression on the RHS, we can easily check that (4.1) holds under continuity assumption of H_ζ. This completes the proof of the lemma. □

Proof of Lemma 4.4. Since $X^\ell(\boldsymbol{t}) - Z^\ell(\boldsymbol{t}) = \sum_{i=1}^{\ell}(N^i(\boldsymbol{t}) - \Lambda^i(\boldsymbol{t}))$, it suffices to show $N^i(\boldsymbol{t}) - \Lambda^i(\boldsymbol{t})$ is a multiparameter martingale relative to the filtration $\mathcal{F}^{X^\ell}$ for $i = 1, ..., \ell$. Observe that $\mathcal{F}_\infty^i$ are mutually independent, $i \in \mathbb{N}$. For $\boldsymbol{s} \le \boldsymbol{t}$, $B_i \in \mathcal{F}_{\boldsymbol{s}}^i$ and $B = \cap_{1\le j\le \ell} B_j$, we then have

$$E[\mathbf{1}(B)(N^i(\boldsymbol{t}) - N^i(\boldsymbol{s}) - \Lambda^i(\boldsymbol{t}) + \Lambda^i(\boldsymbol{s}))] = \left[\prod_{j\le\ell, j\ne i} P(B_j)\right] E[\mathbf{1}(B_i)(N^i(\boldsymbol{t}) - N^i(\boldsymbol{s}) - \Lambda^i(\boldsymbol{t}) + \Lambda^i(\boldsymbol{s}))] = 0.$$

A monotone class argument implies that $E[\mathbf{1}(B)(N^i(\boldsymbol{t}) - N^i(\boldsymbol{s}) - \Lambda^i(\boldsymbol{t}) + \Lambda^i(\boldsymbol{s}))] = 0$ for all $B \in \mathcal{F}_{\boldsymbol{s}}^{X^\ell}$. Thus, $N^i(\boldsymbol{t}) - \Lambda^i(\boldsymbol{t})$ is a multiparameter martingale relative to the filtration $\mathcal{F}^{X^\ell}$, $\boldsymbol{t} \in \mathbb{R}_+^K$. The claim that $E[(Z^\ell(\boldsymbol{t}))^2] < \infty$ for $\boldsymbol{t} \in \mathbb{R}_+^K$ follows from (4.1). This completes the proof of this lemma. □

Appendix B: The Second Approach to the Proof of Tightness in §6.1 In this section, we provide the second approach to the proof of tightness of the processes $(\hat{\boldsymbol{X}}^n, \hat{\boldsymbol{Y}}^n, \hat{S}^n)$ in Lemma 6.2.

To show the tightness of $\{\hat{X}_k^n : n \ge 1\}$, it suffices to show the tightness of $\{\hat{M}_{k,1}^n : n \ge 1\}$ in (6.4) and $\{\hat{M}_{k,2}^n : n \ge 1\}$ in (6.5), for $k = 1, ..., K$. Similarly, the tightness of $\{\hat{Y}_k^n : n \ge 1\}$ (respectively, $\{\hat{S}^n : n \ge 1\}$) follows from the tightness of $\{\hat{Z}_{k,1}^n : n \ge 1\}$ in (6.8) and $\{\hat{Z}_{k,2}^n : n \ge 1\}$ in (6.9) (respectively, $\{\hat{V}_1^n : n \ge 1\}$ in (6.6) and $\{\hat{V}_2^n : n \ge 1\}$ in (6.7)), for $k = 1, ..., K$. The tightness of $\{\hat{M}_{k,1}^n : n \ge 1\}$, $\{\hat{Z}_{k,1}^n : n \ge 1\}$, $k = 1, ..., K$, and $\{\hat{V}_1^n : n \ge 1\}$ follows from the weak convergence of $(\hat{\boldsymbol{M}}_1^n, \hat{\boldsymbol{Z}}_1^n, \hat{V}_1^n)$ in (6.16). Then, it remains to show the tightness of $\{\hat{M}_{k,2}^n : n \ge 1\}$ in (6.5), $\{\hat{Z}_{k,2}^n : n \ge 1\}$ in (6.9) and $\{\hat{V}_2^n : n \ge 1\}$ in (6.7), $k = 1, ..., K$. The proof follows from two steps. First, we give decompositions of these processes into two processes, based on a decomposition of the multiparameter sequential empirical process $\hat{K}^n$. Second, we establish the tightness of each decomposed process via an appropriate tightness criterion. In particular, one decomposed process is dealt with using Aldous' sufficient condition for tightness (see, e.g., Theorem 16.10 in [10]) in Lemma B.3, and the second decomposed process is dealt with a sufficient condition for tightness (see, e.g., Lemma 3.32 of Chapter VI in [29]) in Lemma B.4.

We will first prove Lemma B.3, starting with some preliminary results below. Denote

$$\hat{R}^n(t,\boldsymbol{x}) := \hat{K}^n(\bar{A}^n(t),\boldsymbol{x}), \quad t\geq 0,\ \boldsymbol{x}\in\mathbb{R}_+^K.$$

From Lemma 4.5, for $t\geq 0$ and $\boldsymbol{x}\in\mathbb{R}_+^K$, $\hat{R}^n(t,\boldsymbol{x})$ can be written as

$$\hat{R}^n(t,\boldsymbol{x}) = \hat{W}_0^n(t,\boldsymbol{x}) + \hat{W}_1^n(t,\boldsymbol{x}), \tag{B.1}$$

where

$$\hat{W}_0^n(t,\boldsymbol{x}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)}\left(\mathbf{1}(\boldsymbol{\eta}^i\leq\boldsymbol{x}) - \int_{L_{\boldsymbol{x}}}\frac{\mathbf{1}(\boldsymbol{\eta}^i>\boldsymbol{u})}{F(T_{\boldsymbol{u}})}F(d\boldsymbol{u})\right), \tag{B.2}$$

and

$$\hat{W}_1^n(t,\boldsymbol{x}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)}\int_{L_{\boldsymbol{x}}}\left(\frac{\mathbf{1}(\boldsymbol{\eta}^i>\boldsymbol{u}) - F(T_{\boldsymbol{u}})}{F(T_{\boldsymbol{u}})}\right)F(d\boldsymbol{u}). \tag{B.3}$$

Therefore, by (B.1), (6.5), (6.9) and (6.7), we can decompose $\hat{M}_{k,2}^n$, $\hat{Z}_{k,2}^n$ and $\hat{V}_2^n$ as

$$\hat{M}_{k,2}^n(t) = -\hat{G}_k^{n,1}(t) - \hat{G}_k^{n,2}(t), \tag{B.4}$$
$$\hat{Z}_{k,2}^n(t) = \hat{I}_k^{n,1}(t) + \hat{I}_k^{n,2}(t), \tag{B.5}$$
$$\hat{V}_2^n(t) = \hat{H}^{n,1}(t) + \hat{H}^{n,2}(t), \tag{B.6}$$

for $t\geq 0$ and $k=1,...,K$, where

$$\hat{G}_k^{n,1}(t) := \int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_k\leq t)d\hat{W}_0^n(s,\boldsymbol{x}), \tag{B.7}$$
$$\hat{G}_k^{n,2}(t) := \int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_k\leq t)d\hat{W}_1^n(s,\boldsymbol{x}), \tag{B.8}$$
$$\hat{H}^{n,1}(t) := \int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_j\leq t,\ \forall j)d\hat{W}_0^n(s,\boldsymbol{x}), \tag{B.9}$$
$$\hat{H}^{n,2}(t) := \int_0^t\int_{\mathbb{R}_+^K}\mathbf{1}(s+x_j\leq t,\ \forall j)d\hat{W}_1^n(s,\boldsymbol{x}), \tag{B.10}$$
$$\hat{I}_k^{n,1}(t) := \int_0^t\int_{\mathbb{R}_+^K}(\mathbf{1}(s+x_k\leq t) - \mathbf{1}(s+x_j\leq t,\ \forall j))d\hat{W}_0^n(s,\boldsymbol{x}) = \hat{G}_k^{n,1}(t) - \hat{H}^{n,1}(t), \tag{B.11}$$
$$\hat{I}_k^{n,2}(t) := \int_0^t\int_{\mathbb{R}_+^K}(\mathbf{1}(s+x_k\leq t) - \mathbf{1}(s+x_j\leq t,\ \forall j))d\hat{W}_1^n(s,\boldsymbol{x}) = \hat{G}_k^{n,2}(t) - \hat{H}^{n,2}(t). \tag{B.12}$$

Therefore, for $k=1,...,K$, in order to show the tightness of $\{\hat{M}_{k,2}^n : n\geq 1\}$, it is sufficient to show the tightness of $\{\hat{G}_k^{n,1} : n\geq 1\}$ and $\{\hat{G}_k^{n,2} : n\geq 1\}$. Similarly, the tightness of $\{\hat{V}_2^n : n\geq 1\}$ follows from the tightness of $\{\hat{H}^{n,1} : n\geq 1\}$ and $\{\hat{H}^{n,2} : n\geq 1\}$. Then, by (B.11) and (B.12), the tightness of $\{\hat{Z}_{k,2}^n : n\geq 1\}$ follows directly, $k=1,...,K$. For each k, we start by proving the tightness of $\{\hat{G}_k^{n,1} : n\geq 1\}$ and $\{\hat{H}^{n,1} : n\geq 1\}$ in Lemma B.3, and then prove the tightness of $\{\hat{G}_k^{n,2} : n\geq 1\}$ and $\{\hat{H}^{n,2} : n\geq 1\}$ in Lemma B.4.

We first present some preliminary results to prove Lemma B.3. Denote, for $k=1,...,K$ and $i\in\mathbb{N}$,

$$\begin{aligned} G^{n,1}_{k,i}(t) &:= \mathbf{1}(0\le \eta^i_k \le t-\tau^n_i) - \int_0^{\eta^i_1}\cdots\int_0^{\eta^i_k\wedge(t-\tau^n_i)^+}\cdots\int_0^{\eta^i_K}\frac{1}{F(T_{\boldsymbol{u}})}F(d\boldsymbol{u}),\\ H^{n,1}_i(t) &:= \mathbf{1}(\mathbf{0}\le \boldsymbol{\eta}^i \le (t-\tau^n_i)\boldsymbol{e}) - \int_{L_{\boldsymbol{\eta}^i\wedge(t-\tau^n_i)^+\boldsymbol{e}}}\frac{1}{F(T_{\boldsymbol{u}})}F(d\boldsymbol{u}), \end{aligned} \tag{B.13}$$

where we recall $\boldsymbol{e}$ is the K-dimensional vector with one for each component. By (B.7) and (B.11), we obtain that

$$\hat{G}^{n,1}_k(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)} G^{n,1}_{k,i}(t), \quad \hat{H}^{n,1}(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)} H^{n,1}_i(t). \tag{B.14}$$

Let $\xi^n_i := \tau^n_i - \tau^n_{i-1}$ be the interarrival times between the $(i-1)^{\text{th}}$ and i^{th} jobs arriving to the system, $i\in\mathbb{N}$. Define the σ-fields by

$$\begin{aligned} \mathcal{G}^n_t(k) &:= \sigma(\mathbf{1}(\eta^i_k \le (s-\tau^n_i)),\ s\le t,\ i=1,...,A^n(t)) \vee \sigma(\eta^i_j,\ j\ne k,\ i=1,...,A^n(t)) \\ &\qquad \vee \sigma\{A^n(s),\ s\le t\}\vee\sigma(\xi^n_r, r\ge 1)\vee\mathcal{N}, \end{aligned} \tag{B.15}$$

$$\mathcal{H}^n_t := \sigma(\mathbf{1}(\boldsymbol{\eta}^i \le (s-\tau^n_i)\boldsymbol{e}),\ s\le t,\ i=1,...,A^n(t))\vee\sigma(A^n(s),\ s\le t)\vee\sigma(\xi^n_r, r\ge 1)\vee\mathcal{N}, \tag{B.16}$$

where $\mathcal{N}$ includes all the null sets.

It is easy to verify that $\mathcal{G}^n(k) := \{\mathcal{G}^n_t(k) : t\ge 0\}$ for $k=1,...,K$ and $\mathcal{H}^n := \{\mathcal{H}^n_t : t\ge 0\}$ are filtrations. We state the martingale properties of $\hat{G}^{n,1}_k(\cdot)$ for $k=1,...,K$ and $\hat{H}^{n,1}(\cdot)$ in Lemmas B.1 and B.2, respectively. Since the proofs of the two Lemmas are similar, we only prove Lemma B.2 in detail.

Lemma B.1. *The processes $\hat{G}^{n,1}_k := \{\hat{G}^{n,1}_k(t) : t\ge 0\}$ are $\mathcal{G}^n(k)$-square-integrable martingales for $k=1,...,K$.*

Lemma B.2. *The process $\hat{H}^{n,1} := \{\hat{H}^{n,1}(t) : t\ge 0\}$ is an $\mathcal{H}^n$-square-integrable martingale.*

Proof of Lemma B.2. From (B.14), it suffices to show that, for each $i\in\mathbb{N}$, the process $H^{n,1}_i = \{H^{n,1}_i(t) : t\ge 0\}$ is an $\mathcal{H}^n$-square-integrable martingale. By the construction of the filtration $\mathcal{H}^n$ in (B.16) and the definition of $H^{n,1}_i$, $H^{n,1}_i$ is $\mathcal{H}^n$-adapted. Note that, for every $t\ge 0$,

$$|H^{n,1}_i(t)| \le 1 + \int_{L_{\boldsymbol{\eta}^i}} \frac{1}{F(T_{\boldsymbol{u}})}F(d\boldsymbol{u}), \quad a.s.$$

By Lemma 4.3, we have $E[(H^{n,1}_i(t))^2] < \infty$, for $t\ge 0$. We next will show the martingale property for $H^{n,1}_i$, i.e., for $s<t$,

$$E[H^{n,1}_i(t)|\mathcal{H}^n_s] = H^{n,1}_i(s). \tag{B.17}$$

To show (B.17), it suffices to show

$$\mathbf{1}(\tau^n_i > s)E[H^{n,1}_i(t)|\mathcal{H}^n_s] = 0, \tag{B.18}$$

and

$$\mathbf{1}(\tau^n_i \le s)E[H^{n,1}_i(t)|\mathcal{H}^n_s] = H^{n,1}_i(s). \tag{B.19}$$

We first prove (B.18). By the construction of $\mathcal{H}^n$ in (B.16), τ^n_i is an $\mathcal{H}^n$-stopping time. Thus, the σ-field $\mathcal{H}^n_{\tau^n_i}$ is well-defined. Thus,

$$\mathbf{1}(\tau^n_i > s)E[H^{n,1}_i(t)|\mathcal{H}^n_s] = \mathbf{1}(\tau^n_i > s)E\left[E\left[H^{n,1}_i(t)|\mathcal{H}^n_{\tau^n_i}\right]|\mathcal{H}^n_s\right]. \tag{B.20}$$

Then, we claim that

$$E[H_i^{n,1}(t)|\mathcal{H}^n_{\tau_i^n}] = \frac{E[H_i^{n,1}(t)|\tau_i^n]}{P(\boldsymbol{\eta}^i > \mathbf{0}|\tau_i^n)} = 0, \tag{B.21}$$

where the last equality follows from (B.13) and the independence of $\boldsymbol{\eta}^i$ and τ_i^n. For the first equality in (B.21), by Lemma 3.6 in [37], it suffices to show

$$\mathcal{H}_{\tau_i^n} \cap \{\boldsymbol{\eta}^i > \mathbf{0}\} \subset (\sigma(\xi_r^n, r \geq 1) \vee \sigma(\boldsymbol{\eta}^r, r \geq 1, r \neq i) \vee \sigma(\tau_i^n) \vee \mathcal{N}) \cap \{\boldsymbol{\eta}^i > \mathbf{0}\}. \tag{B.22}$$

It suffices to check (B.22) for sets generating $\mathcal{H}^n_{\tau_i^n}$. By (B.16), we note that

$$\mathcal{H}^n_{\tau_i^n} = \sigma(\xi_r^n, r \geq 1) \vee \sigma(\tau_r^n, \mathbf{1}(\boldsymbol{\eta}^r \leq (s \wedge \tau_i^n - \tau_r^n)\boldsymbol{e}), s \geq 0, r = 1, ..., A^n(\tau_i^n)) \vee \mathcal{N}. \tag{B.23}$$

(Here we use, e.g., the argument in Appendix A.2 of Brémaud [12].) Then, for $s_1, ..., s_l > 0$, $l = i, i+1, ...$, $p = 1, 2, ...$, and Borel sets $Z_1, ..., Z_p$, B_r and C_r with $r = 1, ..., l$, since $A^n(\tau_i^n) \geq l > i$, then $\tau_r^n = \tau_i^n$, $r = i+1, ..., l$, we have

$$\begin{aligned}
&\left(\bigcap_{r=1}^{p}\{\xi_r^n \in Z_r\}\right) \cap \{A^n(\tau_i^n) \geq l\} \cap \left(\bigcap_{r=1}^{l}\{\tau_r^n \in B_r\}\right) \cap \left(\bigcap_{r=1}^{l}\{\mathbf{1}(\boldsymbol{\eta}^r \leq (s_r \wedge \tau_i^n - \tau_r^n)\boldsymbol{e}) \in C_r\}\right) \cap \{\boldsymbol{\eta}^i > \mathbf{0}\} \\
&= \left(\bigcap_{r=1}^{p}\{\xi_r^n \in Z_r\}\right) \cap \left(\bigcap_{r=i+1}^{l}\{\tau_i^n = \tau_r^n\}\right) \cap \left(\bigcap_{r=1}^{i-1}\{\tau_r^n \in B_r\}\right) \cap \left(\bigcap_{r=i}^{l}\{\tau_r^n \in B_r\}\right) \\
&\qquad \cap \left\{\bigcap_{r=1}^{i-1}\{\mathbf{1}(\boldsymbol{\eta}^r \leq (s_r \wedge \tau_i^n - \tau_r^n)\boldsymbol{e}) \in C_r\}\right\} \cap \{\boldsymbol{\eta}^i > \mathbf{0}\},
\end{aligned}$$

where $0 \in C_r$, $i \leq r \leq l$, and the LHS is Ø otherwise. We show that the event on the RHS of the previous equation is in $(\sigma(\xi_r^n, r \geq 1) \vee \sigma(\boldsymbol{\eta}^r, r \geq 1, r \neq i) \vee \sigma(\tau_i^n) \vee \mathcal{N}) \cap \{\boldsymbol{\eta}^i > \mathbf{0}\}$. It is enough to prove that this holds for the event $\bigcap_{r=i+1}^{l}\{\tau_i^n = \tau_r^n\} \cap \{\boldsymbol{\eta}^i > \mathbf{0}\}$. We then can proceed in the same way as the proof of Lemma 3.5 in [37], and we omit the details here for brevity. Thus, we have proved (B.18) holds.

We will next prove (B.19). Note that

$$\begin{aligned}
&\mathbf{1}(\tau_i^n \leq s)E[H_i^{n,1}(t)|\mathcal{H}_s^n] \\
&= \mathbf{1}(\boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e})E[H_i^{n,1}(t)|\mathcal{H}_s^n] + \mathbf{1}(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e})\mathbf{1}((s - \tau_i^n) \geq 0)E[H_i^{n,1}(t)|\mathcal{H}_s^n].
\end{aligned} \tag{B.24}$$

Since $\mathbf{1}(\boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e})$ and $\mathbf{1}(\mathbf{0} \leq \boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e})$ are $\mathcal{H}_s^n$-measurable, the first term on the RHS of (B.24) is

$$\begin{aligned}
&\mathbf{1}(\boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e})E[H_i^{n,1}(t)|\mathcal{H}_s^n] \\
&= \mathbf{1}(\mathbf{0} \leq \boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e}) - \mathbf{1}(\boldsymbol{\eta}^i \leq (s - \tau_i^n)\boldsymbol{e}) \int_{L_{\boldsymbol{\eta}^i \wedge (s - \tau_i^n) + \boldsymbol{e}}} \frac{1}{F(T_{\boldsymbol{u}})} F(d\boldsymbol{u}).
\end{aligned} \tag{B.25}$$

For the second term in (B.24), we claim that

$$\begin{aligned}
&\mathbf{1}(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e})\mathbf{1}((s - \tau_i^n) \geq 0)E[H_i^{n,1}(t)|\mathcal{H}_s^n] \\
&= \mathbf{1}(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e})\mathbf{1}((s - \tau_i^n) \geq 0)\frac{E[\mathbf{1}(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e})\mathbf{1}((s - \tau_i^n) \geq 0)H_i^{n,1}(t)|\tau_i^n]}{P(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e} \text{ and } s - \tau_i^n \geq 0|\tau_i^n)} && \text{(B.26)} \\
&= -\mathbf{1}(\boldsymbol{\eta}^i \nleq (s - \tau_i^n)\boldsymbol{e})\mathbf{1}((s - \tau_i^n) \geq 0) \int_{L_{\boldsymbol{\eta}^i \wedge (s - \tau_i^n)\boldsymbol{e}}} \frac{1}{F(T_{\boldsymbol{u}})} F(d\boldsymbol{u}), && \text{(B.27)}
\end{aligned}$$

where (B.27) follows from the definition of $\hat{H}_i^{n,1}$. For (B.26), by Lemma 3.6 in [37] and the fact that $\mathbf{1}(\boldsymbol{\eta}^i \not\leq (s-\tau_i^n)\boldsymbol{e})$ and $\mathbf{1}((s-\tau_i^n)\geq 0)$ are $\mathcal{H}_s^n$ measurable, it suffices to show

$$\mathcal{H}_s^n \cap \{\boldsymbol{\eta}^i \not\leq (s-\tau_i^n)\boldsymbol{e}, s-\tau_i^n \geq 0\} \subset (\sigma(\xi_r^n, r\geq 1)\vee \sigma(\boldsymbol{\eta}^r, r\geq 1, r\neq i)\vee \sigma(\tau_i^n)\vee \mathcal{N}) \cap \{\boldsymbol{\eta}^i \not\leq (s-\tau_i^n)\boldsymbol{e}, s-\tau_i^n \geq 0\}. \tag{B.28}$$

The verification of (B.28) is similar to the proof of Lemma 3.5 in [37], and we omit the details for brevity. Thus, we can see (B.19) holds. In summary, the martingale property of $H_i^{n,1}$ follows from (B.18) and (B.19). ☐

Now, we can prove the tightness of $\{\hat{G}_k^{n,1}: n\geq 1\}$, $k=1,...,K$, and $\{\hat{H}^{n,1}: n\geq 1\}$.

Lemma B.3. *The sequences $\{\hat{G}_k^{n,1}: n\geq 1\}$, $k=1,...,K$, and $\{\hat{H}^{n,1}: n\geq 1\}$ are tight in $\mathbb{D}$.*

Proof of Lemma B.3. We only show the tightness of $\{\hat{H}^{n,1}: n\geq 1\}$ in detail. The proof for the tightness for the sequence $\{\hat{G}_k^{n,1}: n\geq 1\}$, $k=1,...,K$, is similar.

By Aldous' sufficient condition for tightness (see, e.g., Theorem 16.10 in [10]), we need to verify the following: for $L>0$ and $\epsilon>0$,

$$\lim_{\kappa\to\infty}\limsup_{n\to\infty} P\left(\sup_{t\leq L}|\hat{H}^{n,1}(t)|>\kappa\right)=0, \tag{B.29}$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\tau\in\mathcal{C}_L^n} P\left(\sup_{0\leq t\leq\delta}|\hat{H}^{n,1}(\tau+t)-\hat{H}^{n,1}(\tau)|>\epsilon\right)=0, \tag{B.30}$$

where $\mathcal{C}_L^n$ is the set of all $\mathcal{H}^n$-stopping times bounded by L. Since the proofs of (B.29) and (B.30) are analogous, we only check (B.30) here. Note that, for $\kappa>0$ and $\delta<1$,

$$\begin{aligned}
&P\left(\sup_{0\leq t\leq\delta}|\hat{H}^{n,1}(\tau+t)-\hat{H}^{n,1}(t)|>\epsilon\right)\\
\leq\ & P\left(\sup_{0\leq t\leq\delta}|\hat{H}^{n,1}(\tau+t)-\hat{H}^{n,1}(t)|>\epsilon, \bar{A}^n(L+1)>\kappa\right)\\
&+P\left(\sup_{0\leq t\leq\delta}|\hat{H}^{n,1}(\tau+t)-\hat{H}^{n,1}(t)|>\epsilon, \bar{A}^n(L+1)\leq\kappa\right)\\
\leq\ & P\left(\bar{A}^n(L+1)>\kappa\right)+P\left(\sup_{0\leq t\leq\delta}|\hat{H}_{\kappa n}^{n,1}(\tau+t)-\hat{H}_{\kappa n}^{n,1}(t)|>\epsilon\right),
\end{aligned} \tag{B.31}$$

where $\hat{H}_{\kappa n}^{n,1}(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor \kappa n\wedge A^n(t)\rfloor} H_i^{n,1}(t)$ for $t\geq 0$. By Assumption 1, we see that

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

For the second term in (B.31), by Doob's optional stopping theorem and Lemma B.2, $\{\hat{H}_{\kappa n}^{n,1}(\tau+t)-\hat{H}_{\kappa n}^{n,1}(\tau): t\geq 0\}$ is a locally square-integrable martingale with respect to the filtration $\mathcal{H}^n$ for $\tau\in\mathcal{C}_L^n$. Thus, $\left\{|\hat{H}_{\kappa n}^{n,1}(\tau+t)-\hat{H}_{\kappa n}^{n,1}(\tau)|: t\geq 0\right\}$ is a submartingale with respect to the filtration $\mathcal{H}^n$. Thus, by Doob's inequality (see Theorem 9.2 of Chapter 1 in [43]), we have

$$P\left(\sup_{0\leq t\leq\delta}|\hat{H}_{\kappa n}^{n,1}(\tau+t)-\hat{H}_{\kappa n}^{n,1}(\tau)|>\epsilon\right)\leq\frac{1}{\epsilon^2}E[(\hat{H}_{\kappa n}^{n,1}(\tau+\delta)-\hat{H}_{\kappa n}^{n,1}(\tau))^2]. \tag{B.32}$$

By Lemma B.2 and the independence of $\boldsymbol{\eta}^i$ and $\boldsymbol{\eta}^j$ for $i\neq j$, we have

$$E[(\hat{H}_{\kappa n}^{n,1}(\tau+\delta)-\hat{H}_{\kappa n}^{n,1}(\tau))^2] \tag{B.33}$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{\lfloor A^n(\tau)\wedge\kappa n\rfloor}(H_i^{n,1}(\tau+\delta)-H_i^{n,1}(\tau))^2\right]+E\left[\frac{1}{n}\sum_{i=\lfloor A^n(\tau)\wedge\kappa n\rfloor+1}^{\lfloor A^n(\tau+\delta)\wedge\kappa n\rfloor}(H_i^{n,1}(\tau+\delta))^2\right] \tag{B.34}$$

$$\leq E\left[\frac{1}{n}\sum_{i=1}^{A^n(L)}\sup_{\substack{s\leq L\\0\leq t-s\leq\delta}}(H_i^{n,1}(t)-H_i^{n,1}(s))^2\right]+E\left[\frac{1}{n}\sup_{\substack{s\leq L\\0\leq t-s\leq\delta}}\sum_{i=A^n(s)+1}^{A^n(t)}(\tilde{H}_i^{n,1})^2\right], \tag{B.35}$$

where $\tilde{H}_i^{n,1} := 1+\int_{L_{\boldsymbol{\eta}^i}}\frac{1}{F(T_{\boldsymbol{u}})}F(d\boldsymbol{u})$. Denote $\sigma^2(\delta) := E\left[\sup_{s\leq L,0\leq t-s\leq\delta}(H_i^{n,1}(t)-H_i^{n,1}(s))^2\right]$. By the square integrability of $H_i^{n,1}$, we can see that $\sigma^2(\delta)<\infty$. By the dominated convergence theorem, we obtain that

$$\lim_{\delta\to 0}\sigma^2(\delta)=0. \tag{B.36}$$

Note that, by the FWLLN, for each $\delta>0$ and $0<L<\infty$,

$$\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\sup_{s\leq L,0\leq t-s\leq\delta}(H_i^{n,1}(t)-H_i^{n,1}(s))^2 \Rightarrow \sigma^2(\delta)u \quad \text{in} \quad \mathbb{D} \quad \text{as} \quad n\to\infty. \tag{B.37}$$

By Assumption 1, (B.36) and (B.37), we have

$$\lim_{\delta\to 0}\limsup_{n\to\infty}E\left[\frac{1}{n}\sum_{i=1}^{A^n(L)}\sup_{s\leq L,0\leq t-s\leq\delta}(H_i^{n,1}(t)-H_i^{n,1}(s))^2\right]=0.$$

Because of the $\mathbb{C}$-tightness of $\{\bar{A}^n : n\geq 1\}$ (see Lemma 3.2 in [37]), similarly, we also have

$$\lim_{\delta\to 0}\limsup_{n\to\infty}E\left[\frac{1}{n}\sup_{s\leq L,0\leq t-s\leq\delta}\sum_{i=A^n(s)+1}^{A^n(t)}(\tilde{H}_i^{n,1})^2\right]=0.$$

Therefore, we have shown (B.30), and the lemma is proved. □

To complete the tightness proof of $\{\hat{M}_{k,2}^n : n\geq 1\}$, $\{\hat{Z}_{k,2}^n : n\geq 1\}$ and $\{\hat{V}_2^n : n\geq 1\}$, the final step is to show that the sequences $\{\hat{G}_k^{n,2} : n\geq 1\}$, $k=1,...,K$, and $\{\hat{H}^{n,2} : n\geq 1\}$ are tight.

LEMMA B.4. *The sequences of processes $\{\hat{G}_k^{n,2} : n\geq 1\}$, $k=1,...,K$, and $\{\hat{H}^{n,2} : n\geq 1\}$ are tight in $\mathbb{D}$.*

Proof of Lemma B.4. Since the proof for the tightness of $\{\hat{G}_k^{n,2} : n\geq 1\}$ for $k=1,...,K$ is analogous to that of $\{\hat{H}^{n,2} : n\geq 1\}$, we will focus on the tightness proof of $\{\hat{H}^{n,2} : n\geq 1\}$ in detail.

Denote

$$\hat{T}^n(t,\boldsymbol{x}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}(\mathbf{1}(\boldsymbol{\eta}^i\geq\boldsymbol{x})-F(T_{\boldsymbol{x}})), \quad t\geq 0,\ \boldsymbol{x}\in\mathbb{R}_+^2.$$

By the definition of $\hat{H}^{n,2}$ in (B.10), we could rewrite it as

$$\hat{H}^{n,2}(t)=\int_{\mathbb{R}_+^K}\left[\hat{T}^n\left(\min_{1\leq j\leq K}\bar{A}^n(t-x_j),\boldsymbol{x}\right)\mathbf{1}(x_j\leq t,\forall j)(F(T_{\boldsymbol{x}}))^{-1}\right]F(d\boldsymbol{x}).$$

For each $\epsilon>0$ and $t\geq 0$, we can decompose $\hat{H}^{n,2}$ as

$$\hat{H}^{n,2}(t)=\hat{H}_1^{n,2,\epsilon}(t)+\hat{H}_2^{n,2,\epsilon}(t),$$

where

$$\hat{H}_1^{n,2,\epsilon}(t) = \int_{\mathbb{R}_+^K} \left[\hat{T}^n \left(\min_{1\le j\le K} \bar{A}^n(t-x_j), \boldsymbol{x} \right) \mathbf{1}(x_j \le t, \forall j) \mathbf{1}(F(T_{\boldsymbol{x}}) \ge \epsilon)(F(T_{\boldsymbol{x}}))^{-1} \right] F(d\boldsymbol{x}),$$
$$\hat{H}_2^{n,2,\epsilon}(t) = \int_{\mathbb{R}_+^K} \left[\hat{T}^n \left(\min_{1\le j\le K} \bar{A}^n(t-x_j), \boldsymbol{x} \right) \mathbf{1}(x_j \le t, \forall j) \mathbf{1}(F(T_{\boldsymbol{x}}) < \epsilon)(F(T_{\boldsymbol{x}}))^{-1} \right] F(d\boldsymbol{x}).$$

To show $\{\hat{H}^{n,2} : n \ge 1\}$ is tight, it suffices to prove that the following two conditions hold (see Lemma 3.32 of Chapter VI in [29]): for each $\delta > 0$ and $T > 0$,
1. $\{\hat{H}_1^{n,2,\epsilon} : n \ge 1\}$ is tight,

2. $\lim_{\epsilon\to 0} \limsup_{n\to\infty} P\left(\|\hat{H}_2^{n,2,\epsilon}\|_T > \delta \right) = 0$, where $\|\hat{H}_2^{n,2,\epsilon}\|_T := \sup_{0\le t\le T} |\hat{H}_2^{n,2,\epsilon}(t)|$.

We start by verifying condition 1. Note that $\hat{H}_1^{n,2,\epsilon}$ is the integral representation of $\hat{T}^n(\bar{A}^n(\cdot), \cdot)$. If we can show the weak convergence of $\hat{T}^n(\bar{A}^n(\cdot), \cdot)$ and the mapping defined by this integral representation is continuous in $\mathbb{D}$, condition 1 will follow from the continuous mapping theorem. We first show the weak convergence of $\hat{T}^n(\cdot,\cdot)$. Note that the process $\hat{T}^n(\cdot,\cdot)$ is similar to the multiparameter sequential empirical process $\hat{K}^n(\cdot,\cdot)$ defined in (3.3), which focuses on the lower tail of the random vector. If we can represent $\hat{T}^n(\cdot,\cdot)$ in terms of the multiparameter sequential empirical process $\hat{K}^n(\cdot,\cdot)$, the weak convergence of $\hat{T}^n(\cdot,\cdot)$ follows. The following lemma is a key observation to establish the relationship between $\hat{T}^n(\cdot,\cdot)$ and $\hat{K}^n(\cdot,\cdot)$.

Lemma B.5. *For each i,*

$$\mathbf{1}(\boldsymbol{\eta}^i > \boldsymbol{x}) = 1 + \sum_{k=1}^{K} (-1)^k \sum_{1\le j_1<\ldots<j_k\le K} \mathbf{1}(\eta_{j_l}^i \le x_{j_l},\ l = 1,\ldots,k), \quad a.s. \tag{B.38}$$

Proof of Lemma B.5. We use the induction method on the dimension K to prove the lemma. When $K = 1$, (B.38) holds obviously. Assume that (B.38) holds when $K = N$. When $K = N + 1$, we notice that

$$\mathbf{1}(\eta_j^i > x_j, j = 1,\ldots,N+1) = \mathbf{1}(\eta_j^i > x_j, j = 1,\ldots,N) - \mathbf{1}(\eta_1^i > x_1, \ldots, \eta_N^i > x_N, \eta_{N+1}^i \le x_{N+1}).$$

For the two terms on the right side of the previous equation, using (B.38) on the first term and on the first N components of the second term by induction, we have

$$\begin{aligned}
&\mathbf{1}(\eta_j^i > x_j, j = 1,\ldots,N+1) \\
&= 1 + \sum_{k=1}^{N} (-1)^k \sum_{1\le j_1<\ldots<j_k\le N} \mathbf{1}(\eta_{j_l}^i \le x_{j_l},\ l = 1,\ldots,k) \\
&\quad - \left(\mathbf{1}(\eta_{N+1}^i \le x_{N+1}) + \sum_{k=1}^{N} (-1)^k \sum_{1\le j_1<\ldots<j_k\le N} \mathbf{1}(\eta_{j_l}^i \le x_{j_l},\ l = 1,\ldots,k) \mathbf{1}(\eta_{N+1}^i \le x_{N+1}) \right) \\
&= 1 + \sum_{k=1}^{N+1} (-1)^k \sum_{1\le j_1<\ldots<j_k\le N+1} \mathbf{1}(\eta_{j_l}^i \le x_{j_l},\ l = 1,\ldots,k).
\end{aligned}$$

Therefore, (B.38) holds for $K = N + 1$. In summary, we see that (B.38) holds. □

By Lemma B.5, $\hat{T}^n(\cdot,\cdot)$ can be represented by the multiparameter sequential empirical processes driven by the vector and the subvector of the parallel service times. Following the proof of Theorem 3.1, we can show the weak convergence of $\hat{T}^n(\cdot,\cdot)$. Then, we state the result in Lemma B.6 and omit the proof for brevity.

LEMMA B.6.
$$\hat{T}^n(t,\boldsymbol{x}) \Rightarrow \hat{T}(t,\boldsymbol{x}) \quad in \quad \mathbb{D}([0,\infty),\mathbb{D}_K) \quad as \quad n\to\infty,$$
where $\hat{T}(t,\boldsymbol{x})$ is a continuous Gaussian random field with mean function $E[\hat{T}(t,\boldsymbol{x})]=0$ and covariance function

$$Cov\left(\hat{T}(t,\boldsymbol{x}),\hat{T}(s,\boldsymbol{y})\right) = (t\wedge s)(F(T_{\boldsymbol{x}\vee\boldsymbol{y}}) - F(T_{\boldsymbol{x}})F(T_{\boldsymbol{y}})), \quad t\geq 0,\ \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^K_+.$$

With the weak convergence of $\hat{T}^n(\cdot,\cdot)$, the next step is to establish the weak convergence of $\hat{T}^n(\bar{A}^n(\cdot),\cdot)$. Define the composition mapping $\circ_1 : \mathbb{D}([0,\infty),\mathbb{D}_K)\times\mathbb{D}_\uparrow \to \mathbb{D}([0,\infty),\mathbb{D}_K)$ by

$$x\circ_1 y(s,\boldsymbol{t}) := x(y(s),\boldsymbol{t}), \quad \text{for} \quad s\geq 0,\ \boldsymbol{t}\in\mathbb{R}^K_+,\ x\in\mathbb{D}([0,\infty),\mathbb{D}_K) \text{ and } y\in\mathbb{D}_\uparrow. \tag{B.39}$$

In order to show the weak convergence of $\hat{T}^n(\bar{A}^n(\cdot),\cdot)$, by (2.4) and Lemma B.6, it suffices to show the mapping $\circ_1$ is continuous, which is provided in Lemma B.7.

LEMMA B.7. *The mapping $\circ_1$ defined in (B.39) is continuous in $\mathbb{D}([0,\infty),\mathbb{D}_K)$ at each $(x,y)\in\mathbb{C}([0,\infty),\mathbb{C}_K)\times\mathbb{C}_\uparrow$.*

Proof of Lemma B.7. For each $(x,y)\in\mathbb{C}([0,\infty),\mathbb{C}_K)\times\mathbb{C}_\uparrow$, $T>0$ and a bounded closed set $\mathcal{A}\subset\mathbb{R}^K_+$, we assume that $(x_n,y_n)\in\mathbb{D}([0,\infty),\mathbb{D}_K)\times\mathbb{D}_\uparrow$ satisfies

$$(x_n,y_n)\to(x,y) \quad \text{in} \quad (\mathbb{D}([0,\infty),\mathbb{D}_K),\|\cdot\|_{T,\mathcal{A}}) \quad \text{as} \quad n\to\ \infty. \tag{B.40}$$

Note that (B.40) also implies that

$$x_n\to x \quad \text{in} \quad (\mathbb{D}([0,\infty),\mathbb{D}_K),\|\cdot\|_{T,\mathcal{A}}) \quad \text{as} \quad n\to\ \infty, \tag{B.41}$$

and
$$y_n\to y \quad \text{in} \quad (\mathbb{D}_\uparrow,\|\cdot\|_T) \quad \text{as} \quad n\to\infty. \tag{B.42}$$
By Proposition 5.2 in [23] and Lemma 2.1 in [63], it suffices to show

$$\|x_n\circ_1 y_n - x\circ_1 y\|_{T,\mathcal{A}}\to 0 \quad \text{as} \quad n\to\infty. \tag{B.43}$$

By the triangle inequality, we have

$$\begin{aligned}\|x_n\circ_1 y_n - x\circ_1 y\|_{T,\mathcal{A}} \leq& \|x_n\circ_1 y_n - x\circ_1 y_n\|_{T,\mathcal{A}} + \|x\circ_1 y_n - x\circ_1 y\|_{T,\mathcal{A}},\\ =&\|x_n - x\|_{T',\mathcal{A}} + \|x\circ_1 y_n - x\circ_1 y\|_{T,\mathcal{A}},\end{aligned} \tag{B.44}$$

where $T'=\sup_n\|y_n\|_T$. The first term in (B.44) converges to 0 by the convergence of x_n in (B.41). Noticing that x is continuous and the convergence of y_n in (B.42), the second term in (B.44) converges to 0 as $n\to\infty$. □

From Lemma B.7 and the continuous mapping theorem, we obtain the weak convergence of $\{\hat{T}^n(\bar{A}^n(\cdot),\cdot): n\geq 1\}$ in $\mathbb{D}([0,\infty),\mathbb{D}_K)$, which is summarized in Lemma B.8.

LEMMA B.8.

$$\hat{T}^n\circ_1\bar{A}^n(t,\boldsymbol{x}) \Rightarrow \hat{T}\circ_1\bar{a}(t,\boldsymbol{x}) \quad in \quad \mathbb{D}([0,\infty),\mathbb{D}_K) \quad as \quad n\to\infty, \tag{B.45}$$

where $\hat{T}(\cdot,\cdot)$ is defined in Lemma B.6.

Next, we focus on the proof of the continuity of the integration mappings. Define the mapping $h:\mathbb{D}([0,\infty),\mathbb{D}_K)\to\mathbb{D}[0,\infty)$ by

$$h(u)(t) := \int_{\mathbb{R}_+^K}\left[u\left(\min_{j=1,\dots,K}(t-x_j),\boldsymbol{x}\right)\mathbf{1}(x_j\le t,\forall j)\mathbf{1}(F(T_{\boldsymbol{x}})\ge\epsilon)(F(T_{\boldsymbol{x}}))^{-1}\right]F(d\boldsymbol{x}), \tag{B.46}$$

where $t\ge 0$ and $u\in\mathbb{D}([0,\infty),\mathbb{D}_K)$. Lemma B.9 shows the continuity of the mapping h.

Lemma B.9. *The mapping h defined in (B.46) is continuous at u, where $u\in\mathbb{C}([0,\infty),\mathbb{C}_K)$.*

Proof of Lemma B.9. Let $u\in\mathbb{C}([0,\infty),\mathbb{C}_K)$. Assume $u_n\in\mathbb{D}([0,\infty),\mathbb{D}_K)$ satisfying

$$u_n\to u \quad \text{in} \quad \mathbb{D}([0,\infty),\mathbb{D}_K) \quad \text{as} \quad n\to\infty. \tag{B.47}$$

Since u is continuous, by the definition of h, $h(u)(\cdot)$ is also continuous. Let $T>0$. In order to show the continuity of the mapping h, it suffices to show

$$\|h(u_n)(\cdot)-h(u)(\cdot)\|_T\to 0 \quad \text{as} \quad n\to\infty. \tag{B.48}$$

Let the set $\mathcal{A}:=\{\boldsymbol{x}\in\mathbb{R}_+^K : F(T_{\boldsymbol{x}})\ge\epsilon\}$, we have

$$\begin{aligned}
&\|h(u_n)(\cdot)-h(u)(\cdot)\|_T\\
&\quad=\sup_{t\le T}\Big|\int_{\mathbb{R}_+^K}\Big[(u_n(\min_{j=1,\dots,K}(t-x_j),\boldsymbol{x})-u(\min_{j=1,\dots,K}(t-x_j),\boldsymbol{x}))\mathbf{1}(x_j\le t,\forall j)\\
&\qquad\qquad\times\mathbf{1}(F(T_{\boldsymbol{x}})\ge\epsilon))(F(T_{\boldsymbol{x}}))^{-1}\Big]F(d\boldsymbol{x})\Big|\\
&\le\|u_n-u\|_{T,\mathcal{A}}\int_{\mathcal{A}}\frac{\mathbf{1}(F(T_{\boldsymbol{x}})\ge\epsilon)}{F(T_{\boldsymbol{x}})}F(d\boldsymbol{x}),\\
&\le\|u_n-u\|_{T,\mathcal{A}}\epsilon^{-1}P(\mathcal{A}).
\end{aligned} \tag{B.49}$$

By the convergence of u_n, we see that the RHS of (B.49) converges to 0 as $n\to\infty$. Thus, the mapping h is continuous in $\mathbb{D}([0,\infty),\mathbb{D}_K)$. □

Combining (B.45) and Lemma B.9, we obtain that $\{\hat{H}_1^{n,2,\epsilon}:n\ge1\}$ weak converges from the continuous mapping theorem, implying that $\{\hat{H}_1^{n,2,\epsilon}:n\ge1\}$ is tight in $\mathbb{D}$. Therefore, condition 1 is verified.

Now, we start to prove that condition 2 holds. Note that, for any $\delta>0$, $T>0$ and $\kappa>0$,

$$\begin{aligned}
&P\Big(\sup_{t\le T}\Big|\int_{\mathbb{R}_+^K}\Big[\hat{T}^n(\min_{1\le j\le K}\bar{A}^n(t-x_j),\boldsymbol{x})\mathbf{1}(x_j\le t,\forall j)\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)(F(T_{\boldsymbol{x}}))^{-1}\Big]F(d\boldsymbol{x})\Big|>\delta\Big)\\
&\le P\Big(\sup_{t\le T}\Big|\int_{\mathbb{R}_+^K}\Big[\hat{T}^n(\min_{1\le j\le K}\bar{A}^n(t-x_j),\boldsymbol{x})\mathbf{1}(x_j\le t,\forall j)\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)(F(T_{\boldsymbol{x}}))^{-1}\Big]F(d\boldsymbol{x})\Big|>\delta,\\
&\qquad\qquad\bar{A}^n(T)\le\kappa T\Big)+P\left(\bar{A}^n(T)>\kappa T\right)\\
&\le P\Big(\Big|\int_{\mathbb{R}_+^K}\frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{F(T_{\boldsymbol{x}})}\sup_{t\le\kappa T}|\hat{T}^n(t,\boldsymbol{x})|F(d\boldsymbol{x})\Big|>\delta\Big)+P\left(\bar{A}^n(T)>\kappa T\right).
\end{aligned} \tag{B.50}$$

By Assumption 1, we see that the second term on the RHS of (B.50) converges to 0 as $n\to\infty$ for sufficiently large κ. Thus, by Markov inequality and Fubini's theorem, we only need to show the following

$$\lim_{\epsilon\to0}\limsup_{n\to\infty}\int_{\mathbb{R}_+^K}\frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{F(T_{\boldsymbol{x}})}E\left[\sup_{t\le\kappa T}|\hat{T}^n(t,\boldsymbol{x})|\right]F(d\boldsymbol{x})=0.$$

Notice that, for any fixed $\boldsymbol{x} \in \mathbb{R}_+^K$, $\{\hat{T}^n(t,\boldsymbol{x}) : t \geq 0\}$ is a locally square-integrable martingale with respect to the filtration generated by itself, and it has the predictable quadratic-variation process (see Chapter 3 in [29]), $\langle \hat{T}^n(\cdot,\boldsymbol{x})\rangle(t) = \frac{\lfloor nt \rfloor}{n} F(T_{\boldsymbol{x}})(1-F(T_{\boldsymbol{x}}))$. By Theorem 1.9.5 in [43], we have

$$E\left[\sup_{t\leq \kappa T} |\hat{T}^n(t,\boldsymbol{x})|\right] \leq 3E(\langle \hat{T}^n(\cdot,\boldsymbol{x})\rangle(\kappa T))^{\frac{1}{2}} \leq 3\sqrt{\kappa T}(F(T_{\boldsymbol{x}})(1-F(T_{\boldsymbol{x}})))^{\frac{1}{2}}.$$

Thus,

$$\int_{\mathbb{R}_+^K} \frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{F(T_{\boldsymbol{x}})} E\left[\sup_{t\leq\kappa T} |\hat{T}^n(t,\boldsymbol{x})|\right] F(d\boldsymbol{x}) \leq 3\sqrt{\kappa T}\int_{\mathbb{R}_+^K} \frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{(F(T_{\boldsymbol{x}}))^{\frac{1}{2}}} F(d\boldsymbol{x}).$$

By Lemma B.5 and the definitions of multivariate integral and increments in $\mathbb{R}_+^K$ (see (5.1)), we obtain that

$$\int_{\mathbb{R}_+^K} \frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{(F(T_{\boldsymbol{x}}))^{\frac{1}{2}}} F(d\boldsymbol{x}) = (-1)^K \int_{\mathbb{R}_+^K} \frac{\mathbf{1}(F(T_{\boldsymbol{x}})<\epsilon)}{(F(T_{\boldsymbol{x}}))^{\frac{1}{2}}} F^c(d\boldsymbol{x}). \tag{B.51}$$

where $F^c(\boldsymbol{x}) := P(\boldsymbol{\eta}^i > \boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}_+^K$. The RHS of (B.51) goes to 0 as $\epsilon \to 0$ by the monotone convergence theorem, if we show $\left|\int_{\mathbb{R}_+^K} (F(T_{\boldsymbol{x}}))^{-\frac{1}{2}} F^c(d\boldsymbol{x})\right| < \infty$. This can be easily verified. Thus the proof is complete. □

Acknowledgments. This research is supported, in part, by the Army Research Office (W911NF-14-1-0019) and Marcus Endowment Grant at Penn State University. We thank the Statistical and Applied Mathematical Sciences Institute (SAMSI) for sponsoring the 2012-13 Program on Data-Driven Decisions in Healthcare, where this work was motivated and originated from. We also thank the participants of the workshops of the Program and the working group on fork-join networks for many helpful discussions. We particularly thank Ward Whitt, Avi Mandelbaum, Jim Dai, Amy Ward and Ohad Perry for many helpful discussions and comments on the paper.

References

[1] Antkiewicz, R., A. Gasecki, A. Najgebauer, D. Pierzchała, Z. Tarapata. 2010. Stochastic PERT and CAST logic approach for computer support of complex operation planning. *Analytical and Stochastic Modeling Techniques and Applications*, Vol. 6158. Springer, 159–173.

[2] Aras, A. K., Y. Liu, W. Whitt. 2014. Heavy-traffic limit for the initial content process. Submitted.

[3] Armony, M., S. Israelit, A. Mandelbaum, Y. N. Marmor, Y. Tseytlin, G. B. Yom-Tov. 2011. Patient flow in hospitals: A data-based queueing-science perspective. Preprint.

[4] Atar, R., A. Mandelbaum, A. Zviran. 2012. Control of fork-join networks in heavy traffic. *2012 50th Annual Allerton Conference on Communication, Control, and Computing*. 823–830.

[5] Avi-Itzhak, B., S. Halfin. 1991. Non-preemptive priorities in simple fork-join queues. *Queueing, Performance and Control in ATM* **1**(3).

[6] Baccelli, F., M. Lelarge, S. Foss. 2004. Asymptotics of subexponential max plus networks: The stochastic event graph case. *Queueing Systems* **46**(1-2) 75–96.

[7] Baccelli, F., A. M. Makowski, A. Shwartz. 1989. The fork-join queue and related systems with synchronization constraints: Stochastic ordering and computable bounds. *Advances in Applied Probability* **21**(3) 629–660.

[8] Baccelli, F., W. A. Massey, D. Towsley. 1989. Acyclic fork-join queuing networks. *Journal of the ACM* **36**(3) 615–642.

[9] Bickel, P. J., M. J. Wichura. 1971. Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics* **42** 1656–1670.

[10] Billingsley, P. 2009. *Convergence of Probability Measures*. John Wiley & Sons.

[11] Bouzebda, S., T. Zari. 2013. Strong approximation of empirical copula processes by Gaussian processes. *Statistics* **47**(5) 1047–1063.

[12] Brémaud, P. 1981. *Point Processes and Queues: Martingale Dynamics*. Springer.

[13] Csörgő, M., L. Horváth. 1988. A note on strong approximations of multivariate empirical processes. *Stochastic Processes and their Applications* **28**(1) 101–109.

[14] Csörgő, M, P. Révész. 1975. A strong approximation of the multivariate empirical process. *Studia Sci. Math. Hungar* **10** 427–434.

[15] Dai, H. 2011. Exact Monte Carlo simulation for fork-join networks. *Advances in Applied Probability* **43**(2) 484–503.

[16] Dean, J., S. Ghemawat. 2008. MapReduce: Simplified data processing on large clusters. *Communications of the ACM* **51**(1) 107–113.

[17] Dedecker, J., F. Merlevède, E. Rio. 2014. Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab.* **19**(9) 1–56.

[18] Dieker, A. B., M. Lelarge. 2006. Tails for (max, plus) recursions under subexponentiality. *Queueing Systems* **53**(4) 213–230.

[19] Doukhan, P., J. D. Fermanian, G. Lang. 2009. An empirical central limit theorem with applications to copulas under weak dependence. *Statistical Inference for Stochastic Processes* **12**(1) 65–87.

[20] Durieu, O., M. Tusche. 2014. An empirical process central limit theorem for multidimensional dependent data. *Journal of Theoretical Probability* **27**(1) 249–277.

[21] Durrett, R. 2010. *Probability: Theory and Examples*. Cambridge University Press.

[22] Eick, S. G., W. A. Massey, W. Whitt. 1993. The physics of the $M_t/G/\infty$ queue. *Operations Research* **41**(4) 731–742.

[23] Ethier, S. N., T. G. Kurtz. 2009. *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

[24] Gallien, J., L. M. Wein. 2001. A simple and effective component procurement policy for stochastic assembly systems. *Queueing Systems* **38**(2) 221–248.

[25] Gurvich, I., A. Ward. 2014. On the dynamic control of matching queues. *Stochastic Systems* **4** 1–45.

[26] Huang, J., A. Mandelbaum, H. Zhang, J. Zhang. 2014. Refined models for efficiency-driven queues with applications to delay announcements and staffing. Working paper.

[27] Ivanoff, B. G. 1980. The function space $D([0,\infty)^q, E)$. *Canadian Journal of Statistics* **8**(2) 179–191.

[28] Ivanoff, B. G., E. Merzbach. 1999. *Set-Indexed Martingales*. CRC Press.

[29] Jacod, J., A. N. Shiryaev. 1987. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin.

[30] Jean-Marie, A., L. Gün. 1993. Parallel queues with resequencing. *Journal of the ACM* **40**(5) 1188–1208.

[31] Jiang, L., R. E. Giachetti. 2008. A queueing network model to analyze the impact of parallelization of care on patient cycle time. *Health Care Management Science* **11**(3) 248–261.

[32] Karatzas, I., S. E. Shreve. 1991. *Brownian Motion and Stochastic Calculus*. Springer.

[33] Khoshnevisan, D. 2002. *Multiparameter Processes: An Introduction to Random Fields*. Springer.

[34] Klementowski, L. J. 1978. PERT/CPM and Supplementary Analytical Techniques. An Analysis of Aerospace Usage. Technical report, DTIC Document.

[35] Ko, S.-S., R. F. Serfozo. 2004. Response times in $M/M/s$ fork-join networks. *Advances in Applied Probability* **36**(3) 854–871.

[36] Ko, S.-S., R. F. Serfozo. 2008. Sojourn times in $G/M/1$ fork-join networks. *Naval Research Logistics* **55**(5) 432–443.

[37] Krichagina, E. V., A. A. Puhalskii. 1997. A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. *Queueing Systems* **25**(1-4) 235–280.

[38] Larson, R. C., M. F. Cahn, M. C. Shell. 1993. Improving the New York City arrest-to-arraignment system. *Interfaces* **23**(1) 76–96.

[39] Lawler, G. F., V. Limic. 2010. *Random Walk: A Modern Introduction*. Cambridge University Press.

[40] Leadbetter, M. R., G. Lindgren, H. Rootzén. 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.

[41] Leite, S. C., M. D. Fragoso. 2008. Heavy traffic analysis of state-dependent fork-join queues with triggers. *International Symposium on Performance Evaluation of Computer and Telecommunication Systems*. 488–494.

[42] Lin, M., L. Zhang, A. Wierman, J. Tan. 2013. Joint optimization of overlapping phases in MapReduce. *Performance Evaluation* **70**(10) 720–735.

[43] Lipster, R. S., A. N. Shiryaev. 1989. *Theory of Martingales*. Kluwer.

[44] Lu, H., G. Pang. 2014. Heavy-traffic limits for a fork-join network in the Halfin-Whitt regime. Preprint.

[45] Mandelbaum, A., P. Momčilović. 2008. Queues with many servers: the virtual waiting-time process in the QED regime. *Mathematics of Operations Research* **33**(3) 561–586.

[46] Mandelbaum, A., P. Momčilović. 2012. Queues with many servers and impatient customers. *Mathematics of Operations Research* **37**(1) 41–65.

[47] Marshall, A. W., I. Olkin. 1967. A multivariate exponential distribution. *Journal of the American Statistical Association* **62**(317) 30–44.

[48] Neuhaus, G. 1971. On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics* **42**(4) 1285–1295.

[49] Nguyen, V. 1993. Processing networks with parallel and sequential tasks: heavy traffic analysis and Brownian limits. *Annals of Applied Probability* **3**(1) 28–55.

[50] Nguyen, V. 1994. The trouble with diversity: fork-join networks with heterogeneous customer population. *Annals of Applied Probability* **4**(1) 1–25.

[51] Pang, G., W. Whitt. 2010. Two-parameter heavy-traffic limits for infinite-server queues. *Queueing Systems* **65**(4) 325–364.

[52] Pang, G., W. Whitt. 2012. Infinite-server queues with batch arrivals and dependent service times. *Probability in the Engineering and Informational Sciences* **26**(2) 197–220.

[53] Pang, G., W. Whitt. 2013. Two-parameter heavy-traffic limits for infinite-server queues with dependent service times. *Queueing Systems* **73**(2) 119–146.

[54] Philipp, W., L. Pinzur. 1980. Almost sure approximation theorems for the multivariate empirical process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **54**(1) 1–13.

[55] Pinotsi, D., M. A. Zazanis. 2005. Synchronized queues with deterministic arrivals. *Operations Research Letters* **33**(6) 560–566.

[56] Prabhakar, B., N. Bambos, T. S. Mountford. 2000. The synchronization of Poisson processes and queueing networks with service and synchronization nodes. *Advances in Applied Probability* **32**(3) 824–843.

[57] Puhalskii, A. A., J. E. Reed. 2010. On many-server queues in heavy traffic. *Annals of Applied probability* **20**(1) 129–195.

[58] Reed, J. E. 2009. The $G/GI/N$ queue in the Halfin-Whitt regime. *Annals of Applied Probability* **19**(6) 2211–2269.

[59] Sklar, A. 1959. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

[60] Squillante, M. S., Y. Zhang, A. Sivasubramaniam, N. Gautam. 2008. Generalized parallel-server fork-join queues with dynamic task scheduling. *Annals of Operations Research* **160**(1) 227–255.

[61] Straf, M. L. 1972. Weak convergence of stochastic processes with several parameters. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2. 187–221.

[62] Takahashi, M., H. Ōsawa, T. Fujisawa. 2000. On a synchronization queue with two finite buffers. *Queueing Systems* **36**(1-3) 107–123.

[63] Talreja, R., W. Whitt. 2009. Heavy-traffic limits for waiting times in many-server queues with abandonment. *Annals of Applied Probability* **19**(6) 2137–2175.

[64] Tan, J., X. Meng, L. Zhang. 2012. Delay tails in MapReduce scheduling. *ACM SIGMETRICS Performance Evaluation Review* **40**(1) 5–16.

[65] Tan, X., C. Knessl. 1996. A fork-join queueing model: Diffusion approximation, integral representations and asymptotics. *Queueing Systems* **22**(3-4) 287–322.

[66] Varki, E., L. W. Dowdy. 1996. Analysis of balanced fork-join queueing networks. *ACM SIGMETRICS Performance Evaluation Review* **24**(1) 232–241.

[67] Varma, S. 1990. Heavy and light traffic approximations for queues with synchronization constraints. Ph.D. thesis, Department of Electrical Engineering, University of Maryland.

[68] Wang, W., K. Zhu, L. Ying, J. Tan, L. Zhang. 2013. Map task scheduling in MapReduce with data locality: Throughput and heavy-traffic optimality. *INFOCOM 2013 Proceedings*. 1609–1617.

[69] Whitt, W. 2002. *Stochastic-Process Limits*. Springer.

[70] Willits, C. J., D. C. Dietz. 2001. Nested fork-join queueing network model for analysis of airfield operations. *Journal of Aircraft* **38**(5) 848–855.

[71] Zaied, I. 2012. The offered load in fork-join networks: Calculations and applications to service engineering of emergency departments. Master's thesis, Technion-Israel Institute of Technology, Faculty of Industrial and Management Engineering.

[72] Zviran, A. 2011. Fork-join networks in heavy traffic: Diffusion approximations and control. Master's thesis, Technion-Israel Institute of Technology, Faculty of Industrial and Management Engineering.