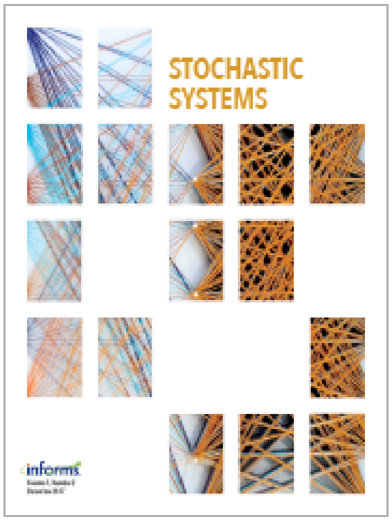


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Analysis of a Large-Scale Machine Repair Model

Petar Momčilović, Amir Motaei

To cite this article:
Petar Momčilović, Amir Motaei (2018) An Analysis of a Large-Scale Machine Repair Model. Stochastic Systems 8(2):91-125. https://doi.org/10.1287/stsy.2018.0010

STOCHASTIC SYSTEMS
Vol. 8, No. 2, June 2018, pp. 91–125
ISSN 1946-5238 (print), ISSN 1946-5238 (online)
http://pubsonline.informs.org/journal/stsy/

An Analysis of a Large-Scale Machine Repair Model

Petar Momčilović,[a] Amir Motaei[a]

[a] Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611
Contact: petar@ise.ufl.edu (PM); motaei@ufl.edu, http://orcid.org/0000-0002-5038-8934 (AM)

Received: November 2017
Revised: January 2018
Accepted: January 2018
Published Online: March 9, 2018

MSC2010 Subject Classification: Primary: 60K25; secondary: 60F17

https://doi.org/10.1287/stsy.2018.0010

Abstract. A machine repair model under general operating/repair distributions is considered in the Quality-and-Efficiency Driven asymptotic (QED) regime: both the number of machines and the number of repairmen are large, while the capacity and offered load relate via the square-root staffing rule. Process-level convergence of the number of broken machines is established—the limit is in terms of the corresponding tractable infinite-repairmen process, a stationary centered Gaussian process.

Funding: This work was supported in part by the National Science Foundation [Award CMMI-1362630].

Keywords: machine repair model • QED regime • heavy traffic

1. Introduction

The machine repair model is a classic operations research model. The system consists of a finite number of machines and repairmen. A machine operates for a random amount of time before it breaks down, and it takes a repairman a random amount of time to fix a machine. Once a machine is repaired, it becomes operational, and the cycle repeats. The machine repair model is relevant in describing systems with a finite population of customers that repeatedly return for service. Examples include systems based on membership/subscription (de Véricourt and Jennings 2011), professional and warranty services, and computer networks (de Véricourt and Jennings 2008). In the health care context, one can think of patients recovering in a hospital bed and requiring assistance from medical personnel multiple times throughout the day (the number of occupied beds is relatively constant during a single shift in some medical departments). Outpatient facilities providing dental, oncology and dialysis services can modeled under this framework as well, albeit at a different time scale. Many primary care physicians manage a finite number of patients (Green et al. 2007) (panel size), and thus our model can be relevant in that setting as well.

We focus on a large-scale (many machines and many repairmen) system operating in the Quality-and-Efficiency-Driven (QED) regime. Informally, under such a regime, the probabilities that a machine finds an available repairman and a repairman finds a broken machine are strictly in $(0,1)$. Analogous to open systems, the capacity of the system is roughly equal to the offered load. There exists a fundamental difference in the behavior of open and closed (QED) systems. In an open system, the arrival rate is independent of the number of customers in the system. On the other hand, due to a feedback present in a closed system, the arrival rate decreases as the number of customers in the system increases. In general, our model might be relevant for large-scale systems with the following features: (i) the system operator has a fixed constituency from which service requests materialize, and (ii) once processed, customers return to the pool of potential users and seek additional service later. Studies of the QED regime are partially motivated by the fact that QED-based approximations adequately describe (finite-size) systems that operate in efficiency- or quality-driven regimes (de Véricourt and Jennings 2008). We further motivate the QED regime for machine repair models in Appendix A. Under specific assumptions, we argue that QED-based approximations of steady-state performance measures remain valid for a broader class of machine repair models, where behavior of a single machine depends on the system size.

Contributions. We study a machine repair model with general repair and working times in the QED regime. Our main result (Theorem 1) characterizes the limiting (centered and scaled) number of broken machines process in terms of a nonlinear operator and an explicitly characterized stationary Gaussian process; this Gaussian process corresponds to a limiting (centered and scaled) infinite-repairmen process, an analog of the infinite-server process. (Theorem 1 is analogous to the main result in Reed 2009, where the case of an open QED system is treated.) That is, we establish a relation (in the limit as the size of the system increases) between a machine repairmen model and the corresponding infinite-repairmen model (a process that is characterized explicitly). Due to the closed nature of our system, this relation is more intricate than for the open system. In particular, the corresponding infinite-server model provides a lower bound on the number of customers in the open system with finitely many servers, since arrivals in the two open models occur at the same time. On the other hand, this is not the case for closed systems: breakdown times in the two closed models do not occur at the same time—any wait for repair causes a time shift of future breakdown times for a given machine in the model with finitely many repairmen (see Figure 1 in Section 4.1 for an illustration). Hence, our nonlinear operator accounts not only for waiting times (like in the open system), but for time shifts in breakdown times as well. This operator is defined in terms of the repair and working time distributions. Hence, we demonstrate the impact of those distributions on model performance. Specifically, the nonlinear operator is defined by a conditional probability that an on-off process is in the on state; on and off periods in this process are distributed according to the repair and working time distributions, respectively.

From the technical point of view, we express the difference between the finite and infinite models as a countable sum of processes with some nontrivial dependency structure. These processes are analyzed by means of an extension of a martingale method used to study an open system in Reed (2009) (originally developed in Krichagina and Puhalskii 1997). The extension is due to the dependency structure of processes we consider. In particular, we define these processes recursively in such a way that nonrecursive parts are similar (we work with time-varying "arrival" processes) to a process that arises in the analysis of an open system. In our case, a martingale decomposition is used not only to prove tightness, but also to bound relevant supremum norms.

Related research. As far as the model is concerned, our work most closely relates to de Véricourt and Jennings (2008). There, a memoryless machine repair model in the QED regime was considered. The authors analyzed both transient process-level convergence as well as convergence of steady-state distributions. Corresponding results for an open system in the QED regime (with exponential service times) were earlier obtained in Halfin and Whitt (1981). In Reed (2009), process-level weak convergence for an open system with a general service time distribution in the QED regime was established; steady-state measures were not considered, since no closed form expressions are known to exist. We consider a generalization of the model in de Véricourt and Jennings (2008)—we allow for general repair and working times. Our main result generalizes (de Véricourt and Jennings 2008) in the same way that Reed (2009) generalizes (Halfin and Whitt 1981). Both Halfin and Whitt (1981) and de Véricourt and Jennings (2008) use very similar approaches: (i) steady-state results are based on analyses of the corresponding birth-and-death chains, and (ii) (transient) process-level convergence is established based on Stone's criteria (Stone 1963). On the other hand, our work and Reed (2009) share some similarities as well: (i) they relate the corresponding finite- and infinite-server/repairmen models, and (ii) parts of the analyses are based on the technique developed in Krichagina and Puhalskii (1997). (We provide a detailed outline of our approach in Section 4.1.)

Next, we discuss references that relate to the QED regime and closed queueing networks in heavy traffic. There exist many extensions of the basic machine repair problem (standby machines, spare repairmen, etc.)—readers are referred to a survey (Haque and Armstrong 2007). Origins of the QED regime can be traced back to Erlang (1948). An early mathematical analysis related to the regime can be found in Jagerman (1974). The regime was formalized by Halfin and Whitt (1981), where they analyzed a system with exponential service times. Subsequent studies focused on systems with phase-type (Puhalskii and Reiman 2000), deterministic (Jelenković et al. 2004) (see also Sigman and Whitt 2011) and discrete (Mandelbaum and Momčilović 2008) distributions. The case of general service times was solved in Reed (2009) (see also Puhalskii and Reed 2010). An alternative treatment of the general case based on measure-valued processes can be found in Kaspi and Ramanan (2013). A model that covers impatient customers is analyzed in Dai and He (2010), Mandelbaum and Momčilović (2012). Steady-state distributions of QED systems was studied in Gamarnik and Momčilović (2008), Gamarnik and Goldberg (2013b). Asymptotic refinements of the QED regime can be found in Janssen et al. (2011), Zhang et al. (2011). A speed of convergence to a stationary distribution is a subject of van Leeuwaarden and Knessl (2012), Gamarnik and Goldberg (2013a). Some of results on the control of various queueing models under the QED regime include Atar et al. (2004), Atar (2005a, b), Harrison and Zeevi (2004), Tezcan (2008), Gurvich and Whitt (2009), Armony and Ward (2010), Atar et al. (2011), Mandelbaum et al. (2012), Ata and Gurvich (2012).

The above mentioned papers examined open systems. Certain closed systems can be analyzed by considering open systems with state-dependent arrival rates (Armony and Maglaras 2004). Models with reentrant customers, such as the Erlang-R model (Yom-Tov and Mandelbaum 2014), share features with both open and closed systems. As it was demonstrated in Yom-Tov and Mandelbaum (2014), methods that do not take into account the repetitive nature of service (such as the Piecewise Stationary Approximation (Jennings et al. 1996)) may lead to poor results in time-varying systems. A machine repair problem with spares (an additional feature) in the Efficiency-Driven (ED) regime was studied in Iglehart (1965). Closed many-server systems with state-dependent drifts are explored in Mandelbaum and Pats (1995). In Randhawa and Kumar (2009), the authors consider a closed Markovian many-server loss system with retrials. The case of general service times is covered in Bassamboo et al. (2009), where a transient regime in a closed system is considered; in this model, each customer receives "service" only once and the analysis reduces to a study of an infinite-server process. Examples of heavy-traffic limit theorems for closed queueing systems with a fixed number of servers can be found in Kogan et al. (1986), Harrison and Williams (1996), Kumar (2000).

Notation. Denote by $D[0,\infty)$ the space of all real-valued functions on $[0,\infty)$ that are right-continuous with left limits (r.c.l.l.), endowed with the standard J_1 topology. The J_1 metric is denoted by $d_{J_1}(\cdot,\cdot)$, $|\cdot|$ stands for the Euclidian metric on $\mathbb{R}$, and the uniform metric u is defined by the supremum norm (for $x\in D[0,\infty)$ and $t\geq 0$)

$$\|x\|_t = \sup_{0\leq s\leq t} |x(s)|.$$

Borel σ-field generated by the Skorokhod J_1 topology is denoted by $\mathscr{D}$ (e.g., see Billingsley 1999, Chapter 16); $\mathscr{B}(\mathbb{R})$ stand for the Borel σ-field on $\mathbb{R}$. Product metric spaces $(D^l[0,\infty), d^l)$ are defined by $(D[0,\infty)\times\cdots\times D[0,\infty), d\times\cdots\times d)$, where $d\times\cdots\times d$ referes to the corresponding maximum metric. For a sequence of variables $\{x_i\}$ we use $x_{i:j}$ to represent $\sum_{m=i}^{j} x_m$. The symbols $*$, $\wedge$ and $\vee$ denote the convolution, minimum and maximum operators, respectively. Given a distribution function H, $H^{(n)}$ denotes the n-fold convolution of H with itself, i.e., $H^{(2)} = H*H$; $H^{(0)}\equiv 1$. For $x\in D[0,\infty)$ or $x\in\mathbb{R}$, define $\bar{x}=1-x$, $x^+ = x\vee 0$ and $x^- = -(x\wedge 0)$ (that is, $x = x^+ - x^-$). Let $\Rightarrow$ denote weak convergence—for stochastic processes in $D^l[0,\infty)$ and for random variables in $\mathbb{R}^l$. Furthermore, let $1_{\{\cdot\}}$ and Φ be the indicator function and the standard normal distribution, respectively. The composition map is denoted by $\circ$; namely, for $(x,y)\in D[0,\infty)\times D[0,\infty)$, $x\circ y$ is defined by $(x\circ y)(t) = x(y(t))$, $t\geq 0$. We use $\stackrel{\mathrm{d}}{=}$ to indicate equality in distribution. The symbol $\stackrel{\text{f.d.}}{\Rightarrow}$ is used to denote finite-dimensional convergence; $\stackrel{\mathbb{P}}{\rightarrow}$ stands for convergence in probability. For a nondecreasing function f on $\mathbb{R}$, we define the (left continuous) inverse of f as $f^{\leftarrow}(x) = \inf\{s\colon f(s)\geq x\}$; the infimum of an empty set is set to $+\infty$.

Organization. The paper is organized as follows. In the next section, we describe the model. Our main results can be found in Section 3. The last two sections contain technical proofs. The appendices contain additional motivation for the QED regime, ancillary results, as well as a dependency graph that illustrates how various results in the paper relate.

2. Model

In this section, we describe the model and our assumptions. We consider a sequence of machine repair models indexed by the number of machines n. The number of repairmen in the nth model is given by k^n. At any instant of time, a machine is either working, or it is broken. A machine works for a random amount of time distributed according to a distribution F (with mean $1/\lambda<\infty$, $F(0)=0$). Once broken, a machine requires a repair that lasts a random amount of time distributed according to a distribution G (with mean $1/\mu<\infty$, $G(0)=0$). A machine enters a repair phase as soon as a repairman is available; repairmen service machines according to the FCFS rule. For convenience, define $p := \lambda/(\lambda+\mu)$ as the long-run fraction of time a machine would spend being broken (repaired) in the case it had a dedicated repairman. Working and repair times are independent both across time and machines/repairmen.

Let $X^n(t)$ be the number of broken machines (being repaired or awaiting service) in the nth system at time t; $X^n := \{X^n(t),\ t\geq 0\}$. Similarly, we let $X_i^n(t)$ be an indicator of the ith machine not working in the nth system at time t; then, we have

$$X^n = \sum_{i=1}^{n} X_i^n, \tag{1}$$

where $X_i^n := \{X_i^n(t),\ t\geq 0\}$. Initially, at time $t=0$, ξ^n machines are broken in the nth system, i.e., $X^n(0)=\xi^n$. In particular, $\xi^n\wedge k^n$ machines are being repaired (those with indices $i=1,\ldots,\xi^n\wedge k^n$), $(\xi^n-k^n)^+$ machines are awaiting service (those with indices $i=\xi^n\wedge k^n+1,\ldots,\xi^n$), and $n-\xi^n$ machines are working (those with indices $i=\xi^n+1,\ldots,n$). We assume that the remaining working (repair) times of machines working (being repaired) at time $t=0$ are independent and distributed according to the residual distribution of F (G). Recall that the residual distributions $\ddot{F}$ and $\ddot{G}$ are defined by

$$\ddot{F}(x) := \lambda\int_0^x \bar{F}(s)\,\mathrm{d}s \qquad\text{and}\qquad \ddot{G}(x) := \mu\int_0^x \bar{G}(s)\,\mathrm{d}s. \tag{2}$$

Now, consider an arbitrary machine i. For this machine, we define the jth cycle ($j \geq 2$) to be the time period between the jth and $(j+1)$st breakdowns. During the jth cycle, the ith machine first awaits for service for $w_{i,j} \geq 0$ time units, then receives service for $s_{i,j}$ time units, and finally remains operational for $a_{i,j}$ time units.

For notational convenience, we adopt the convention that the first cycle for machines working at time $t = 0$ ($\xi^n < i \leq n$) consists of their remaining working times ($a_{i,1}$ is distributed according to $\ddot{F}$, $w_{i,1} \equiv 0$, $s_{i,1} \equiv 0$); for machines being repaired at time $t = 0$ ($1 \leq i \leq \xi^n \wedge k^n$), the first cycle consists of the rest of the repair time and the following working period ($s_{i,1}$ is distributed according to $\ddot{G}$, $a_{i,1}$ according to F, $w_{i,1} \equiv 0$); the first cycle for machines awaiting repair at time $t = 0$ ($\xi^n \wedge k^n < i \leq \xi^n$) is the time period between $t = 0$ and their first breakdown that occurs after $t = 0$ ($s_{1,i}$ and $a_{i,1}$ are distributed according to G and F, respectively). Finally, let $c_{i,j} = s_{i,j} + a_{i,j}$, and note that $\{c_{i,j}\}_{j \geq 2}$ have the distribution $H = G * F$. Then, X_i^n can be written as follows:

$$X_i^n(t) = \sum_{j=0}^{\infty} 1_{\{0 \leq t - c_{i,1:j} - w_{i,1:j} < w_{i,j+1} + s_{i,j+1}\}}; \tag{3}$$

recall that $c_{i,1:j} := \sum_{m=1}^{j} c_{i,m}$ and $w_{i,1:j} := \sum_{m=1}^{j} w_{i,m}$.

We consider a sequence of machine repair models in the QED regime. In such a regime, the capacity and the offered load are related via the square-root rule:

$$\beta^n := \frac{k^n - np}{\sqrt{np\bar{p}}} \to \beta, \tag{4}$$

as $n \to \infty$, for some $\beta \in (-\infty, \infty)$; recall that $\bar{p} = 1 - p$. Analogous to open systems in the QED regime, the capacity of the system is approximately equal to the offered load. The scaled total number of broken machines is given by

$$\hat{X}^n := \frac{X^n - k^n}{\sqrt{np\bar{p}}}. \tag{5}$$

Remark 1. Typically, the scaling parameter used for QED regime is $\sqrt{k^n}$ rather than $\sqrt{np\bar{p}}$; $\sqrt{n}$ was used in de Véricourt and Jennings (2008). However, k^n, n and $np\bar{p}$ are of the same order of magnitude, and $np\bar{p}$ is more convenient in terms of stating our results. In particular, under the $\sqrt{np\bar{p}}$ scaling, the stationary component of the corresponding infinite-repairmen process has unit variance (see Lemma 1 in Section 3). In addition, such a scaling allows one to consider more general settings, where p or $\bar{p}$ vanish while n increases. For example, in the model described in this section, the distributions F and G do not vary with the size of the system (n); this is a standard assumption, e.g., see de Véricourt and Jennings (2008). Consequently, the parameter p is fixed. However, in some applications, the value of p can be close to 0, while the product np is large (for example, $p = 0.001$ and $n = 50{,}000$). For such cases, it is of interest to consider a model where p varies with n. We discuss one aspect of such a model in Appendix A.

We assume that the system is in the QED regime at time $t = 0$ (since we focus on diffusion rather than fluid limits). In particular, the number of broken machines at time $t = 0$ satisfies, as $n \to \infty$,

$$\hat{\xi}^n := \frac{\xi^n - k^n}{\sqrt{np\bar{p}}} \Rightarrow \hat{\xi}. \tag{6}$$

Note that this limit and the assumption on the distribution of residual repair and working times at $t = 0$ (see (2)) specify the state of the system at time $t = 0$. These two relatively simple assumptions ensure that the system remains in the QED regime on finite time intervals (the weak limit of $\hat{X}^n$ is nondegenerate). In the analysis (Reed 2009) of the QED GI/GI/k queue, the open analogue of our model, the same assumption ($\ddot{G}$) on residual service (repair) times was used. An alternative way to specify the system at $t = 0$ would be to initialize it in its steady state. However, the limiting (QED) steady-state distribution for the machine repair model is not known; moreover, even the steady-state distribution for the QED GI/GI/k queue is not known (only partial results exists, e.g., see Gamarnik and Goldberg 2013b).

3. Results

An infinite-repairmen process is central in our analysis. We define $Z_i^n := \{Z_i^n(t), t \geq 0\}$ to be an indicator process of whether machine i is broken in the case when a dedicated repairman is always available to it. In terms of earlier introduced notation, one has, for $t \geq 0$,

$$Z_i^n(t) = \sum_{j=0}^{\infty} 1_{\{0 \leq t - c_{i,1:j} < s_{i,j+1}\}}. \tag{7}$$

We note that both X_i^n and Z_i^n describe the behavior of the same machine i (a machine is described by two sequences of working and repair times). Indeed, the right-hand side of (7) is obtained from the right-hand side of (3) by eliminating waiting times $\{w_{i,j}\}$. In particular, a machine i, $\xi^n \wedge k^n < i \le \xi^n$, is waiting repair at time $t = 0$ in the original model (X_i^n), while it repair starts at $t = 0$ in the infinite-repairmen model (Z_i^n).

Several functions can be associated with the above definition (7). In particular, let $P_1(t)$, $t \ge 0$, be the conditional probability of a machine being broken at time t, given that a repair process of the machine started at time $t = 0$. Then, $P_1 := \{P_1(t),\ t \ge 0\}$ is given by

$$P_1 = R - G * R = \bar{G} * R, \tag{8}$$

where R is the renewal function associated with the distribution H, i.e., $R = 1 + H + H * H + \cdots$ (e.g., see Çinlar 1975, p. 286). Similarly, let $\ddot{P}_1(t)$ and $\ddot{P}_0(t)$ be the conditional probabilities of a machine being broken at time t, given that it is being repaired at $t = 0$ (with the distribution of the remaining repair time equal to $\ddot{G}$) and it is working at $t = 0$ (with the distribution of the remaining working time equal to $\ddot{F}$), respectively. Consequently, if $\ddot{P}_1 := \{\ddot{P}_1(t),\ t \ge 0\}$ and $\ddot{P}_0 := \{\ddot{P}_0(t),\ t \ge 0\}$, one has

$$\ddot{P}_1 = 1 + \ddot{G} * F * R - \ddot{G} * R = 1 - \ddot{G} * \bar{F} * R, \tag{9}$$

$$\ddot{P}_0 = \ddot{F} * R - \ddot{F} * G * R = \ddot{F} * \bar{G} * R. \tag{10}$$

We note that, due to the construction of stationary on-off processes in Heath (1998), $p\ddot{P}_1 + \bar{p}\ddot{P}_0 = p$ holds. Moreover, if the distribution H is nonlattice, than $P_1(t) \to p$, as $t \to \infty$ (Asmussen 2003, p. 173); the same applies to $\ddot{P}_1$ and $\ddot{P}_0$ (Resnick 1992, p. 237). Finally, based on the preceding and the indexing of the machines (see Section 2), it follows that

$$\mathbb{E}Z_i^n(t) = \begin{cases} \ddot{P}_1(t), & 1 \le i \le \xi^n \wedge k^n, \\ P_1(t), & \xi^n \wedge k^n < i \le \xi^n, \\ \ddot{P}_0(t), & \xi^n < i \le n. \end{cases} \tag{11}$$

Note that $\mathbb{E}Z_i^n(t)$ depends on the machine index i, since i determines the state of a machine at time $t = 0$ in the original finite-repairmen system: machines $1 \le i \le \xi^n \wedge k^n$ are being repaired, machines $\xi^n \wedge k^n < i \le \xi^n$ are awaiting repair, and machines $\xi^n < i \le n$ are working.

The next lemma describes the limiting behavior of a centered and scaled version of the infinite-repairmen process

$$Z^n := \{Z^n(t),\ t \ge 0\} := \sum_{i=1}^{n} Z_i^n. \tag{12}$$

Lemma 1 (Infinite-Repairmen Process). *Suppose*

$$\limsup_{t \downarrow 0} \frac{F(t) \vee G(t)}{t} < \infty. \tag{13}$$

Then, as $n \to \infty$,

$$\hat{Z}^n \equiv \frac{Z^n - np}{\sqrt{np\bar{p}}} \quad \Rightarrow \quad \hat{Z} \equiv \hat{Y} + (\hat{\xi} + \beta - \hat{Y}(0))(\ddot{P}_1 - \ddot{P}_0) + \hat{\xi}^+(P_1 - \ddot{P}_1), \tag{14}$$

where $\hat{\xi}$ and $\hat{Y} := \{\hat{Y}(t), t \ge 0\}$ are independent, and $\hat{Y}$ is a stationary centered Gaussian process with the covariance function

$$\mathbb{E}\hat{Y}(t)\hat{Y}(t+s) = \frac{\ddot{P}_1(s) - p}{1 - p}, \quad t, s \ge 0. \tag{15}$$

Proof. See Section 5.1. ◻

The statement of the lemma is consistent with $\hat{Z}(0) = \hat{\xi} + \beta$, since $P_1(0) = \ddot{P}_1(0) = 1$ and $\ddot{P}_0(0) = 0$. The term $(\hat{\xi} + \beta - \hat{Y}(0))(\ddot{P}_1 - \ddot{P}_0)$ stems from the difference between $\hat{Z}(0)$ and $\hat{Y}(0)$. The term $\hat{\xi}^+(P_1 - \ddot{P}_1)$ is due to $(\xi^n - k^n)$ machines awaiting repair at $t = 0$—their remaining repair times at $t = 0$ are distributed according to G rather than $\ddot{G}$ (for exponential G, one has $P_1 = \ddot{P}_1$).

Remark 2. The condition (13) is needed for application of Hahn's theorem (Hahn 1978), a central limit theorem for $D[0,1]$ variables (stationary on-off processes in our context). In particular, (13) ensures that a condition of Hahn's theorem are satisfied (Szczotka 1980) (see also Dębicki and Palmowski 1999). In the case of renewal counting processes (rather than alternating renewal processes), it was shown in Whitt (1985) that an analogous condition is necessary for the condition of Hahn's theorem to hold (see also Whitt 2002, p. 229). Such a condition does not appear in Reed (2009), where process-level weak convergence was considered for an open QED analogue of our model (G/GI/N model)—there, it was assumed that a sequence of (exogenous) arrival processes satisfies a FCLT. Interestingly, condition (13) for the service time distribution appears in the study (Gamarnik and Goldberg 2013b) of the limiting *steady-state* distribution of the GI/GI/N system.

The second component required to state our main result is the nonlinear convolution operator defined below. The operator φ_G plays a crucial rule in describing the behavior of open QED systems with the service distribution G.

Definition 1 (Reed 2009)**.** Let P be either G or $\bar{P}_1 = 1 - P_1$. The mapping $\varphi_P\colon D[0,\infty) \to D[0,\infty)$ is such that $\varphi_P(x)$, for each $x \in D[0,\infty)$, is the unique solution y to

$$y(t) = x(t) + \int_0^t y^+(t-s)\,\mathrm{d}P(s). \tag{16}$$

Proposition 1. *For each $x \in D[0,\infty)$, there exists a unique $\varphi_{\bar{P}_1}(x)$. The function $\varphi_{\bar{P}_1}\colon D[0,\infty) \to D[0,\infty)$ is Lipschitz continuous in the topology of uniform convergence over bounded intervals, and it is measurable with respect to the Borel σ-field generated by the Skorokhod J_1 topology.*

Proof. See Section 5.2. □

The following theorem is the main result of the paper. It describes the limiting number-of-broken-machines process in terms of the corresponding limiting infinite-repairman process $\hat{Z}$ (see Lemma 1). We note that, in general, $\bar{P}_1$ is not a monotonic function. This reflects the fact that $\hat{Z}$ does not serve as a lower bound for $\hat{X}$.

Theorem 1. *Consider a sequence of machine repair models in the QED regime. If the sequence of scaled and centered infinite-repairmen processes converges weakly ($\hat{Z}^n \Rightarrow \hat{Z}$, as $n \to \infty$), then, as $n \to \infty$,*

$$\hat{X}^n \Rightarrow \hat{X} \equiv \varphi_{\bar{P}_1}(\hat{Z} - \beta). \tag{17}$$

Proof. See Section 4. □

Remark 3 (Open System)**.** Let $\grave{X}^k := \{\grave{X}^k(t),\, t \ge 0\}$ be the number-in-system process in an open system with k servers, Poisson arrivals and the same service distribution G. Here, we index systems with the number of servers, since the number of customers (machines) is not fixed. The arrival rate $\mu\theta^k$ in the k-server system is such that the sequence of systems is in the QED regime: $(k - \theta^k)/\sqrt{k} \to \beta$, as $k \to \infty$. The initial number of customers in the system $\grave{\xi}^k$ satisfies $(\grave{\xi}^k - k)/\sqrt{k} \Rightarrow \grave{\xi}$, as $k \to \infty$. The main result in Reed (2009) implies, as $k \to \infty$,

$$\frac{\grave{X}^k - k}{\sqrt{k}} \Rightarrow \grave{X} \equiv \varphi_G(\grave{Z} - \beta), \tag{18}$$

where

$$\grave{Z} = \grave{Y} + (\grave{\xi} + \beta - \grave{Y}(0))\bar{\bar{G}} + \grave{\xi}^+(\bar{G} - \bar{\bar{G}}) \tag{19}$$

and $\grave{Y} := \{\grave{Y}(t),\, t \ge 0\}$ is a stationary centered Gaussian process (independent of $\grave{\xi}$) with $\mathbb{E}\grave{Y}(t)\grave{Y}(t+s) = 1 - \ddot{G}(s)$, $t, s \ge 0$. (Compare (19) to (14). Process $\grave{Y}$ is a stationary limiting (centered and scaled) $M/G/\infty$ process defined by the service distribution G.) The result in Reed (2009) is stated in terms of a time-changed Brownian bridge and a nonstationary limiting (centered and scaled) infinite-server process; those two terms correspond to customers initially (at time $t = 0$) present in the system and those arriving after $t = 0$, respectively. However, considering all customers together results in (18) and (19).

We note that comparison $\hat{X}$ and $\grave{X}$ via (17) and (18) is not straightforward, in general. The reason is that the difference in $\hat{X}$ and $\grave{X}$ is not only due to the different nonlinear operators φ_G and $\varphi_{\bar{P}_1}$, but different correlation structures of $\hat{Z}$ and $\grave{Z}$ (or rather different covariance functions of $\hat{Y}$ and $\grave{Y}$) as well.

Next, we discuss an example where explicit results can be obtained.

Example 1 (Memoryless System)**.** Assume $\bar{G}(t) = e^{-\mu t}$ and $\bar{F}(t) = e^{-\lambda t}$ for $t \ge 0$. In this case, $P_1(t) = p + \bar{p}e^{-(\lambda+\mu)t}$, for $t \ge 0$ (e.g., see Ross 1997, p. 243), and it follows that $\hat{X}$ satisfies, for $t \ge 0$,

$$\hat{X}(t) = \hat{Z}(t) - \beta + \mu \int_0^t \hat{X}^+(t-s)e^{-(\lambda+\mu)s}\,\mathrm{d}s, \tag{20}$$

where $\hat{Z}(t) = \hat{Y}(t) + (\hat{\xi} + \beta - \hat{Y}(0))e^{-(\lambda+\mu)t}$ and $\hat{Y}$ is a stationary centered Gaussian process (independent of $\hat{\xi}$) with $\mathbb{E}\hat{Y}(t)\hat{Y}(t+s) = e^{-(\lambda+\mu)s}$, $t, s \geq 0$. Alternatively, $\hat{Z}$ is characterized by (e.g., see Browne and Whitt 1995)

$$\mathrm{d}\hat{Z}(t) = -(\lambda + \mu)\hat{Z}(t)\,\mathrm{d}t + \sqrt{2(\lambda + \mu)}\,\mathrm{d}B(t); \tag{21}$$

here, $B := \{B(t),\ t \geq 0\}$ is a standard Brownian motion. Relation (20) implies

$$\mathrm{d}\hat{X}(t) = \mathrm{d}\hat{Z}(t) + \mu\hat{X}^{+}(t)\,\mathrm{d}t - (\lambda + \mu)(\hat{X}(t) - \hat{Z}(t) + \beta)\,\mathrm{d}t,$$

which combined with (21) yields

$$\mathrm{d}\hat{X}(t) = (\mu\hat{X}^{+}(t) - (\lambda + \mu)(\hat{X}(t) + \beta))\,\mathrm{d}t + \sqrt{2(\lambda + \mu)}\,\mathrm{d}B(t). \tag{22}$$

That is, $\hat{X}$ is a diffusion, as earlier obtained in de Véricourt and Jennings (2008, Theorem 2).

Based on Remark 3, the limiting number of customers $\grave{X} := \{\grave{X}(t),\ t \geq 0\}$ in the corresponding (the same initial conditions: $\grave{\xi} = \hat{\xi}$) open QED system satisfies $\grave{X} = \varphi_G(\grave{Z} - \beta)$, where $\grave{Z}(t) = \grave{Y}(t) + (\grave{\xi} + \beta - \grave{Y}(0))e^{-\mu t}$ and $\grave{Y} := \{\grave{Y}(t),\ t \geq 0\}$ is a stationary centered Gaussian process (independent of $\hat{\xi}$) with $\mathbb{E}\grave{Y}(t)\grave{Y}(t+s) = e^{-\mu s}$, $t, s \geq 0$. Defining a version of $\grave{X}$ with a time speedup $\tilde{X} := \{\tilde{X}(t) = \grave{X}(t/\bar{p}),\ t \geq 0\}$ yields that $\tilde{X}$ is equal in distribution to $\varphi_{G/\bar{p}}(\hat{Z} - \beta)$, since $\mathbb{E}\grave{Y}(t/\bar{p})\grave{Y}(t/\bar{p} + s/\bar{p}) = \mathbb{E}\hat{Y}(t)\hat{Y}(t+s)$. Finally, we observe $\varphi_{G/\bar{p}}(\hat{Z} - \beta)(t) \geq \varphi_{\bar{P}_1}(\hat{Z} - \beta)(t)$, which implies that $\tilde{X}(t)$ is stochastically larger than $\hat{X}(t)$, for any $t > 0$. We remark that this is not the only way to compare the closed and open systems. Comparison of the $M/M/k^n/\infty/n$ (machine repair model with memoryless repair and working times) and $M/M/k$ systems (with the same service/repair distribution, and the same staffing parameter β) in the QED regime was considered in de Véricourt and Jennings (2008). In particular, the authors established that the stationary probability of delay is smaller in the closed system than in the open system.

4. Proof of Theorem 1

In the first subsection we provide an outline of the proof. We state and prove some preliminary results in the second subsection. The proof of Theorem 1 can be found in the last subsection.

4.1. Outline of the Proof

We describe the proof in two parts. At the end of the subsection, we provide some intuition behind our result.

Part I: *Decomposition.* Theorem 1 relates the number-of-broken-machines process (X^n) and the corresponding infinite-repairmen process (Z^n), in the limit as $n \to \infty$. In order to quantify $X^n - Z^n$, we consider such differences for individual machines, i.e., $X_i^n - Z_i^n$. Recall that Z_i^n is an on-off process with independent on and off periods. While X_i^n is an on-off process as well, its structure is more intricate. In particular, any wait for repair after a breakdown causes a time shift in all future times when the machine stops and starts working. We illustrate this phenomenon in Figure 1, where we plot X_i^n, Z_i^n and related processes for a machine i, $\xi^n \wedge k^n < i \leq \xi^n$ (a machine that is awaiting repair at $t = 0$; for other initial conditions the figure is similar). As stated earlier, Z_i^n does not provide a lower bound for X_i^n; $(X_i^n - Z_i^n)(t)$ takes values in $\{-1, 0, 1\}$. The difference in X_i^n and Z_i^n is due to the sequence of machine i waiting times $\{w_{i,j}\}_{j\geq 1}$. Since the structure of $X_i^n - Z_i^n$ is involved, we decompose this process into processes based on the cause of discrepancy. To this end, we define $\Upsilon_{i,j}^n = \{\Upsilon_{i,j}^n(t),\ t \geq 0\}$, $j \geq 0$, to account for differences in X_i^n and Z_i^n due to the waiting time $w_{i,j+1}$ (see Figure 1), such that

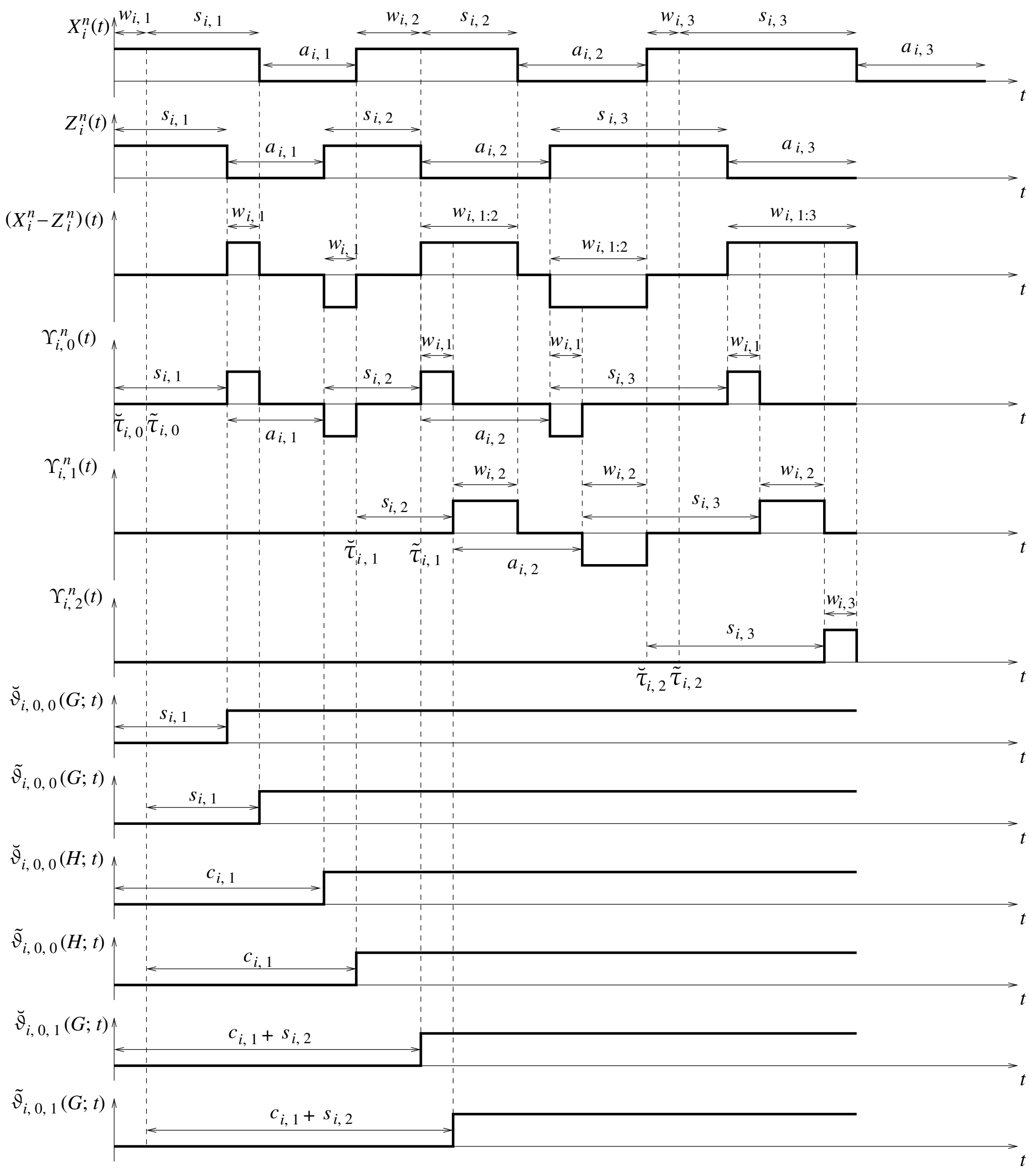
$$X_i^n - Z_i^n = \sum_{j=0}^{\infty} \Upsilon_{i,j}^n.$$

The three main features of $\Upsilon_{i,j}^n$ are (i) length of all $+1$ and -1 periods are equal to $w_{i,j+1}$; (ii) beginning of $+1$ and -1 periods are determined by $\{s_{i,j+l}\}_{l\geq 1}$, $\{a_{i,j+l}\}_{l\geq 1}$ and the "start" time $\breve{\tau}_{i,j} := c_{i,1:j} + w_{i,1:j}$; and $\Upsilon_{i,j}^n$ does *not* depend on $\{w_{i,l}\}_{l>j+1}$. Next, we further decompose $\Upsilon_{i,j}^n$ itself in terms of as difference of "one-point" point processes:

$$\Upsilon_{i,j}^n = \sum_{L\in\{H,G\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}} (\breve{\vartheta}_{i,j,l}^n(L) - \tilde{\vartheta}_{i,j,l}^n(L));$$

here, $\breve{\vartheta}_{i,j,l}^n(L) := \{\breve{\vartheta}_{i,j,l}^n(L;t),\ t \geq 0\}$ and $\tilde{\vartheta}_{i,j,l}^n(L) := \{\tilde{\vartheta}_{i,j,l}^n(L;t),\ t \geq 0\}$. In Figure 1 we illustrate the decomposition of $\Upsilon_{i,0}^n$. We note that one-point point processes are well-studied objects (for example, compensators for such processes are well known—see Jacod and Shiryaev 2003, p. 98).

Figure 1. An illustration of the decomposition for $X_i^n - Z_i^n$ ($X_i^n(t)$ and $Z_i^n(t)$ are indicators of machine i being broken at time t in our machine repair model and the corresponding infinite-repairmen model, respectively). This difference is represented as a sum of processes $\{\Upsilon_{i,j}^n\}_{j\ge 0}$. Each of those processes, in turn, is decomposed into one-point point processes $\{\breve{\vartheta}_{i,j,l}(L)\}_{l\ge 0}$ and $\{\tilde{\vartheta}_{i,j,l}(L)\}_{l\ge 0}$, $L \in \{G, H\}$. Only the decomposition for $\Upsilon_{i,0}$ is shown.



The next step is to center and scale appropriately relevant processes. In particular, we introduce (see (26))

$$\breve{M}_{l,j}^n(L) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big(\breve{\vartheta}_{i,j,l}^n(L) - \mathbb{E}[\breve{\vartheta}_{i,j,l}^n(L) \mid \breve{\tau}_{i,j})]\big), \tag{23}$$

where $L \in \{G, H\}$; $\tilde{M}_{l,j}^n(L)$ is introduced analogously—replace $\breve{\ }$ with $\tilde{\ }$. Note that a conditional expectation is used for centering in the preceding equation, since the same $\breve{\tau}_{i,j}$ is used in the definition of $\{\breve{M}_{l,j}(L)\}_{l\ge 0, L\in\{G,H\}}$ (equivalently, this "preserves" the structure of $\Upsilon_{i,j}^n$; recall that $\breve{\tau}_{i,j}$ is the "start" time of $\Upsilon_{i,j}^n$). Capturing this dependency is crucial. Then, one has

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\Upsilon_{i,j}^n - \mathbb{E}[\Upsilon_{i,j}^n \mid \breve{\tau}_{i,j}, \tilde{\tau}_{i,j}]) = \sum_{L\in\{G,H\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}} (\breve{M}_{l,j}^n(L) - \tilde{M}_{l,j}^n(L));$$

observe that the conditioning on $(\check{\tau}_{i,j}, \tilde{\tau}_{i,j})$ in the preceding equality is equivalent to conditioning on the "start" time and the duration of +1 and −1 periods of $\Upsilon^n_{i,j}$, since that duration is $w_{i,j+1} = \tilde{\tau}_{i,j} - \check{\tau}_{i,j}$ (see Figure 1).

Finally, by using the above, $\hat{X}^n$ (see (5)) can be written in terms of $\hat{Z}^n$ (see (14)) by means of $\varphi_{\bar{P}_1}$ (see Definition 1 and (8)):

$$\hat{X}^n = \varphi_{\bar{P}_1}(\hat{Z}^n - \beta^n + \hat{\Delta}^n),$$

where β^n is defined in (4), and (see (28))

$$\hat{\Delta}^n = \frac{1}{\sqrt{p\bar{p}}} \sum_{L\in\{H,G\}} \sum_{l,j=0}^{\infty} (-1)^{1_{\{L=H\}}} (\check{M}^n_{l,j}(L) - \tilde{M}^n_{l,j}(L)). \tag{24}$$

We note that $\{\check{M}^n_{l,j}(L)\}_{l,j,L}$ and $\{\tilde{M}^n_{l,j}(L)\}_{l,j,L}$ are two sequences with dependent elements; moreover, the two sequences are not mutually independent.

Part II: *Convergence.* In view of continuity of $\varphi_{\bar{P}_1}$ (Proposition 1), $\hat{Z}^n \Rightarrow \hat{Z}$ (Lemma 1) and $\beta^n \to \beta$ (see (4)), the rest of the proof addresses $\hat{\Delta}^n \Rightarrow 0$ (Lemma 5). The main idea is to show that $\check{M}^n_{l,j}(L) - \tilde{M}^n_{l,j}(L)$ vanishes for fixed l and $j \ge 1$ (Lemma 2), as $n \to \infty$, by arguing

$$(\check{M}^n_{l,j}(L), \tilde{M}^n_{l,j}(L)) \;\;\Rightarrow\;\; (M_{l,j}(L), M_{l,j}(L)),$$

where $M_{l,j}(L)$ is an explicitly characterized Gaussian process with a.s. continuous sample paths. The case $j = 0$ is handled in Lemma 4—it is slightly different due to initial conditions of machines at time $t = 0$. The remaining case (large $l + j$) is addressed in Lemma 3. There, instead of considering joint convergence of $\check{M}^n_{l,j}(L)$ and $\tilde{M}^n_{l,j}(L)$, we bound each term separately (based on a martingale decomposition described below).

Next, we discuss the proof of Lemma 2 in detail. The starting point is the fact that finite-dimensional convergence and tightness imply weak convergence, e.g., see Billingsley (1999, p. 139). The finite-dimensional convergence is obtained by utilizing a technique from Krichagina and Puhalskii (1997): stochastic integrals are approximated by finite sums (fluid limits from Lemma 9 are utilized there as well; in turn, those limits are based on fluid limits in Corollary 1 and Lemma 8). Aldous' sufficient criterion (Liptser and Shiryaev 1989, p. 515) is the main tool for establishing tightness (Lemma 12 and Remark 7 in Appendix B.2). One feature of our proof is that tightness for $\check{M}^n_{l,j}$ and $\tilde{M}^n_{l,j}$ is established recursively, which is due to a martingale decomposition (see Lemma 10 in Appendix B.1; the martingale term is characterized in Lemma 11). The basic idea behind this part of our analysis is to define analogues (the main technical difficulty) of arrival processes in the open system so that machinery developed in Krichagina and Puhalskii (1997) is applicable. For this reason, the first two terms in the martingale decomposition of $\check{M}^n_{l,j}$ are of the same form as the terms that appear in the analysis of the open system (in our case "arrivals" are time-varying). However, the decomposition has a recursive term that involves $\check{M}^n_{l-1_{\{L=G\}},j}(\underline{L})$, where $\{\underline{L}\} = \{G, H\}\backslash\{L\}$. This is due to our system being closed—such a structure does not appear in an open system. (In fact, the analogue of $\hat{\Delta}^n$ for an open system contains a single difference rather than infinitely many as in (24).) While in an open systems "arrivals" do correspond to physical arrivals of customers to the service facility, analogues of "arrivals" for our model do not correspond to arrivals in the physical system (either breakdowns or times when machines start working). Rather, "arrivals" in our analysis stem from a decomposition of points that define one-point point processes $\check{\vartheta}_{i,j,l}(L)$ and $\tilde{\vartheta}_{i,j,l}(L)$. Such a representation is needed to obtain orthogonality of martingales that correspond to individual machines (see Remark 6 in Appendix B.1).

Now, we describe the intuition behind our result. Here, we draw a parallel with an open system. There, $\grave{X}^k(t) - \grave{Z}^k(t)$ represents the number of customers that are still in the finite-server system, but not in the infinite-server one, at time t. ($\grave{Z}^k = \{\grave{Z}^k(t), t \ge 0\}$ is the corresponding (unscaled) infinite-server process.) All those customers were delayed upon their arrival, i.e., they were awaiting service at some point in the interval $[0, t]$. Consider customers awaiting service at time $s \in [0, t)$—those customer arrived to the systems at time $\approx s$, since waiting times vanish in the QED regime. Hence, out of those $(\grave{X}(s) - k)^+$ customers, only customers with service times $\approx (t - s)$ are in the group of $\grave{X}^k(t) - \grave{Z}^k(t)$ customers that are still in the finite-server system, but not in the infinite-server one, at time t. This leads to

$$\grave{X}^k(t) - \grave{Z}^k(t) \approx \int_0^t (\grave{X}^k(t-s) - k)^+ \, \mathrm{d}G(s);$$

here, we took advantage of the fact that $(\grave{X}(s) - k)^+$ is large (proportional to $\sqrt{k}$), implying that the fraction of jobs with service times $(t - s)$ is $\approx \mathrm{d}G(t - s)$. The intuition for the closed system is similar, yet one needs to account for more scenarios. A machine awaiting repair at time $s \in [0, t)$ can cause a discrepancy in $X^n(t)$ and $Z^n(t)$ in multiple ways. Namely, such a machine contributes to the positive part of $X^n(t) - Z^n(t)$, if its next repair time, say $s_{i,j}$, is $\approx (t-s)$ (like in the open system), but also if $c_{i,j:m} + s_{i,m+1} \approx (t-s)$, for some $m \geq j$. This reflects the fact that a machine can complete multiple cycles in the time interval $[s,t]$. Similarly, this same machine contributes to the negative part of $X^n(t) - Z^n(t)$, if $c_{i,j:m} \approx (t-s)$, for some $m \geq j$ (see Figure 1). Consequently, one has (see (8))

$$\begin{aligned} X^n(t) - Z^n(t) &\approx \sum_{m=0}^{\infty} \int_0^t (X^n(t-s) - k^n)^+ \, \mathrm{d}G * H^{(m)}(s) - \sum_{m=1}^{\infty} \int_0^t (X^n(t-s) - k^n)^+ \, \mathrm{d}H^{(m)}(s) \\ &= \int_0^t (X^n(t-s) - k^n)^+ \, \mathrm{d}\bar{P}_1(s), \end{aligned}$$

which after centering and scaling yields $\hat{X}^n \approx \varphi_{\bar{P}_1}(\hat{Z}^n - \beta^n)$.

4.2. Preliminary Results

For notational simplicity, let

$$\breve{\tau}_{i,j} := c_{i,1:j} + w_{i,1:j} \qquad \text{and} \qquad \tilde{\tau}_{i,j} := c_{i,1:j} + w_{i,1:j+1}; \tag{25}$$

these two quantities can be interpreted as the time of the jth breakdown and the time when the jth repair starts for the ith machine, respectively; for $j = 0$, we have $\breve{\tau}_{i,0} = 0$ and $\tilde{\tau}_{i,0} = w_{i,1}$ (recall that $w_{i,1} > 0$ only for $(\xi^n - k^n)^+$ machines). Also, for $L \in \{G, H\}$ and $l, j \geq 0$, define processes $\breve{M}^n_{l,j}(L) := \{\breve{M}^n_{l,j}(L;t),\ t \geq 0\}$ by (see (23))

$$\breve{M}^n_{l,j}(L;t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1_{\{c_{i,j+1:j+l} + s_{i,j+l+1} + a_{i,j+l+1} 1_{\{L=H\}} \leq t - \breve{\tau}_{i,j}\}} - L * H^{(l)}(t - \breve{\tau}_{i,j}) \right). \tag{26}$$

Processes $\tilde{M}^n_{l,j}(L) := \{\tilde{M}^n_{l,j}(L;t),\ t \geq 0\}$ are defined similarly—$\breve{\tau}_{i,j}$'s are replaced with $\tilde{\tau}_{i,j}$'s in (26). An asymptotic description of these processes is provided in the following lemma.

Lemma 2. *For $L \in \{G, H\}$ and $j \geq 1$, we have $(\breve{M}^n_{l,j}(L), \tilde{M}^n_{l,j}(L)) \Rightarrow (M_{l,j}(L), M_{l,j}(L))$, as $n \to \infty$, where $M_{l,j}(L) := \{M_{l,j}(L;t), t \geq 0\}$ is a centered Gaussian process with a.s. continuous paths and (for $t, s \geq 0$)*

$$\mathbb{E} M_{l,j}(L;t) M_{l,j}(L;t+s) = \int_0^t L * H^{(l)}(t-u) \overline{L * H^{(l)}}(t+s-u) \, \mathrm{d}\ddot{H} * H^{(j-1)}(u).$$

Proof. See Section 5.3. □

Observe that the variance of $M_{l,j}(L;t)$ vanishes as indices $l \geq 0$ and/or $j \geq 1$ increase:

$$\mathbb{E}(M_{l,j}(L;t))^2 \leq L * \ddot{H} * H^{(l+j-1)}(t);$$

indeed, there exists an $m \geq 1$ such that $H^{(m)}(t) < 1$, and hence

$$\mathbb{E}(M_{l,j}(L;t))^2 \leq H^{(l+j-1)}(t) \leq (H^{(m)}(t))^{\lfloor (l+j-1)/m \rfloor} \to 0,$$

as $l + j \to \infty$. The fact that the variance of $M_{l,j}(L;t)$ vanishes motivates the following result.

Lemma 3. *For $L \in \{G, H\}$ and $\epsilon > 0$, we have*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left[\sum_{\substack{l+j>m \\ j\neq 0}} (\|\breve{M}^n_{l,j}(L)\|_t + \|\tilde{M}^n_{l,j}(L)\|_t) > \epsilon \right] = 0.$$

Proof. See Section B.1. □

Next, we consider $\breve{M}^n_{l,j}(L)$ and $\tilde{M}^n_{l,j}(L)$ when $j = 0$. In particular, we examine a sum of their differences, since the two processes are not centered (only for $j = 0$) due to the distribution of $\{c_{i,1}\}$ (see Section 2); however, their difference is centered. The case $j = 0$ is different from the case $j \geq 1$, because $\breve{\tau}_{i,0}$ and $\tilde{\tau}_{i,0}$ are not equal for $(\xi^n - k^n)^+$ machines awaiting service at time $t = 0$ only (for all other machines $\breve{\tau}_{i,0} = \tilde{\tau}_{i,0} = 0$).

Lemma 4. *For $L \in \{G, H\}$, we have, as $n \to \infty$,*

$$\sum_{L \in \{G,H\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}} (\breve{M}^n_{l,0}(L) - \tilde{M}^n_{l,0}(L)) \quad \Rightarrow \quad 0.$$

Proof. See Section B.1. □

Remark 4. Note that (25) and (26) yield

$$\|\breve{M}^n_{l,j}(L)\|_t + \|\tilde{M}^n_{l,j}(L)\|_t \le \frac{2}{\sqrt{n}} \sum_{i=1}^{n} (1_{\{c_{i,2:j+l} \le t\}} + H^{(l)}(t - c_{i,2:j})),$$

which, in turn, implies

$$\mathbb{E}\left[\sum_{L\in\{G,H\}} \sum_{j,l=0}^{\infty} \|\breve{M}^n_{l,j}(L) - \tilde{M}^n_{l,j}(L)\|_t\right] \le 4\sqrt{n}(H * R(t) + R^{(2)}(t) + 3) < \infty. \tag{27}$$

Therefore, the infinite sum in the statement of Lemma 4 is well defined.

The following lemma is the main preliminary result. Let $\hat{\Delta}^n := \{\hat{\Delta}^n(t),\ t \ge 0\}$, where (see (24))

$$\hat{\Delta}^n(t) := \hat{X}^n(t) - \hat{Z}^n(t) + \beta^n - \int_0^t (\hat{X}^n(t-s))^+ \, \mathrm{d}\bar{P}_1(s). \tag{28}$$

Lemma 5. *We have* $\hat{\Delta}^n \Rightarrow 0$, *as* $n \to \infty$.

Proof. First, note that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=0}^{\infty}(P_1(t-\tilde{\tau}_{i,j}) - P_1(t-\breve{\tau}_{i,j})) &= \sum_{i=1}^{n}\sum_{j=0}^{\infty}\mathbb{E}\left[\Upsilon^n_{i,j}(t) \mid \breve{\tau}_{i,j}, \tilde{\tau}_{i,j}\right] = \sum_{i=1}^{n}\sum_{j=0}^{\infty}\int_0^t 1_{\{\breve{\tau}_{i,j} \le t-s < \tilde{\tau}_{i,j}\}} \, \mathrm{d}\bar{P}_1(s) \\ &= \int_0^t \sum_{i=1}^{n}\sum_{j=0}^{\infty} 1_{\{\breve{\tau}_{i,j} \le t-s < \tilde{\tau}_{i,j}\}} \, \mathrm{d}\bar{P}_1(s) = \int_0^t (X^n(t-s) - k)^+ \, \mathrm{d}\bar{P}_1(s),\end{aligned} \tag{29}$$

where the last equality follows from the fact that the double sum represents the number of machines awaiting service at time $t-s$. Combining (1), (3), (7) and (12) renders (see also Figure 1)

$$\begin{aligned}X^n(t) - Z^n(t) &= \sum_{i=1}^{n}(X^n_i(t) - Z^n_i(t)) = \sum_{i=1}^{n}\sum_{j=0}^{\infty}(1_{\{0 \le t - c_{i,1:j} - s_{i,j+1} < w_{i,1:j+1}\}} - 1_{\{0 \le t - c_{i,1:j} < w_{i,1:j}\}}) \\ &= \sum_{i=1}^{n}\sum_{j=0}^{\infty}\sum_{l=0}^{j}(1_{\{w_{i,1:l} \le t - c_{i,1:j} - s_{i,j+1} < w_{i,1:l+1}\}} - 1_{\{w_{i,1:l-1} \le t - c_{i,1:j} < w_{i,1:l}\}}) = \sum_{i=1}^{n}(X^n_i - Z^n_i)(t).\end{aligned}$$

Next, a change in the order of summation (l and j) and (25) result in

$$\begin{aligned}X^n(t) - Z^n(t) &= \sum_{i=1}^{n}\sum_{j,l=0}^{\infty}(1_{\{w_{i,1:j} \le t - c_{i,1:l+j} - s_{i,l+j+1} < w_{i,1:j+1}\}} - 1_{\{w_{i,1:j} \le t - c_{i,1:l+j+1} < w_{i,1:j+1}\}}) = \sum_{i=1}^{n}\sum_{j=0}^{\infty}\Upsilon^n_{i,j}(t) \\ &= \sum_{i=1}^{n}\sum_{j,l=0}^{\infty}\left(1_{\{c_{i,j+1:j+l} + s_{i,j+l+1} \le t - \breve{\tau}_{i,j} < c_{i,j+1:j+l+1}\}} - 1_{\{c_{i,j+1:j+l} + s_{i,j+l+1} \le t - \tilde{\tau}_{i,j} < c_{i,j+1:j+l+1}\}}\right) \\ &= \sum_{i=1}^{\infty}\sum_{j,l=0}^{\infty}\left((\breve{\vartheta}^n_{i,j,l}(G;t) - \breve{\vartheta}^n_{i,j,l}(H;t)) - (\tilde{\vartheta}^n_{i,j,l}(G;t) - \tilde{\vartheta}^n_{i,j,l}(H;t))\right);\end{aligned} \tag{30}$$

see Figure 1 for an illustration. This, together with (5), (8), (14) and (29), yields

$$\sqrt{p\bar{p}}\,\hat{\Delta}^n = \hat{\Delta}^n_{1,m} + \hat{\Delta}^n_{2,m} + \hat{\Delta}^n_{3,m}, \tag{31}$$

where

$$\begin{aligned}\hat{\Delta}^n_{1,m} &:= \sum_{L\in\{G,H\}} \sum_{\substack{j+l\le m \\ j\ne 0}} (-1)^{1_{\{L=H\}}}(\breve{M}^n_{l,j}(L) - \tilde{M}^n_{l,j}(L)), \\ \hat{\Delta}^n_{2,m} &:= \sum_{L\in\{G,H\}} \sum_{\substack{j+l> m \\ j\ne 0}} (-1)^{1_{\{L=H\}}}(\breve{M}^n_{l,j}(L) - \tilde{M}^n_{l,j}(L)), \\ \hat{\Delta}^n_{3,m} &:= \sum_{L\in\{G,H\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}}(\breve{M}^n_{l,0}(L) - \tilde{M}^n_{l,0}(L));\end{aligned}$$

observe that $\hat{\Delta}^n_{2,m}$ and $\hat{\Delta}^n_{3,m}$ are well defined due to (27) in Remark 4.

Note that, for a fixed m, Lemma 2 and the continuity of the addition operator result in, as $n \to \infty$,

$$\hat{\Delta}^n_{1,m} \Rightarrow 0. \tag{32}$$

Next, the nonnegativity of the supremum norm and Lemma 3 imply

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}[\|\hat{\Delta}^n_{2,m}\|_t > \epsilon] = 0. \tag{33}$$

Combining (31)–(33) and Lemma 4 completes the proof of the lemma. □

4.3. Proof of Theorem 1

Recall the definition (28) of $\hat{\Delta}^n$. Now, observe that

$$\hat{X}^n = \hat{Z}^n + \hat{X}^n - \hat{Z}^n = \varphi_{\bar{P}_1}(\hat{Z}^n - \beta^n + \hat{\Delta}^n), \tag{34}$$

and note that, as $n \to \infty$,

$$\hat{Z}^n - \beta^n + \hat{\Delta}^n \Rightarrow \hat{Z} - \beta,$$

due to (4), the assumption of the theorem, Lemma 5, β being deterministic (see Theorem 11.4.5 in Whitt 2002, p. 379) and the continuity of the addition operator. By the Skorokhod representation theorem (Skorokhod 1956) ($(D[0,\infty),\mathscr{D})$ is a separable space), there exists an alternative probability space with $\{\tilde{Z}^n - \beta^n + \tilde{\Delta}^n\}$ and $\tilde{Z} - \beta$ defined on it, such that

$$\tilde{Z}^n - \beta^n + \tilde{\Delta}^n \to \tilde{Z} - \beta, \tag{35}$$

in $(D[0,\infty), u)$ almost surely, as $n \to \infty$; here,

$$\tilde{Z}^n - \beta^n + \tilde{\Delta}^n \stackrel{d}{=} \hat{Z}^n - \beta^n + \hat{\Delta}^n \qquad \text{and} \qquad \tilde{Z} - \beta \stackrel{d}{=} \hat{Z} - \beta. \tag{36}$$

Furthermore, due to $\varphi_{\bar{P}_1}\colon (D[0,\infty),\mathscr{D}) \to (D[0,\infty),\mathscr{D})$ being measurable (Proposition 1), it follows from (34) and (36) that, for each $n \geq 1$,

$$\hat{X}^n \stackrel{d}{=} \varphi_{\bar{P}_1}(\tilde{Z}^n - \beta^n + \tilde{\Delta}^n). \tag{37}$$

On the other hand, the continuity part of Proposition 1 and (35) result in

$$\varphi_{\bar{P}_1}(\tilde{Z}^n - \beta^n + \tilde{\Delta}^n) \to \varphi_{\bar{P}_1}(\tilde{Z} - \beta) \quad \text{a.s.}, \tag{38}$$

in $(D[0,\infty), u)$, as $n \to \infty$. The facts that convergence in $(D[0,\infty), u)$ implies convergence in $(D[0,\infty), d_{J_1})$ and that almost sure convergence implies weak convergence, together with (37), (38) and Proposition 1 (measurability part), yield $\hat{X}^n \Rightarrow \varphi_{\bar{P}_1}(\hat{Z} - \beta)$, as $n \to \infty$. This completes the proof. □

5. Proofs

This section contains proofs of Lemma 1, Proposition 1, and Lemma 2.

5.1. Proof of Lemma 1

Let α^n be a binomial random variable with parameters (n, p), independent of all other variables in the nth model; set $\hat{\alpha}^n := (\alpha^n - np)/\sqrt{np\bar{p}}$. We construct a stationary process $\hat{Y}^n := \{\hat{Y}^n(t),\ t \geq 0\}$ by modifying the infinite-repairmen process Z^n. In particular, let

$$\hat{Y}^n := \frac{1}{\sqrt{np\bar{p}}} \sum_{i=1}^{n} (Y^n_i - p), \tag{39}$$

where

$$Y^n_i := \begin{cases} \tilde{Z}^n_i, & \xi^n \wedge k^n \wedge \alpha^n < i \leq \xi^n \vee \alpha^n, \\ Z^n_i, & \text{otherwise,} \end{cases}$$

with $\{\tilde{Z}^n_i,\ \xi^n \wedge k^n \wedge \alpha^n < i \leq \alpha^n\}$ and $\{\tilde{Z}^n_i,\ \alpha^n < i \leq \xi^n \vee \alpha^n\}$ being i.i.d. processes corresponding to a broken and working machine at time $t = 0$, respectively. Namely, let $\tilde{c}_{i,j} = \tilde{a}_{i,j} + \tilde{s}_{i,j}$, $j \geq 1$, where $\{\tilde{a}_{i,j}, j \geq 2\}$ and $\{\tilde{s}_{i,j}, j \geq 2\}$ be mutually independent i.i.d. sequences defined by distributions F and G, respectively. Furthermore, let $\tilde{s}_{i,1}$,

$i \le \alpha^n$, be distributed according to $\ddot{G}$, and $\tilde{s}_{i,1} \equiv 0$ for $i > \alpha^n$. Similarly, let $\tilde{a}_{i,1}$ be distributed according to F and $\ddot{F}$ for $i \le \alpha^n$ and $i > \alpha^n$, respectively. Then, we have

$$\tilde{Z}_i^n(t) = \sum_{j=0}^{\infty} 1_{\{0 \le t - \tilde{c}_{i,1:j} < \tilde{s}_{i,j+1}\}}.$$

Based on the above construction, it follows that

$$\mathbb{E}\tilde{Z}_i^n(t) = \begin{cases} \ddot{P}_1(t), & \xi^n \wedge k^n \wedge \alpha^n < i \le \alpha^n, \\ \ddot{P}_0(t), & \alpha^n < i \le \xi^n \vee \alpha^n. \end{cases} \tag{40}$$

Observe that, due to the preceding and the binomial distribution of α^n, the process $\hat{Y}^n$ is stationary with $\hat{Y}^n(0) = \hat{\alpha}^n$ (see (39)), and

$$\hat{Y}^n \stackrel{\mathrm{d}}{=} \tilde{Y}^n \equiv \frac{1}{\sqrt{np\bar{p}}} \sum_{i=1}^{n} (\tilde{Y}_i - p),$$

where $\{\tilde{Y}_i\}$ is a sequence of i.i.d. stationary on-off processes described by the distributions G and F (e.g., see Heath 1998). Since $\tilde{Y}_i$ is a stationary process, it follows that

$$\mathbb{E}(\tilde{Y}_i(t) - p)(\tilde{Y}_i(t+s) - p) = \mathbb{E}\tilde{Y}_i(t)\tilde{Y}_i(t+s) - p^2 = p(\ddot{P}_1(s) - p),$$

for $t, s \ge 0$. Under (13), conditions of the Hahn's theorem (Hahn 1978) (see also Whitt 2002, p. 226) hold (Szczotka 1980), and thus $\tilde{Y}^n \Rightarrow \hat{Y}$, as $n \to \infty$, where $\hat{Y}$ is as in the statement of the lemma; consequently, $\hat{Y}^n \Rightarrow \hat{Y}$, as $n \to \infty$. Moreover, due to the continuous mapping theorem and continuity of the mapping $(D[0,\infty), u) \to (D^2[0,\infty), u^2)$ such that $x \mapsto (x, x(0)(\ddot{P}_0 - \ddot{P}_1))$, for $x \in D[0,\infty)$, one has $(\hat{Y}^n, \hat{\alpha}^n(\ddot{P}_0 - \ddot{P}_1)) \Rightarrow (\hat{Y}, \hat{Y}(0)(\ddot{P}_0 - \ddot{P}_1))$, as $n \to \infty$, which, in turn, implies Whitt (1980)

$$\hat{Y}^n + \hat{\alpha}^n(\ddot{P}_0 - \ddot{P}_1) \Rightarrow \hat{Y} + \hat{Y}(0)(\ddot{P}_0 - \ddot{P}_1), \tag{41}$$

as $n \to \infty$, since $\hat{Y}$ is a Gaussian process.

Next, let $\mathbb{E}Z_i^n := \{\mathbb{E}Z_i^n(t), t \ge 0\}$ and $\mathbb{E}\tilde{Z}_i^n := \{\mathbb{E}\tilde{Z}_i^n(t), t \ge 0\}$. Note that, based on (4), (6), (11) and (40), one can conclude that

$$\sum_{i=\xi^n \wedge k^n \wedge \alpha^n + 1}^{\xi^n \vee \alpha^n} (\mathbb{E}Z_i^n - \mathbb{E}\tilde{Z}_i^n) - \alpha^n(\ddot{P}_0 - \ddot{P}_1) = (\xi^n \wedge k^n)\ddot{P}_1 + (\xi^n - k^n)^+ P_1 - \xi^n \ddot{P}_0 \tag{42}$$

and, therefore,

$$\frac{1}{\sqrt{np\bar{p}}} \sum_{i=\xi^n \wedge k^n \wedge \alpha^n + 1}^{\xi^n \vee \alpha^n} (\mathbb{E}Z_i^n - \mathbb{E}\tilde{Z}_i^n) - \hat{\alpha}^n(\ddot{P}_0 - \ddot{P}_1) \Rightarrow (\beta - \hat{\xi}^-)\ddot{P}_1 + \hat{\xi}^+ P_1 - (\hat{\xi} + \beta)\ddot{P}_0, \tag{43}$$

as $n \to \infty$. The independence of ξ^n and α^n, Lemma 6 (at the end of this subsection) and

$$\frac{1}{\sqrt{np\bar{p}}}(\xi^n \vee \alpha^n - \xi^n \wedge k^n \wedge \alpha^n) \Rightarrow \hat{\xi} \vee (\hat{\alpha} - \beta) + \hat{\xi}^- \vee (\hat{\alpha} - \beta)^-,$$

as $n \to \infty$, yield, as $n \to \infty$,

$$\frac{1}{\sqrt{np\bar{p}}} \sum_{i=\xi^n \wedge k^n \wedge \alpha^n + 1}^{\xi^n \vee \alpha^n} (Z_i^n - \mathbb{E}Z_i^n - \tilde{Z}_i^n + \mathbb{E}\tilde{Z}_i^n) \Rightarrow 0. \tag{44}$$

Finally, (14) and (39) imply

$$\hat{Z}^n = \hat{Y}^n + (\hat{Z}^n - \hat{Y}^n) = \hat{Y}^n + \frac{1}{\sqrt{np\bar{p}}} \sum_{i=\xi^n \wedge k^n \wedge \alpha^n + 1}^{\xi^n \vee \alpha^n} (Z_i^n - \tilde{Z}_i^n). \tag{45}$$

Combining (41), (43)–(45), Theorems 11.4.4 and 11.4.5 in Whitt (2002, pp. 378–379) (note that in view of (42), the left-hand side of (43) and $\hat{\alpha}^n$ are independent) and the continuous mapping theorem (and its extension for summation Whitt 1980) yields the statement of the lemma. □

Lemma 6. *Let $\{U_i\}$ be an i.i.d. sequence of processes and $u := \{\mathbb{E}U_i(t),\ t \ge 0\}$. Suppose $\|U_i - u\|_t \le 1$ and*

$$S_n := \sum_{i=1}^{n}(U_i - u).$$

If $\{\zeta^n\}$ is a sequence of nonnegative random variables independent of $\{U_i\}$ such that $\zeta^n \Rightarrow \zeta$, as $n \to \infty$, then, as $n \to \infty$,

$$\|S_{\lfloor \zeta^n n\rfloor}/n\|_t \xrightarrow{\mathbb{P}} 0.$$

Proof. Let $\delta_1 < \epsilon$ be a continuity point of the distribution of ζ. For $0 < \delta_1 < \delta_2 < \delta_3$ and $\epsilon > 0$, the union bound and $\|S_{\lfloor \zeta^n n\rfloor}/n\|_t \le |\zeta^n|$ result in

$$\mathbb{P}[\|S_{\lfloor \zeta^n n\rfloor}/n\|_t > \epsilon] \le \mathbb{P}[\delta_1 < \zeta^n < \delta_2] + \mathbb{P}\left[\sup_{\delta_2 < x < \delta_3} \|S_{\lfloor xn\rfloor}/\lfloor xn\rfloor\|_t > \frac{\epsilon}{\delta_3}\right] + \mathbb{P}[\zeta^n > \delta_3];$$

the second term on the right-hand side vanishes, as $n \to \infty$, due to Daffer and Taylor (1979). Letting $n \to \infty$, and then letting $\delta_2 \to \delta_1$ and $\delta_3 \to \infty$ yields the statement of the lemma. □

5.2. Proof of Proposition 1

The proof of the proposition proceeds along the lines of the proof of the corresponding proposition for φ_G in Reed (2009). The measurability part is argued at the end of the proof. The rest of the proof covers three cases: degenerate G and F; nondegenerate G; and degenerate G and nondegenerate F (these cases impact properties of $\bar{P}_1$).

(*Degenerate G and F*). Assume that G and F are concentrated at c_G and c_F, respectively. Then $\bar{P}_1$ is periodic with period $c = c_G + c_F$:

$$\bar{P}_1(t) = \sum_{l=0}^{\infty} 1_{\{lc + c_G \le t < lc + c\}}.$$

A solution to (16) satisfies the following recursion: $y(t) = x(t)$, for $0 \le t < c_G$, and

$$y(t) = x(t) + \sum_{i=1}^{\lfloor (t+c_F)/c\rfloor} y^+(t - (i-1)c - c_G) - \sum_{i=1}^{\lfloor t/c\rfloor} y^+(t - ic), \tag{46}$$

for $t \ge c_G$; this implies existence and uniqueness. Let $y_0 := 0$ and recursively define

$$y_{m+1}(t) := x(t) + \int_0^t y_m^+(t-s)\,\mathrm{d}\bar{P}_1(s), \quad t \ge 0, \tag{47}$$

for $m \ge 0$; that is, $y_m \mapsto y_{m+1}$. Then, from (46) it follows that $\|y_{m+1} - y_m\|_t = 0$ for all $m \ge \lceil t/(c_G \wedge c_F)\rceil$. In addition, (46) results in $\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t = \|x_1 - x_2\|_t$, for $0 \le t < c_G$, and $\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t \le \|x_1 - x_2\|_t + \|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_{t-c_G}$ for $c_G \le t < c$ (see (46)). If $t - c_G \ge c_G$, this argument can be applied multiple times to obtain $\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t \le \lceil c/c_G\rceil \|x_1 - x_2\|_t$; the last inequality serves as a basis for our induction. Now, suppose for every positive integer $l \le j$ for some integer j one has

$$\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t \le \begin{cases} a_l\|x_1 - x_2\|_t, & (l-1)c \le t < (l-1)c + c_G, \\ b_l\|x_1 - x_2\|_t, & (l-1)c + c_G \le t < lc. \end{cases} \tag{48}$$

Then, the inductive hypothesis (48) and (46) imply, for $jc \le t < jc + c_G$,

$$\begin{aligned}\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t &\le \|x_1 - x_2\|_t + \sum_{i=1}^{j} \|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_{t-(i-1)c-c_G} + \sum_{i=1}^{j} \|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_{t-ic} \\ &\le \|x_1 - x_2\|_t + \sum_{i=1}^{j} a_i\|x_1 - x_2\|_t + \sum_{i=1}^{j} b_i\|x_1 - x_2\|_t = (1 + a_{1:j} + b_{1:j})\|x_1 - x_2\|_t \equiv a_{j+1}\|x_1 - x_2\|_t;\end{aligned}$$

also, for $jc + c_G \le t < (j+1)c$,

$$\begin{aligned}\|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_t &\le \|x_1 - x_2\|_t + \sum_{i=1}^{j+1} \|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_{t-(i-1)c-c_G} + \sum_{i=1}^{j} \|\varphi_{\bar{P}_1}(x_1) - \varphi_{\bar{P}_1}(x_2)\|_{t-ic} \\ &\le \|x_1 - x_2\|_t + \sum_{i=1}^{j+1} a_i\|x_1 - x_2\|_t + \sum_{i=1}^{j} b_i\|x_1 - x_2\|_t = (1 + a_{1:(j+1)} + b_{1:j})\|x_1 - x_2\|_t \equiv b_{j+1}\|x_1 - x_2\|_t.\end{aligned}$$

Therefore, the inductive assumption (48) holds for all j, and $\varphi_{\bar{P}_1}$ is Lipschitz continuous when G and F are degenerate.

(*Nondegenerate G*). To show uniqueness, we use the method of successive approximations. Recall (47); that definition implies

$$y_{m+1}(t) - y_m(t) = \int_0^t (y_m^+(t-s) - y_{m-1}^+(t-s))\,\mathrm{d}\bar{P}_1(s). \tag{49}$$

In Lemma 7 (see the end of this subsection) select a $\delta > 0$ such that t/δ is an integer. Next, we use induction to argue that, for each integer $1 \le k \le t/\delta$,

$$\|y_{m+1} - y_m\|_{j\delta} \le j^j m^j \epsilon^m \|x\|_t, \quad j = 1, \ldots, k, \;\; m \ge 1. \tag{50}$$

The case $k = 1$ is straightforward: $\|y_1 - y_0\|_\delta = \|x - 0\|_\delta \le \|x\|_t$, (49) and Lemma 7 result in $\|y_{m+1} - y_m\|_\delta \le \epsilon \|y_m - y_{m-1}\|_\delta$, which, in turn, renders $\|y_{m+1} - y_m\|_\delta \le \epsilon^m \|x\|_t \le m\epsilon^m \|x\|_t$, for $m \ge 1$. Now, assume that (50) holds for some k. Based on this inductive hypothesis, Lemma 7 and (49), one has

$$\begin{aligned}\|y_{m+1} - y_m\|_{(k+1)\delta} &\le \sum_{j=1}^{k} \epsilon \|y_m - y_{m-1}\|_{j\delta} + \epsilon \|y_m - y_{m-1}\|_{(k+1)\delta} \\ &\le \sum_{j=1}^{k} j^j (m-1)^j \epsilon^m \|x\|_t + \epsilon \|y_m - y_{m-1}\|_{(k+1)\delta} \\ &\le k^{k+1} m^k \epsilon^m \|x\|_t + \epsilon \|y_m - y_{m-1}\|_{(k+1)\delta}.\end{aligned} \tag{51}$$

Note that $\|y_1 - y_0\|_{l\delta} = \|x - 0\|_{l\delta} \le \|x\|_t$, for all $l = 1, \ldots, t/\delta$, by definition. This, combined with (51), yields

$$\|y_{m+1} - y_m\|_{(k+1)\delta} \le k^{k+1} \epsilon^m \|x\|_t \sum_{j=0}^{m} j^k \le k^{k+1}(m^{k+1} + 1)\epsilon^m \|x\|_t \le (k+1)^{k+1} m^{k+1} \epsilon^m \|x\|_t.$$

Hence, the inductive hypothesis holds. Finally, $\{y_m\}$ is a Cauchy sequence, because

$$\sum_{m=1}^{\infty} \|y_{m+1} - y_m\|_t \le \sum_{m=1}^{\infty} \|y_{m+1} - y_m\|_{(t/\delta)\delta} \le \sum_{m=1}^{\infty} (t/\delta)^{t/\delta} m^{t/\delta} \epsilon^m \|x\|_t < \infty.$$

In view of the fact that $D[0,\infty)$ equipped with the supremum metric is a Banach space (but not under J_1 Whitt 2002, p. 84), there exists a limit point y of $\{y_m\}$ (e.g., see Rudin 1991, p. 4). Letting $m \to \infty$ on both sides of (47), one concludes that the limit y is a solution to (16). This completes the proof of existence in case of nondegenerate distribution G.

As far as uniqueness is concerned, suppose that both u and y solve (16), and define $\Delta := \{\Delta(t),\, t \ge 0\}$, where

$$\Delta(t) := u(t) - y(t) = \int_0^t (u^+(t-s) - y^+(t-s))\,\mathrm{d}\bar{P}_1(s), \quad t \ge 0.$$

Then, Lemma 7 yields

$$\|\Delta\|_\delta \le \sup_{0 \le t \le \delta} \left| \int_0^t (u^+(t-s) - y^+(t-s))\,\mathrm{d}\bar{P}_1(s) \right| \le \epsilon \|\Delta\|_\delta,$$

which implies $\Delta(t) = 0$ for $t \in [0,\delta]$. Similarly, one has $\|\Delta\|_{2\delta} \le \epsilon\|\Delta\|_\delta + \epsilon\|\Delta\|_{2\delta} = \epsilon\|\Delta\|_{2\delta}$, and, hence, $\Delta(t) = 0$ for $t \in [0, 2\delta]$. Iterating the above argument multiple times completes the proof of uniqueness.

Finally, we argue Lipschitz continuity. Note that by Lemma 7 one has, for $0 \le t < \delta$,

$$\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \le \|x_2 - x_1\|_t + \epsilon \|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t,$$

which implies $\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \le (1-\epsilon)^{-1}\|x_2 - x_1\|_t$. Similarly, for $\delta \le t < 2\delta$, one has

$$\begin{aligned}\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t &\le \|x_2 - x_1\|_t + \epsilon\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_{t-\delta} + \epsilon\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \\ &\le (1-\epsilon)^{-1}\|x_2 - x_1\|_t + \epsilon\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t.\end{aligned}$$

which, in turn, implies $\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \le (1-\epsilon)^{-2}\|x_2 - x_1\|_t$, for $0 \le t < 2\delta$. Applying the above argument multiple times completes this part of the proof.

(*Degenerate G and nondegenerate F*). Suppose G is concentrated at c_G. From Lemma 7, it follows that $\bar{P}_1(t) = B(t) - 1_{\{t<c_G\}}$, where B is such that (i) $B(t) = 1$ for $t \le c_G$, and (ii) there exist $\delta > 0$ and $\epsilon < 1$ such that, for every $t \in [0, \infty)$,

$$\sup_{t \le t_1, t_2 \le t+\delta} |B(t_1) - B(t_2)| < \epsilon. \tag{52}$$

Therefore, (16) can be rewritten:

$$y(t) = \begin{cases} x(t), & t < c_G, \\ x(t) + y^+(t - c_G) + \int_{c_G}^t y^+(t-s)\,\mathrm{d}B(s), & t \ge c_G. \end{cases} \tag{53}$$

From the preceding equation, it is apparent that the solution exists and that it is unique for $t < c_G$. Now, suppose that, for some integer $i \ge 1$, y_i is the unique solution to (53) for $t < ic_G$. We argue that there exists a unique y_{i+1} satisfying (53) for $t < (i+1)c_G$. To this end, let

$$x_i(t) := \begin{cases} x_{i-1}(t), & t < ic_G, \\ x(t) + y_{i-1}^+(t - c_G), & ic_G \le t < (i+1)c_G, \end{cases}$$

where $x_0 \equiv x$. Observe that $\{x_i(t), 0 \le t < ic_G\}$ is uniquely defined, since there exists a unique y_i for $t < ic_G$. Then, one can rewrite the second part of (53) as follows:

$$y(t) = x_i(t) + \int_{c_G}^t y^+(t-s)\,\mathrm{d}B(s), \quad t < (i+1)c_G. \tag{54}$$

The same argument used to show existence and uniqueness of the solution in the case of nondegenerate G is applicable in the case of (54). In particular, (52) should be used instead of Lemma 7. Hence, (54) has the unique solution on $t < (i+1)c_G$—term this solution y_{i+1}. Moreover, one has $y_{i+1}(t) = y_i(t)$ for $t < ic_G$, since y_i is unique on $[0, ic_G)$, and y_{i+1} and y_i solve the same equation on $[0, ic_G)$. Iterating the preceding reasoning multiple times yields that (16) has a unique solution when G is degenerate and F is nondegenerate.

The proof of Lipschitz continuity utilizes (53). It is straightforward that $\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \le \|x_2 - x_1\|_t$, for $t < c_G$. Now, suppose $\|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t \le e_i\|x_2 - x_1\|_t$, for $(i-1)c_G \le t < ic_G$, i.e., $e_1 = 1$. Then, (53) implies, for $ic_G \le t < (i+1)c_G$,

$$\begin{aligned} \|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_t &\le \|x_2 - x_1\|_t + \|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_{ic_G} + \lceil c_G/\delta \rceil \epsilon \|\varphi_{\bar{P}_1}(x_2) - \varphi_{\bar{P}_1}(x_1)\|_{ic_G} \\ &\le ((\lceil c_G/\delta \rceil \epsilon + 1)e_i + 1)\|x_2 - x_1\|_t \equiv e_{i+1}\|x_2 - x_1\|_t. \end{aligned}$$

This completes this part of the proof.

(*Measurability*). Let $\phi_P\colon D[0,\infty) \to D[0,\infty)$ be defined by the following, for $t \ge 0$:

$$\phi_P(y)(t) := \int_0^t y^+(t-s)\,\mathrm{d}P(s).$$

The proof consists of two parts. In the first part, measurability of $\phi_{\bar{P}_1}$ is argued (note that ϕ_P is "simpler" than φ_P). The second part deals with measurability of the original operator $\varphi_{\bar{P}_1}$.

For the first part, one needs to argue that $\phi_{\bar{P}_1}$ is measurable with respect to the Borel σ-field $\mathscr{D}$ generated by the Skorokhod J_1 topology. Given that $\mathscr{D}$ is equal to the Kolmogorov σ-field generated by the finite-dimensional cylinder sets (Whitt 2002, p. 385), it is sufficient to check that for each $m \ge 1$ and $A_1, \ldots, A_m \in \mathscr{B}(\mathbb{R})$ one has $\{y \in D[0,\infty)\colon (\phi_{\bar{P}_1}(y)(t_1), \ldots, \phi_{\bar{P}_1}(y)(t_m)) \in (A_1, \ldots, A_m)\} \in \mathscr{D}$, for $0 \le t_1 < t_2 < \cdots < t_m$. Moreover, it is sufficient to verify that $\phi_{\bar{P}_1}(\cdot)(t)$ is measurable, for all $t \ge 0$, since σ-fields are closed under finite intersections. To this end, we consider a decomposition of $\bar{P}_1$ (see (8)):

$$\begin{aligned} \bar{P}_1 &= G * R - H * R \\ &= \mathsf{C}\{G * R\} - \mathsf{C}\{H * R\} + \mathsf{D}\{G * R\} - \mathsf{D}\{H * R\} \\ &\equiv \bar{P}_\mathsf{C} + \bar{P}_\mathsf{D}, \end{aligned} \tag{55}$$

where by $\mathsf{C}\{\cdot\}$ and $\mathsf{D}\{\cdot\}$ denote continuous and discrete parts, respectively (both $G*H$ and $H*R$ are monotonic). In what follows, it is shown that both $\phi_{\bar{P}_\mathsf{C}}$ and $\phi_{\bar{P}_\mathsf{D}}$ are measurable functions, and thus $\phi_{\bar{P}_1} = \phi_{\bar{P}_\mathsf{C}} + \phi_{\bar{P}_\mathsf{D}}$ is measurable as well, since the sum of two measurable functions $(D[0,\infty), \mathscr{D}) \to (D[0,\infty), \mathscr{D})$ is measurable (Dudley 2002, p. 119).

As far as measurability of $\phi_{\bar{P}_{\mathsf{C}}}$ is concerned, it is enough to show that $\phi_{\bar{P}_{\mathsf{C}}}(\cdot)(t)$ is a continuous $(D[0,\infty), d_{J_1}) \to (\mathbb{R}, |\cdot|)$ function, for each $t \geq 0$. Let $\{y_m\}$ be a sequence, such that $y_m \to y$ under the J_1 metric. This implies that $y_m(t) \to y(t)$ for all but a countable number of t (e.g., see Billingsley 1999, p. 247). In addition, by the definition of $\bar{P}_{\mathsf{C}}$ above, $\int_{\mathscr{S}} \mathrm{d}\bar{P}_{\mathsf{C}} = 0$ for any countable set $\mathscr{S}$. Thus, since for each $t \geq 0$, the sequence $\{\|y_m\|_t, m \geq 1\}$ is bounded, it follows that

$$|\phi_{\bar{P}_{\mathsf{C}}}(y_m)(t) - \phi_{\bar{P}_{\mathsf{C}}}(y)(t)| = \left|\int_0^t (y_m^+(t-s) - y^+(t-s))\,\mathrm{d}\bar{P}_{\mathsf{C}}(s)\right| \leq \int_0^t |y_m(t-s) - y(t-s)|\,\mathrm{d}\bar{P}_{\mathsf{C}}(s) \to 0,$$

as $m \to \infty$, i.e., $\phi_{\bar{P}_{\mathsf{C}}}$ is measurable.

Let $\{d_i\}$ be the set of discontinuity points of $\bar{P}_{\mathsf{D}}$ on $[0,t]$. This set is countable, since it is a subset of countable union of sets of discontinuity points of $G * H^{(i)}$ and $H^{(i)}$ (assuming the axiom of choice); the set of discontinuity points of $G * H^{(i)}$ or $H^{(i)}$, for each i, is a countable set, because they are distribution functions. For measurability of $\phi_{\bar{P}_{\mathsf{D}}}$, two cases will be considered. First, suppose that the cardinality of $\{d_i\}$ is $m < \infty$. In that case, $\bar{P}_{\mathsf{D}}$ is of the form $\bar{P}_{\mathsf{D}}(t) = \sum_{i=1}^m c_i 1_{\{t \geq d_i\}}$, and, therefore,

$$\phi_{\bar{P}_{\mathsf{D}}}(y)(t) = \sum_{i=1}^m c_i y^+(t-d_i) 1_{\{t \geq d_i\}}.$$

That is, $\phi_{\bar{P}_{\mathsf{D}}}$ is a finite sum of measurable functions (translation by a constant and multiplication by a constant as well their composition are measurable), making it measurable as well. When the number of points in $\{d_i\}$ is infinite, then

$$\phi_{\bar{P}_{\mathsf{D}}}(y)(t) = \sum_{i=1}^\infty c_i y^+(t-d_i) 1_{\{t \geq d_i\}}.$$

Next, define

$$\phi^m_{\bar{P}_{\mathsf{D}}}(y)(t) := \sum_{i=1}^m c_i y^+(t-d_i) 1_{\{t \geq d_i\}},$$

in order to consider, for every $y \in D[0,\infty)$ and $t \geq 0$,

$$\|\phi^m_{\bar{P}_{\mathsf{D}}}(y) - \phi_{\bar{P}_{\mathsf{D}}}(y)\|_t = \sup_{0 \leq s \leq t}\left|\sum_{i=m+1}^\infty c_i y^+(s-d_i) 1_{\{s \geq d_i\}}\right| \leq \|y\|_t \sum_{i=m+1}^\infty |c_i| 1_{\{s \geq d_i\}} \to 0,$$

as $m \to \infty$; the limit is justified by (see (55))

$$\sum_{i=1}^\infty |c_i| 1_{\{t \geq d_i\}} \leq G * R(t) + H * R(t) < \infty.$$

Therefore, $\phi_{\bar{P}_{\mathsf{D}}}$ is a pointwise limit of $\{\phi^m_{\bar{P}_{\mathsf{D}}}\}_m$ in $(D[0,\infty), u)$. Since $\phi^m_{\bar{P}_{\mathsf{D}}}$'s are measurable, it means that $\phi_{\bar{P}_{\mathsf{D}}}$ is measurable as well (Dudley 2002, p. 125).

Now, we are ready to start the second part of the measurability proof. The basic idea is to exploit the preceding part as well as the part of the proof that relates to existence. Let Ξ: $(D[0,\infty), \mathscr{D}) \to (D[0,\infty), \mathscr{D})$ be defined by, for $t \geq 0$,

$$\Xi(y)(t) := x(t) + \phi_{\bar{P}_1}(y)(t);$$

Ξ is measurable since $\phi_{\bar{P}_1}$ is measurable. The method of successive approximations (see the existence part of the proof) yields, for each $x \in D[0,\infty)$,

$$\varphi_{\bar{P}_1}(x) = \lim_{m\to\infty} \Xi^m(0),$$

where Ξ^m is the m-fold composition of Ξ with itself, and the limit is taken with respect to the uniform metric u. Note that Ξ^m is measurable for every m, because the composition of two measurable functions is measurable (Billingsley 1995, p. 182). This, in turn, implies that $\varphi_{\bar{P}_1}$ measurable, since a pointwise limit of measurable functions is measurable. This completes the proof. □

Lemma 7. *For nondegenerate G, there exist $\delta > 0$ and $\epsilon < 1$ such that*

$$\sup_{\substack{|t_1 - t_2| < \delta \\ t_1, t_2 \in [0,T]}} |P_1(t_1) - P_1(t_2)| \leq \epsilon.$$

Proof. Equation (8), together with monotonicity of R and $G*R$, implies

$$|P_1(t_1)-P_1(t_2)| \le |R(t_1)-R(t_2)| \vee |G*R(t_1)-G*R(t_2)|$$

and

$$\sup_{\substack{|t_1-t_2|<\delta\\ t_1,t_2\in[0,T]}} |P_1(t_1)-P_1(t_2)| \le \sup_{t\in[\delta,T]}\{R(t)-R(t-\delta)\} \vee \sup_{t\in[\delta,T]}\{G*R(t)-G*R(t-\delta)\}.$$

Therefore, in order to prove the lemma, it is sufficient to show that there exist $\delta>0$ and $\epsilon<1$ such that

$$\sup_{t\in[\delta,T]}\{R(t)-R(t-\delta)\}\le\epsilon \qquad \text{and} \qquad \sup_{t\in[\delta,T]}\{G*R(t)-G*R(t-\delta)\}\le\epsilon. \tag{56}$$

We start with the first inequality in (56). This part of the proof has several steps. First, note that the first equation in (56) holds if

$$\sup_{t\in(0,T]}\{R(t)-R(t-)\}<1. \tag{57}$$

Here, we remark that there exists a $t_\star$ that achieves the supremum in (57): $\epsilon_0 := R(t_\star)-R(t_\star-)$. The case $\epsilon_0=0$ is trivial. Otherwise ($\epsilon_0>0$), there exist only finitely many t_i's such that $R(t_i)-R(t_i-)>\epsilon_0/2$—in fact, at most $\lceil 2R(T)/\epsilon_0\rceil$ such points exist. The conclusion follows.

The second step establishes a bound on jumps in R. To this end, one has

$$H^{(i+1)}(t)=\int_{[0,\theta]} H^{(i)}(t-s)\,\mathrm{d}H(s)+\int_{(\theta,t]} H^{(i)}(t-s)\,\mathrm{d}H(s)\le H(\theta)+H^{(i)}(t-\theta),$$

for $\theta>0$ and $i\ge1$. Letting $\theta\to0$ and recalling $H(0)=F(0)\,G(0)=0$ yields $H^{(i+1)}(t)\le H^{(i)}(t-)$. Now, the definition of R and $R(t)<\infty$ imply

$$R(t)-R(t-)=H(t)+\sum_{i=1}^{\infty}(H^{(i+1)}(t)-H^{(i)}(t-))\le H(t)+\sum_{i=1}^{m}(H^{(i+1)}(t)-H^{(i)}(t-)), \tag{58}$$

for any $m\ge0$; the inequality follows from the nonpositivity of the elements of the sum.

The third step establishes properties of H, when $H(T)=1$. For that purpose, define $s_\star := \inf\{s\colon H(s)=1\}<\infty$, and note that $H(s_\star-\theta)<1$, for any $\theta>0$. Due to $1\ge H^{(m)}(ms_\star)\ge H^m(s_\star)=1$ and $H^{(m)}(ms_\star-\theta)\le 1-\bar{H}^m(s_\star-\theta/m)$ <1, the constant $s_\star$ satisfies

$$ms_\star=\inf\{s\colon H^{(m)}(s)=1\}. \tag{59}$$

The nondegenerate nature of H (since G is nondegenerate by assumption) ensures that $0=H(0)<H(s_\star-)$, i.e., there exists a $\theta>0$ such that $H(s_\star-)-H(\theta)>0$. Then, conditioning yields, for some $\theta>0$,

$$H^{(m+1)}((m+1)s_\star-)-H^{(m+1)}(ms_\star)\ge(H(s_\star-)-H(\theta))(H^{(m)}(ms_\star)-H^{(m)}(ms_\star-\theta))>0,$$

where the last (strict) inequality follows from (59). Hence, since $H^{(m+1)}$ is right-continuous, there exists a $\theta>0$ small enough such that

$$H^{(m+1)}((m+1)s_\star-)>H^{(m+1)}(ms_\star+\theta). \tag{60}$$

Now, by using $[0,s_\star]=[0,\eta)\cup[\eta,s_\star-\theta)\cup[s_\star-\theta,s_\star]$ one has

$$H^{(m+1)}(ms_\star)=\int_0^{s_\star}H^{(m)}(ms_\star-s)\,\mathrm{d}H(s)\le H(\eta)+H^{(m)}(ms_\star-\eta)H(s_\star-\theta)+H^{(m)}((m-1)s_\star+\theta)\bar{H}(s_\star-\theta),$$

which, after letting $\eta\downarrow0$, renders

$$H^{(m+1)}(ms_\star)\le H^{(m)}(ms_\star-)H(s_\star-\theta)+H^{(m)}((m-1)s_\star+\theta)\bar{H}(s_\star-\theta)<H^{(m)}(ms_\star-), \tag{61}$$

where the second inequality is due to (60).

The fourth step addresses (56). The case $H(T)<1$ is straightforward: (58) yields $R(t)-R(t-)\le H(t)\le H(T)<1$, for all $t\in(0,T]$, and the desired result follows. On the other hand, the case $H(T)=1$ is slightly more complicated. In particular, recall from the first step that there exists a $t_\star\in(0,T]$ that achieves the supremum in (56). When $t_\star/s_\star\in\{1,2,\ldots,\lfloor T/s_\star\rfloor\}$ (that is, $t_\star=ms_\star$), (58), (59) and (61) yield

$$R(t_\star)-R(t_\star-)\le 1+H^{(m+1)}(ms_\star)-H^{(m)}(ms_\star-)<1. \tag{62}$$

However, when $t_\star/s_\star \notin \{1,2,\dots,\lfloor T/s_\star\rfloor\}$, it must be that $t_\star/s_\star \in ((m-1),m)$ for some m. Therefore, (58) and (59) result in

$$R(t_\star)-R(t_\star-)\le H(t)+\sum_{i=1}^{m-1}(H^{(i+1)}(t)-H^{(i)}(t-))=H^{(m)}(t_\star)<1. \tag{63}$$

Combining (62) and (63) implies that the first equation in (56) holds.

Finally, we argue that the second equation in (56) holds as well. To this end, let $\delta_1>0$ be such that

$$\sup_{t\in[\delta_1,T]}\{G(t)-G(t-\delta_1)\}=\epsilon_1<1;$$

such a δ_1 exists since G is nondegenerate by the assumption of the lemma. Also, let δ_2 be such that

$$\sup_{t\in[\delta_2,T]}\{R(t)-R(t-\delta_2)\}=\epsilon_2<1;$$

see the first part of the proof of this lemma. Then, for $0<\delta\le\delta_1\wedge\delta_2$ and all $t\in[\delta,T]$, one has

$$\begin{aligned}G*R(t)-G*R(t-\delta)&\le\int_0^{t-\delta}(R(t-s)-R(t-\delta-s))\,\mathrm{d}G(s)+\int_{t-\delta}^{t}R(t-s)\,\mathrm{d}G(s)\\&\le\epsilon_2G(t-\delta)+(\epsilon_1\wedge\bar G(t-\delta))R(\delta)\le\epsilon_2(1-\epsilon_1)+\epsilon_1R(\delta),\end{aligned} \tag{64}$$

where the last inequality is obtained by from considering which element achieves the minimum in $\epsilon_1\wedge\bar G(t-\delta)$ and $R(\delta)\ge1>\epsilon_2$. Since $R(0)=1$ and R is right-continuous, it follows that

$$\epsilon_2(1-\epsilon_1)+\epsilon_1R(\delta)\to1-(1-\epsilon_1)(1-\epsilon_2)<1, \tag{65}$$

as $\delta\downarrow0$. Combining (64) and (65) yields the existence of a $\delta_\star>0$ such that

$$\sup_{t\in[\delta_\star,T]}\{G(t)-G(t-\delta_\star)\}<1,$$

i.e., (56) holds, as does the statement of the lemma. □

5.3. Proof of Lemma 2

Before we turn our attention to the proof of the lemma, we state a few results that relate to fluid limits of our model. To this end, let $\check X^n:=X^n/n$, $\check Z^n:=Z^n/n$ and $\check\Delta^n:=\{\check\Delta^n(t),\,t\ge0\}$, where

$$\check\Delta^n(t):=\check X^n-\check Z^n-\int_0^t(\check X^n-k^n/n)^+\,\mathrm{d}\bar P_1(s). \tag{66}$$

Our first lemma shows that $\check\Delta^n$ vanishes asymptotically.

Lemma 8. *We have* $\check\Delta^n\Rightarrow0$, *as* $n\to\infty$.

Proof. First, the same straightforward algebra as in the proof of Lemma 5 (see (29) and (30)) yields

$$\check\Delta^n=\frac{1}{\sqrt n}\sum_{L\in\{G,H\}}\sum_{j,l=0}^{\infty}(-1)^{1_{\{L=H\}}}(\check M^n_{l,j}(L)-\tilde M^n_{l,j}(L)).$$

Therefore, for a fixed m, the preceding equality, $\tilde\tau_{i,j}\ge\check\tau_{i,j}\ge c_{i,1:j}$ (see (25)) and monotonicity result in

$$\begin{aligned}|\check\Delta^n(t)|&\le2\frac{(\xi^n-k^n)^+}{n}+\frac{1}{\sqrt n}\sum_{L\in\{G,H\}}\sum_{\substack{j+l\le m\\ j>0}}(|\check M^n_{l,j}(L;t)|+|\tilde M^n_{l,j}(L;t)|+\frac4n\sum_{i=1}^{n}\sum_{\substack{j+l>m\\ j>0}}(1_{\{c_{i,1:j+l}\le t\}}+H^{(l)}(t-c_{i,1:j}))\\&=:2\frac{(\xi^n-k^n)^+}{n}+\check\Delta^n_{1,m}(t)+\check\Delta^n_{2,m}(t);\end{aligned} \tag{67}$$

set $\check\Delta^n_{1,m}:=\{\check\Delta^n_{1,m}(t),t\ge0\}$ and $\check\Delta^n_{2,m}:=\{\check\Delta^n_{2,m}(t),t\ge0\}$. In the preceding inequality, we used (i) the fact that $\check\tau_{i,0}$ and $\tilde\tau_{i,0}$ differ for $(\xi^n-k^n)^+$ indices $(\xi^n\wedge k^n<i\le\xi^n)$; and (ii) the fact that a centered on-off process is bounded in absolute value by 1.

Second, recall that $\{c_{i,1}\}$ are independent by construction, with the distribution $\ddot{G} * F$ $(1 \le i \le \xi^n \wedge k^n)$, $G * F$ $(\xi^n \wedge k^n < i \le \xi^n)$ or $\ddot{F}$ $(i > \xi^n)$. A strong law of large numbers for $D[0,1]$-valued random variables (Daffer and Taylor 1979) implies (almost surely), as $n \to \infty$,

$$\|\check{\Delta}^n_{2,m}\|_t \to 8 \sum_{\substack{j+l>m \\ j>0}} \ddot{H} * H^{(j+l-1)}(t). \tag{68}$$

In addition, we have

$$\sum_{\substack{j+l>m \\ j>0}} \ddot{H} * H^{(j+l-1)}(t) \le \sum_{i>m} i H^{(i-1)}(t) = m H^{(m)} * R(t) + H^{(m)} * R^{(2)}(t) \to 0, \tag{69}$$

as $m \to \infty$; the limit is due to the fact that $H^{(m)}(t)$ decreases exponentially in m (since $H^{(m)}(t) \le (H(t))^m$). Relations (68) and (69) result in, for any $\epsilon > 0$,

$$\lim_{m\to\infty} \lim_{n\to\infty} \mathbb{P}[\|\check{\Delta}^n_{2,m}\|_t > \epsilon] = 0. \tag{70}$$

On the other hand, processes $\check{M}^n_{l,j}(L;t)$ and $\tilde{M}^n_{l,j}(L;t)$ are tight (see Lemma 12 and Remark 7 in Appendix B.2), and hence $\check{\Delta}^n_{1,m} \Rightarrow 0$, as $n \to \infty$. The preceding limit, together with (6), (67) and (70), yields the statement of the lemma. □

Corollary 1 (Fluid Limit). *We have $\check{X}^n \Rightarrow p$, as $n \to \infty$.*

Proof. Recall Definition 1. From (66), it follows that

$$\check{X}^n = \varphi_{\bar{P}_1}(\check{Z}^n + \check{\Delta}^n - k^n/n) + k^n/n.$$

Lemma 8 ($\check{\Delta}^n \Rightarrow 0$), Rao (1963) ($\check{Z}^n \to p$) and (4) ($k^n/n \to p$) yield (see also Theorem 11.4.5 in Whitt 2002, p. 379) $\check{Z}^n + \check{\Delta}^n - k^n/n \Rightarrow 0$, as $n \to \infty$. Now, the same argument as in Theorem 1 (based on properties of $\varphi_{\bar{P}_1}$ stated in Proposition 1) results in, as $n \to \infty$,

$$\check{X}^n \Rightarrow \varphi_{\bar{P}_1}(0) + p = p,$$

where $\varphi_{\bar{P}_1}(0) = 0$ is due to the fact that 0 is a unique solution to $y(t) = \int_0^t y^+(t-s)\,\mathrm{d}\bar{P}_1(s)$ (see the proof of Proposition 1). □

The following lemma characterizes fluid limits of $\check{A}^n_j := \{\check{A}^n_j(t) = \#\{i\colon \check{\tau}_{i,j} \le t\}, t \ge 0\}$ and $\tilde{A}^n_j := \{\tilde{A}^n_j(t) = \#\{i\colon \tilde{\tau}_{i,j} \le t\}, t \ge 0\}$, for $j \ge 1$. Recall that $H = F * G$ and, consequently,

$$\ddot{H} = p\ddot{G} * F + \bar{p}\ddot{F} = p\ddot{G} + \bar{p}\ddot{F} * G.$$

Lemma 9. *For $j \ge 1$, we have $\check{A}^n_j/n \Rightarrow \ddot{H} * H^{(j-1)}$ and $\tilde{A}^n_j/n \Rightarrow \ddot{H} * H^{(j-1)}$, as $n \to \infty$.*

Proof. First, note that $0 \le \check{A}^n_j(t) - \tilde{A}^n_j(t) \le (X^n(t) - k^n)^+$, for $t \ge 0$. Thus, Corollary 1 implies, as $n \to \infty$,

$$\check{A}^n_j/n - \tilde{A}^n_j/n \Rightarrow 0. \tag{71}$$

The rest of the proof is based on induction. A base is due to the construction of our model (see (4) and (6)): $\check{A}^n_1/n \Rightarrow \ddot{H}$, as $n \to \infty$. Consequently, (71) implies $\tilde{A}^n_1/n \Rightarrow \ddot{H}$, as $n \to \infty$. Now, for $\epsilon > 0$, let $a^{\uparrow}_j(t) = (\ddot{H} * H^{(j-1)}(t) + \epsilon) \wedge 1$ and $a^{\downarrow}_j(t) = (\ddot{H} * H^{(j-1)}(t) - \epsilon)^+$. Then our inductive assumption yields, as $n \to \infty$,

$$\mathbb{P}[\mathscr{E}] \to 1, \tag{72}$$

where $\mathscr{E} := \{a^{\downarrow}_j(t) \le \tilde{A}^n_j(t)/n \le a^{\uparrow}_j(t), \forall t \in [0,T]\}$. Next, note that

$$\check{A}^n_{j+1}(t) = \sum_{i=1}^{\tilde{A}^n_j(t)} 1_{\{\tilde{\eta}_{i,j} + \check{c}_{i,j+1} \le t\}} = \sum_{i=1}^{n} 1_{\{\tilde{\eta}_{i,j} + \tilde{c}_{i,j+1} \le t\}},$$

where $\tilde\eta_{i,j} := \inf\{t \ge 0\colon \tilde A^n_j(t) \ge i\}$ and $\tilde c_{i,j+1}$ is the duration of the next cycle of a customer that starts its $(j+1)$st cycle at time $\tilde\eta_{i,j}$. The preceding equality and the definition of $\mathscr{E}$ result in

$$\mathscr{E} \subseteq \left\{ \sum_{i=1}^{n} 1_{\{(a_j^{\downarrow})^{\leftarrow}(i/n)+\tilde c_{i,j+1}\le t\}} \le \breve A^n_{j+1}(t) \le \sum_{i=1}^{n} 1_{\{(a_j^{\uparrow})^{\leftarrow}(i/n)+\tilde c_{i,j+1}\le t\}},\ \forall\, t\in[0,T]\right\}; \tag{73}$$

recall that $f^{\leftarrow}(x) = \inf\{s\colon f(s)\ge x\}$ is an inverse of a nondecreasing f.

A law of large numbers (e.g., see Daffer and Taylor 1979) and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}[(a_j^{\uparrow})^{\leftarrow}(i/n)+\tilde c_{i,j+1}\le t] &= \frac{1}{n}\sum_{i=1}^{n} H(t-(a_j^{\uparrow})^{\leftarrow}(i/n)) = \frac{1}{n}\int_0^t H(t-s)\,\mathrm{d}\sum_{i=1}^{n} 1_{\{na_j^{\uparrow}(s)\ge i\}}\\ &= \int_0^t H(t-s)\,\mathrm{d}\frac{\lfloor na_j^{\uparrow}(s)\rfloor}{n} \to \int_0^t H(t-s)\,\mathrm{d}a_j^{\uparrow}(s),\end{aligned}$$

as $n\to\infty$, yield

$$\left\{\frac{1}{n}\sum_{i=1}^{n} 1_{\{(a_j^{\uparrow})^{\leftarrow}(i/n)+\tilde c_{i,j+1}\le t\}},\ t\ge 0\right\} \Rightarrow \left\{\int_0^t H(t-s)\,\mathrm{d}a_j^{\uparrow}(s),\ t\ge 0\right\}, \tag{74}$$

as $n\to\infty$. Using the same approach results in, as $n\to\infty$,

$$\left\{\frac{1}{n}\sum_{i=1}^{n} 1_{\{(a_j^{\downarrow})^{\leftarrow}(i/n)+\tilde c_{i,j+1}\le t\}},\ t\ge 0\right\} \Rightarrow \left\{\int_0^t H(t-s)\,\mathrm{d}a_j^{\downarrow}(s),\ t\ge 0\right\}. \tag{75}$$

Finally, combining (72)–(75) with

$$\ddot H * H^{(j)}(t) - \epsilon \le \int_0^t H(t-s)\,\mathrm{d}a_j^{\downarrow}(s) \le \int_0^t H(t-s)\,\mathrm{d}a_j^{\uparrow}(s) \le \ddot H * H^{(j)}(t) + \epsilon$$

concludes the proof of the lemma. □

Finally, we are ready to present the proof of Lemma 2.

Proof of Lemma 2. The proof of the lemma is very similar to the proof of Lemma 5.3 in Krichagina and Puhalskii (1997) (see also the proof of Proposition 5.1 in Reed 2009). The basic idea is to express $\breve M^n_{l,j}(L)$ and $\tilde M^n_{l,j}(L)$ in terms of empirical processes that converge to a Kiefer process (the proof also utilizes tightness of the relevant processes—see Lemma 12 and Remark 7 in Appendix B.2; it is based on a martingale decomposition—see Pang et al. 2007 for a survey). The main difference is that the fluid limits of $\breve A^n_j$ and $\tilde A^n_j$ are not identity functions as in Krichagina and Puhalskii (1997), Reed (2009), but $\ddot H * H^{(j-1)}$ (see Lemma 9 in Section 5.3). A diagram that outlines relations between various objects used in the proof is shown in Figure 2.

First, we consider machines in the increasing order of their jth breakdown times (the jth breakdown time of machine i occurs at $\breve\tau_{i,j}$). That is, we introduce a new labeling of machines based on $\breve A^n_j$. (Due to the FCFS policy, the labeling based on $\tilde A^n_j$ is the same.) To this end, we introduce some notation. Let $s^m_{i,m+j}$ and $a^m_{i,m+j}$ be the service and working times in the $(m+j)$th cycle of a machine that was the ith (among all the machines) to start its mth cycle (also set $c^m_{i,m+j} := s^m_{i,m+j} + a^m_{i,m+j}$); for this customer let $\breve\eta_{i,j} := \inf\{t\ge 0\colon \breve A^n_j(t)\ge i\}$ be the time of its jth breakdown. Now, for $L\in\{G,H\}$, let $\breve V^n_{l,j}(L) := \{\breve V^n_{l,j}(L;t,x),\ t\ge 0,\ x\ge 0\}$ and $U^n_{l,j}(L) := \{U^n_{l,j}(L;t,x),\ t\ge 0,\ x\in[0,1]\}$ be defined by

$$\breve V^n_{l,j}(L;t,x) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\breve A^n_j(t)}\left(1_{\{c^j_{i,j+1:j+l}+s^j_{i,j+l+1}+a^j_{i,j+l+1}1_{\{L=H\}}\le x\}} - L * H^{(l)}(x)\right)$$

Figure 2. A diagram illustrating relations between quantities used in the proof of Lemma 2.

$$\begin{array}{ccccc}\breve M^n_{l,j}(L) & & M^n_{l,j}(L) & & M_{l,j}(L)\\ \updownarrow (90) & & \updownarrow (91) & & \updownarrow (85)\\ \breve M^n_{l,j,m}(L;t) & \overset{(86)}{\longleftrightarrow} & M^n_{l,j,m}(L;t) & \overset{(92)}{\longleftrightarrow} & M_{l,j,m}(L;t)\end{array}$$

and

$$U^n_{l,j}(L;t,x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \left(1_{\{L*H^{(l)}(c^j_{i,j+1:j+l}+s^j_{i,j+l+1}+a^j_{i,j+l+1}1_{\{L=H\}}) \le x\}} - x\right), \tag{76}$$

respectively; here $\{c^j_{i,j+1:j+l} + s^j_{i,j+l+1} + a^j_{i,j+l+1}1_{\{L=H\}},\ i > n\}$ is a sequence of i.i.d. random variables distributed according to $L * H^{(l)}$ and independent of all other random variables in model. Then, $\breve{V}^n_{l,j}(L)$ and $U^n_{l,j}(L)$ relate via

$$\breve{V}^n_{l,j}(L;t,x) = U^n_{l,j}(L;\breve{A}^n_j(t)/n,\ L * H^{(l)}(x)). \tag{77}$$

Moreover, the processes of interest, $\breve{M}^n_{l,j}(L)$, can be written as (see (26))

$$\breve{M}^n_{l,j}(L) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\breve{A}^n_j(t)} \left(1_{\{c^j_{i,j+1:j+l}+s^j_{i,j+l+1}+a^j_{i,j+l+1}1_{\{L=H\}} \le t-\breve{\eta}_{i,j}\}} - L * H^{(l)}(t-\breve{\eta}_{i,j})\right) \tag{78}$$

and it admits the following representation (see Krichagina and Puhalskii 1997 for the definition of the integral):

$$\breve{M}^n_{l,j}(L;t) = \int_0^t \int_0^t 1_{\{s+x\le t\}}\, \mathrm{d}\breve{V}^n_{l,j}(L;s,x).$$

Showing directly that processes $\breve{M}^n_{l,j}(L)$ and $M_{l,j}(L)$ are "close" (for large n) is difficult. Instead, we introduce random variables that approximate $\breve{M}^n_{l,j}(L;t)$ and $M_{l,j}(L;t)$. To this end, consider a sequence $\{r^m_q\}$ such that $0 = r^m_0 < r^m_1 < r^m_2 < \cdots < r^m_m = t$ and

$$\max_{1\le q\le m} |r^m_q - r^m_{q-1}| \to 0, \tag{79}$$

as $m \to \infty$. In turn, this sequence is used to define $\breve{M}^n_{l,j,m}(L;t)$, $M^n_{l,j,m}(L;t)$ and $M_{l,j,m}(L;t)$ as follows:

$$\breve{M}^n_{l,j,m}(L;t) := \sum_{q=1}^{m} \left(\breve{V}^n_{l,j}(L;r^m_q, t-r^m_q) - \breve{V}^n_{l,j}(L;r^m_{q-1}, t-r^m_q)\right), \tag{80}$$

$$M^n_{l,j,m}(L;t) := \sum_{q=1}^{m} \left(U^n_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_q), L*H^{(l)}(t-r^m_q)) - U^n_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_{q-1}), L*H^{(l)}(t-r^m_q))\right) \tag{81}$$

and

$$M_{l,j,m}(L;t) := \sum_{q=1}^{m} \left(U_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_q), L*H^{(l)}(t-r^m_q)) - U_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_{q-1}), L*H^{(l)}(t-r^m_q))\right), \tag{82}$$

where $U_{l,j}(L) := \{U_{l,j}(L;t,x), t \ge 0, x \in [0,1]\}$ is a Kiefer process (e.g., see Kiefer 1972 or Csörgő and Révész 1981, Section 4.2), a two-parameter continuous centered Gaussian process on $[0,\infty)\times[0,1]$ with the covariance function

$$\mathbb{E}U_{l,j}(L;s,x)U_{l,j}(L;t,y) = (s\wedge t)(x \wedge y - xy);$$

$U_{l,j}(L;\cdot,x)$ is a Brownian motion for fixed x, and $U_{l,j}(L;t,\cdot)$ is a Brownian bridge for fixed t. The last process we introduce is $M^n_{l,j}(L) := \{M^n_{l,j}(L;t), t \ge 0\}$:

$$M^n_{l,j}(L;t) := \int_0^t \int_0^t 1_{\{s+x\le t\}}\, \mathrm{d}V^n_{l,j}(L;s,x), \tag{83}$$

where $V^n_{l,j}(L;t,x) := U^n_{l,j}(L;\ddot{H} * H^{(j-1)}(t), L*H^{(l)}(x))$. Here, we note that the centered Gaussian process of interest (with a.s. continuous paths) (Krichagina and Puhalskii 1997, Lemma 2.1), $M_{l,j}(L)$, can be expressed in terms of a Kiefer process as well (Krichagina and Puhalskii 1997):

$$M_{l,j}(L;t) = \int_0^t \int_0^t 1_{\{s+x\le t\}}\, \mathrm{d}V_{l,j}(L;s,x) := \underset{m\to\infty}{\text{l.i.m.}}\ M_{l,j,m}(L;t), \tag{84}$$

where $V_{l,j}(L;t,x) := U_{l,j}(L;\ddot{H} * H^{(j-1)}(t), L*H^{(l)}(x))$ and l.i.m. stands for the mean-square limit. Then, from (82) and (84), it follows that, as $m \to \infty$,

$$M_{l,j,m}(L;t) \xrightarrow{\mathbb{P}} M_{l,j}(L;t). \tag{85}$$

In addition, (80), (81), Lemma 9, Lemma 3.1 in Krichagina and Puhalskii (1997) ($U^n_{l,j}(L) \Rightarrow U_{l,j}(L)$) as well as the continuity of $U_{l,j}(L)$ and $\ddot{H} * H^{(j-1)}$ yield, for $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[|\breve{M}^n_{l,j,m}(L;t) - M^n_{l,j,m}(L;t)| > \epsilon] = 0. \tag{86}$$

Second, we argue that $\breve{M}^n_{l,j,m}(L;t)$ converges to $\breve{M}^n_{l,j}(L;t)$ (in a particular sense, as $n \to \infty$ first, and $m \to \infty$ then). To this end, define $\breve{R}^n_{l,j,m,t}(L) := \{\breve{R}^n_{l,j,m,t}(L;s), s \ge 0\}$ by

$$\breve{R}^n_{l,j,m,t}(L;s) := \sum_{i=1}^{\breve{A}^n_j(s)} \sum_{q=1}^{m} 1_{\{r^m_{q-1} < \breve{\eta}_{i,j} \le r^m_q\}} \big(1_{\{t - r^m_q < c^j_{i,j+1:j+l} + s^j_{i,j+l+1} + a^j_{i,j+l+1} 1_{\{L=H\}} \le t - \breve{\eta}_{i,j}\}} - L * H^{(l)}(t - \breve{\eta}_{i,j}) + L * H^{(l)}(t - r^m_q)\big).$$

Then, (76)–(78) and (80) imply

$$\breve{R}^n_{l,j,m,t}(L;t) = \sqrt{n}(\breve{M}^n_{l,j}(L;t) - \breve{M}^n_{l,j,m}(L;t)). \tag{87}$$

Two additional objects that relate to $\breve{R}^n_{l,j,m,t}(L)$ are of interest: the σ-algebra

$$\breve{\mathscr{F}}^n_{s,l,j}(L) = \sigma\big\{\breve{\eta}_{i,j} \wedge \breve{\eta}_{\breve{A}^n_j(s)+1,j},\ c^j_{i\wedge \breve{A}^n_j(s),j+1:j+l} + s^j_{i\wedge \breve{A}^n_j(s),j+l+1} + a^j_{i\wedge \breve{A}^n_j(s),j+l+1} 1_{\{L=H\}},\ 1 \le i \le n\big\} \vee \mathcal{N}$$

and the process $\langle \breve{R}^n_{l,j,m,t}(L)\rangle := \{\langle \breve{R}^n_{l,j,m,t}(L)\rangle(s), s \ge 0\}$ defined by

$$\begin{aligned}
\langle \breve{R}^n_{l,j,m,t}(L)\rangle(s) &:= \sum_{i=1}^{\breve{A}^n_j(s)} \sum_{q=1}^{m} 1_{\{r^m_{q-1} < \breve{\eta}_{i,j} \le r^m_q\}} (L * H^{(l)}(t - \breve{\eta}_{i,j}) - L * H^{(l)}(t - r^m_q))(1 - L * H^{(l)}(t - \breve{\eta}_{i,j}) + L * H^{(l)}(t - r^m_q)) \\
&\le \sum_{i=1}^{\breve{A}^n_j(s)} \sum_{q=1}^{m} 1_{\{r^m_{q-1} < \breve{\eta}_{i,j} \le r^m_q\}} (L * H^{(l)}(t - r^m_{q-1}) - L * H^{(l)}(t - r^m_q)) \\
&\le \sum_{q=1}^{m} (L * H^{(l)}(t - r^m_{q-1}) - L * H^{(l)}(t - r^m_q))(\breve{A}^n_j(r^m_q) - \breve{A}^n_j(r^m_{q-1})) \\
&\le L * H^{(l)}(t) \sup_{1\le q\le m} \{\breve{A}^n_j(r^m_q) - \breve{A}^n_j(r^m_{q-1})\}.
\end{aligned} \tag{88}$$

Then, the same argument as in the proof of the third part of Lemma 5.2 in Krichagina and Puhalskii (1997) yields that $\breve{R}^n_{l,j,m,t}$ is an $\{\breve{\mathscr{F}}^n_{s,l,j}(L), s \ge 0\}$-square-integrable martingale with the process $\langle \breve{R}^n_{l,j,m,t}(L)\rangle$ as predictable quadratic-variation process. This fact, together with (87), (88) and the Lenglart-Rebolledo inequality (e.g., see Liptser and Shiryaev 1989, p. 66 or Whitt 2007), yields, for $\delta, \epsilon > 0$,

$$\begin{aligned}
\mathbb{P}[|\breve{M}^n_{l,j}(L;t) - \breve{M}^n_{l,j,m}(L;t)| > \delta] &= \mathbb{P}[|\breve{R}^n_{l,j,m,t}(L;t)|/\sqrt{n} > \delta] \\
&\le \epsilon/\delta^2 + \mathbb{P}\Big[\sup_{1\le q\le m} \{\breve{A}^n_j(r^m_q)/n - \breve{A}^n_j(r^m_{q-1})/n\} > \epsilon\Big].
\end{aligned} \tag{89}$$

Now, Lemma 9, the continuity of $\ddot{H} * H^{(l)}$ and (79) result in, for $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big[\sup_{1\le q\le m} \{\breve{A}^n_j(r^m_q)/n - \breve{A}^n_j(r^m_{q-1})/n\} > \epsilon\Big] = 0.$$

The preceding double limit and (89) imply, for $\delta > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\breve{M}^n_{l,j}(L;t) - \breve{M}^n_{l,j,m}(L;t)| > \delta] = 0. \tag{90}$$

Third, a somewhat simpler argument can be used to compare $M^n_{l,j}(L;t)$ and $M^n_{l,j,m}(L;t)$. In particular, (81) and (83) render

$$\begin{aligned}
&\sqrt{n}(M^n_{l,j}(L;t) - M^n_{l,j,m}(L;t)) \\
&\quad = \sum_{i=1}^{\lfloor n\ddot{H}*H^{(j-1)}(t)\rfloor} \sum_{q=1}^{m} 1_{\{r^m_{q-1} < \eta_{i,j} \le r^m_q\}} \big(1_{\{t - r^m_q < c^j_{i,j+1:j+l} + s^j_{i,j+l+1} + a^j_{i,j+l+1} 1_{\{L=H\}} \le t - \eta_{i,j}\}} - L * H^{(l)}(t - \eta_{i,j}) + L * H^{(l)}(t - r^m_q)\big),
\end{aligned}$$

where $\eta_{i,j} = (\ddot{H} * H^{(j-1)})^{\leftarrow}(i/n)$. Due to the independence of the summands in the preceding expression, one has

$$n\,\mathrm{Var}(M^n_{l,j}(L;t) - M^n_{l,j,m}(L;t)) \le \sup_{1\le q\le m}\left\{\lfloor n\ddot{H} * H^{(j-1)}(r^m_q)\rfloor - \lfloor n\ddot{H} * H^{(j-1)}(r^m_{q-1})\rfloor\right\},$$

and therefore, due to the continuity of $\ddot{H} * H^{(j-1)}$ and (79), one has, for $\delta > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}[|M^n_{l,j}(L;t) - M^n_{l,j,m}(L;t)| > \delta] = 0. \tag{91}$$

Fourth, consider a set $\{t_i, i = 1,\dots,h\} \subseteq \{r^m_i, i = 1,\dots,m\}$. Since the only condition of $\{r^m_i\}$ is (79), the set $\{t_i\}$ can be arbitrary. Then, (81), (82), Krichagina and Puhalskii (1997, Lemma 3.1) ($U^n_{l,j}(L) \Rightarrow U_{l,j}(L)$) and the continuity of $U_{l,j}(L)$ yield, as $n\to\infty$,

$$(M^n_{l,j,m}(L;t_1),\dots,M^n_{l,j,m}(L;t_h)) \Rightarrow (M_{l,j,m}(L;t_1),\dots,M_{l,j,m}(L;t_h)), \tag{92}$$

where $r^m_{m_i} = t_i$,

$$M^n_{l,j,m}(L;t_i) := \sum_{q=1}^{m_i}\left(U^n_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_q), L * H^{(l)}(t_i - r^m_q)) - U^n_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_{q-1}), L * H^{(l)}(t_i - r^m_q))\right)$$

and

$$M_{l,j,m}(L;t_i) := \sum_{q=1}^{m_i}\left(U_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_q), L * H^{(l)}(t_i - r^m_q)) - U_{l,j}(L;\ddot{H} * H^{(j-1)}(r^m_{q-1}), L * H^{(l)}(t_i - r^m_q))\right).$$

Fifth, (85), (86), (90)–(92) and Billingsley (1999, Theorem 3.2) imply, as $n\to\infty$,

$$(\breve{M}^n_{l,j}(L), M^n_{l,j}(L)) \overset{\text{f.d.}}{\Rightarrow} (M_{l,j}(L), M_{l,j}(L)). \tag{93}$$

Furthermore, repeating the above argument, one can show that, as $n\to\infty$,

$$(\tilde{M}^n_{l,j}(L), M^n_{l,j}(L)) \overset{\text{f.d.}}{\Rightarrow} (M_{l,j}(L), M_{l,j}(L)). \tag{94}$$

Combining (93) and (94) renders, as $n\to\infty$,

$$(\breve{M}^n_{l,j}(L), \tilde{M}^n_{l,j}(L), M^n_{l,j}(L)) \overset{\text{f.d.}}{\Rightarrow} (M_{l,j}(L), M_{l,j}(L), M_{l,j}(L)).$$

Finally, the preceding limit, Lemma 12 in the appendix (see also Remark 7), Whitt (2007, Lemma 3.2) (it assumes the axiom of choice) and Billingsley (1999, p. 139) result in, as $n\to\infty$,

$$(\breve{M}^n_{l,j}(L), \tilde{M}^n_{l,j}(L), M^n_{l,j}(L)) \Rightarrow (M_{l,j}(L), M_{l,j}(L), M_{l,j}(L)).$$

Sixth, we argue that $M_{l,j}(L)$ has a.s. continuous paths. To begin, we show that $M_{l,j}(L)$ is stochastically continuous (i.e., continuity in the mean square sense). The definitions (82) and (84) yield, for $s < t$ (the same partition is used),

$$\begin{aligned}
&\mathbb{E}(M_{l,j}(L;t) - M_{l,j}(L;s))^2\\
&\quad = \lim_{m\to\infty}\mathbb{E}(M_{l,j,m}(L;t) - M_{l,j,m}(L;s))^2\\
&\quad = \lim_{m\to\infty}\sum_{q=1}^{m}\left((\ddot{H} * H^{(j-1)}(r^m_q) - \ddot{H} * H^{(j-1)}(r^m_{q-1}))(L * H^{(l)}(t - r^m_q) - L * H^{(l)}(s - r^m_q))(1 + L * H^{(l)}(s - r^m_q) - L * H^{(l)}(t - r^m_q))\right)\\
&\quad = \int_0^t (L * H^{(l)}(t-u) - L * H^{(l)}(s-u))(1 + L * H^{(l)}(s-u) - L * H^{(l)}(t-u))\,\mathrm{d}\ddot{H} * H^{(j-1)}(u),
\end{aligned} \tag{95}$$

where the last equality follows from the continuity of $\ddot{H} * H^{(j-1)}$, the fact that $L * H^{(l)}$ has at most a countable number of jumps and the fact that $\sup_{q\le m}|r^m_q - r^m_{q-1}| \to 0$, as $m\to\infty$. Now, by Lemma 4.9.6 in Liptser and Shiryaev

(1989, p. 279), to conclude that $M_{l,j}(L)$ has a.s. continuous path, it is enough to show that the following holds, for any partition $\{r_q^m\}$ of $[0,t]$ such that $\sup_{q\le m}|r_q^m - r_{q-1}^m|\to 0$, as $m\to\infty$,

$$\lim_{\kappa\to\infty}\limsup_{m\to\infty}\mathbb{P}\left[\sum_{i=1}^m (M_{l,j}(L;r_q^m)-M_{l,j}(L;r_{q-1}^m))^2>\kappa\right]=0.$$

The preceding double limit follows from Markov's inequality and the following:

$$\sum_{i=1}^m \mathbb{E}(M_{l,j}(L;r_q^m)-M_{l,j}(L;r_{q-1}^m))^2 \le \sum_{i=1}^m \int_0^t (L*H^{(l)}(r_q^m-u)-L*H^{(l)}(r_{q-1}^m-u))\,\mathrm{d}\ddot{H}*H^{(j-1)}(u)\le 1,$$

where (95) is used in the first inequality.

This concludes the proof. □

Acknowledgments

The first author thanks Mor Armony and Avishai Mandelbaum for helpful discussions.

Appendix A. (Additional) Motivation for QED

Just in this appendix, we restrict our attention to a model with exponential repair and working times, but allow p to vary with n. Our focus is on the limiting stationary distribution, and, thus, it is immaterial whether p varies due to changes in F, G or both.

Let $X^n(\infty)$ be a random variable with the distribution equal to the stationary distribution of $\{X^n(t), t\ge 0\}$. The next result states that the orders of magnitude of stochastic fluctuations of the stationary number of idle repairman $((X^n(\infty)-k^n)^-)$ and waiting machines $((X^n(\infty)-k^n)^+)$ are different—$\sqrt{np\bar{p}}$ and $\sqrt{np\bar{p}}/(\sqrt{p}\vee\beta^n)$, respectively. More importantly, comparing the proposition below with Theorem 2 in de Véricourt and Jennings (2008), indicates that the stationary QED approximation for finite p remains valid even in the case when p varies with n.

Proposition 2. *Consider a sequence of memoryless machine repair models in the QED regime. Let σ be the solution to $p(n-\sigma)=\bar{p}(\sigma\wedge k^n)$, and $\beta^n/\sqrt{p}\to\psi$, as $n\to\infty$. If $np\bar{p}\to\infty$, then*

$$\lim_{n\to\infty}\mathbb{P}[X^n(\infty)\le k^n]=\left(1+\frac{\psi e^{\psi^2/2}\Phi(-\psi)}{\beta e^{\beta^2/2}\Phi(\beta)}\right)^{-1}; \tag{A.1}$$

moreover, for $x\le\beta^+$,

$$\lim_{n\to\infty}\mathbb{P}\left[\frac{X^n(\infty)-\sigma\wedge k^n}{\sqrt{np\bar{p}}}<x\,\middle|\,X^n(\infty)\le k^n\right]=\frac{\Phi(x-\beta^-)}{\Phi(\beta)} \tag{A.2}$$

and, for $x\ge -(1\vee\psi)\psi^-$,

$$\lim_{n\to\infty}\mathbb{P}\left[\frac{X^n(\infty)-\sigma\vee k^n}{\sqrt{np\bar{p}}}(\sqrt{p}\vee\beta^n)>x\,\middle|\,X^n(\infty)\ge k^n\right]=\frac{\bar{\Phi}(x/(1\vee\psi)+\psi^+)}{\bar{\Phi}(\psi)}. \tag{A.3}$$

Remark 5. There exist several cases of interest:
- If $\beta=\psi=0$, then the ratio ψ/β in (A.1) should be interpreted as the limit of $1/\sqrt{p}$.
- If $p\to 0$ and $\beta>0$, then $\psi\to\infty$ and (A.1) reduces to

$$\lim_{n\to\infty}\mathbb{P}[X^n(\infty)\le k]=\frac{\sqrt{2\pi}\beta e^{\beta^2/2}\Phi(\beta)}{1+\sqrt{2\pi}\beta e^{\beta^2/2}\Phi(\beta)}.$$

- If $p\to 0$ and $\beta<0$, then $\psi\to-\infty$ and (A.1) reduces to $\mathbb{P}[X^n(\infty)\le k^n]\to 0$, as $n\to\infty$.
- If $\psi\to\infty$, then the right-hand side of (A.3) reduces to e^{-x}, $x\ge 0$.

Proof. Let $\psi^n:=\beta^n/\sqrt{p}$, $\pi_i:=\mathbb{P}[X^n(\infty)=i]$ and note that (e.g., see Kleinrock 1975, Section 3.10)

$$\pi_i=\begin{cases}\dfrac{\binom{n}{i}(p/\bar{p})^{i-k^n}}{\binom{n}{k^n}}\pi_{k^n}, & i\le k^n,\\[2ex] \dfrac{\binom{n}{i}i!(p/\bar{p})^{i-k^n}}{(k^n)^{i-k^n}k^n!\binom{n}{k^n}}\pi_{k^n}, & i\ge k^n.\end{cases} \tag{A.4}$$

Next, we consider $\pi_{\lfloor k^n+x\sqrt{np\bar{p}}\rfloor}$ for $x<0$. Relation (A.4), Stirling's approximation, (4) and some tedious algebra result in, for $x<0$,

$$\pi_{\lfloor k^n+x\sqrt{np\bar{p}}\rfloor}=e^{-x^2/2-\beta^n x}\pi_{k^n}(1+o(1)), \tag{A.5}$$

as $n\to\infty$. This estimate, together with $\sigma\wedge k^n = k^n-(\beta^n)^+\sqrt{np\bar{p}}$ and the assumption $np\bar{p}\to\infty$, yields

$$\mathbb{P}\left[\frac{X^n(\infty)-\sigma\wedge k^n}{\sqrt{np\bar{p}}}<x\,\middle|\,X^n(\infty)\le k^n\right]\to\frac{\int_{-\infty}^{x-\beta^+}e^{-y^2/2-\beta y}\,dy}{\int_{-\infty}^{0}e^{-y^2/2-\beta y}\,dy},$$

as $n\to\infty$, for $x\le\beta^+$, and (A.2) follows. Similarly, (A.4), Stirling's approximation, (4), the Taylor expansion of $\log(1+x)$ and $\sigma\vee k^n=k^n+(\psi^n)^-\sqrt{n\bar{p}}$ imply

$$\pi_{\lfloor\sigma\vee k^n+x\sqrt{n\bar{p}}/(1\vee\psi^n)\rfloor}=e^{-(\psi^n)^+x/(1\vee\psi^n)-x^2/(2(1\vee\psi^n)^2)}\pi_{\lfloor\sigma\vee k^n\rfloor}(1+o(1)),\tag{A.6}$$

as $n\to\infty$, for $x\ge-(1\vee\psi^n)(\psi^n)^-$. Consequently, passing $n\to\infty$ result in (A.3). Finally, (A.1) can be obtained as well. To this end, (A.6) renders

$$\pi_{k^n}=e^{-((\psi^n)^-)^2/2}\pi_{\lfloor\sigma\vee k^n\rfloor}(1+o(1)),$$

as $n\to\infty$, which, combined with (A.5), (A.6) and

$$\mathbb{P}[X^n(\infty)\le k^n]=\left(1+\frac{\sum_{i=k^n+1}^{n}\pi_i}{\sum_{i=0}^{k^n}\pi_i}\right)^{-1},$$

implies

$$\mathbb{P}[X^n(\infty)\le k^n]\to\left(1+\frac{\psi e^{(\psi^-)^2/2}\int_{-\psi^-}^{\infty}e^{-x^2/2-\psi^+x}\,dx}{\beta\int_{-\infty}^{0}e^{-x^2/2-\beta x}\,dx}\right)^{-1},$$

as $n\to\infty$. This concludes the proof. □

Appendix B. Ancillary Results

B.1. Decomposition

In this section, we state a decomposition for $\check{M}^n_{l,j}(L)$ (Lemma 10; there is an analogous decomposition for $\tilde{M}^n_{l,j}(L)$—we omit details), and provide some properties of elements of this decomposition (Lemma 11). (The proofs of Lemmas 3 and 4 can be found at the end of this section.) In order to state it, we introduce some additional notation. To this end, let

$$\check{\gamma}_{i,l,j}(L):=\check{\eta}_{i,j}+c^j_{i,j+1:j+l}+s^j_{i,j+l+1}1_{\{L=H\}}\tag{B.1}$$

and

$$v_{i,l,j}(L):=1_{\{L=G\}}s^j_{i,j+l+1}+1_{\{L=H\}}a^j_{i,j+l+1}.\tag{B.2}$$

Random variables $c^j_{i,j+1:j+l}$, $s^j_{i,j+l+1}$ and $\check{\eta}_{i,j}$ were introduced in the proof of Lemma 2. Recall that $s^m_{i,m+j}$ and $a^m_{i,m+j}$ are the service and working times in the $(m+j)$th cycle of a machine that was the ith (among all the machines) to start its mth cycle ($c^m_{i,m+j}:=s^m_{i,m+j}+a^m_{i,m+j}$); for this customer let $\check{\eta}_{i,j}:=\inf\{t\ge0\colon \check{A}^n_j(t)\ge i\}$ be the time of its jth breakdown (the fluid limit of $\check{A}^n_j$ is obtained in Lemma 9). Note that $\check{M}^n_{l,j}(L)$ (see (78)) can be written as follows:

$$\check{M}^n_{l,j}(L;t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1_{\{v_{i,l,j}(L)\le t-\check{\gamma}_{i,l,j}(L)\}}-L*H^{(l)}(t-\check{\eta}_{i,j})).$$

Furthermore, let $\check{E}^n_{l,j}(L):=\{\check{E}^n_{l,j}(L;t),\,t\ge0\}$,

$$\check{E}^n_{l,j}(L;t):=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\check{E}^n_{i,l,j}(L;t),$$

with $\check{E}^n_{i,l,j}(L):=\{\check{E}^n_{i,l,j}(L;t),\,t\ge0\}$ and

$$\check{E}^n_{i,l,j}(L;t):=1_{\{v_{i,l,j}(L)\le t-\check{\gamma}_{i,l,j}(L)\}}-\int_0^{v_{i,l,j}(L)\wedge(t-\check{\gamma}_{i,l,j}(L))^+}\frac{dG_L(s)}{\bar{G}_L(s-)},\tag{B.3}$$

where

$$G_L:=1_{\{L=G\}}G+1_{\{L=H\}}F$$

is the distribution function of $v_{i,l,j}(L)$. The integral in (B.3) is the standard compensator for the "one-point" point process (the indicator in (B.3)) (Jacod and Shiryaev 2003, p. 98); it is well defined for all distribution functions G_L, since $\bar{G}_L(s-)$ is zero only on the null set of G_L. We also define a counting process $\check{J}^n_{l,j}(L):=\{\check{J}^n_{l,j}(L;t),\,t\ge0\}$ with

$$\check{J}^n_{l,j}(L;t):=\sum_{i=1}^{\check{A}^n_j(t)}1_{\{\check{\gamma}_{i,l,j}\le t\}}.\tag{B.4}$$

The last process we define here is $\breve{B}^n_{l,j}(L) := \{\breve{B}^n_{l,j}(L;t,x), t\ge 0, x\in[0,1]\}$:

$$\breve{B}^n_{l,j}(L;t,x) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}(1_{\{G_L^{\leftarrow}(\breve{v}_{i,l,j}(L))\le x\}} - x),$$

where, for $i\le n$, $\breve{v}_{i,l,j}(L) = v_{i^\star,l,j}(L)$, $i\mapsto i^\star$ is a bijection and $\breve{\gamma}_{i^\star,l,j}(L) = \inf\{t\ge 0\colon \breve{J}^n_{l,j}(L;t)\ge i\}$; for $i>n$, $\{\breve{v}_{i,l,j}(L)\}$ is a sequence of i.i.d. random variables distributed according to G_L, independent of all other random variables in model.

Let $\underline{L}\in\{G,H\}$ be defined by $\{\underline{L}\} = \{G,H\}\setminus\{L\}$. We state the following decomposition without a proof, since it is based on straightforward algebra. (By definition $\breve{M}_{-1,j}(L)\equiv 0$.)

Lemma 10. *We have*

$$\breve{M}^n_{l,j}(L;t) = \breve{E}^n_{l,j}(L;t) - \int_0^t \frac{\breve{B}^n_{l,j}(L;\breve{J}^n_{l,j}(L;t-u)/n, G_L(u-))}{\bar{G}_L(u-)}\,\mathrm{d}G_L(u) + \int_0^t \breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L};t-u)\,\mathrm{d}G_L(u).$$

The above result (in conjunction with Lemma 11) is an extension of the martingale decomposition in Reed (2009). The first two terms of the decompositions are similar to those in Reed (2009), in the sense that their martingale properties can be proved with the same technique, provided that appropriate filtrations can be defined. The approach in Reed (2009) is itself an extension of a martingale method for empirical process (e.g., see Jacod and Shiryaev 2003, p. 99) for the case of dependent one-point point processes (the dependency, in that cases, is due to the finiteness of service capacity of an open system, i.e., waiting times). The main idea employed (originally used in Krichagina and Puhalskii 1997) is to find a σ-algebra for a given time in the filtration, such that terms corresponding to individual customers become conditionally independent (the corresponding martingale terms are orthogonal). In particular, arrivals to the system, service times of future arrivals, and residual service times of customers in service are conditionally independent given this σ-algebra. In the case of a machine repair model, dependencies are more intricate due to the closed nature of the system: breakdown times (arrivals) and repair (service) times are dependent. Thus, we first decompose $X^n - Z^n$ (see Section 4.1), and then find appropriate σ-algebras with desirable properties for individual pieces ($\breve{M}^n_{l,j}(L)$ and $\tilde{M}^n_{l,j}(L)$) in the decomposition of $X^n - Z^n$. It turns out, the σ-algebra that facilitates the analysis of $\breve{M}^n_{l,j}(L)$ (resp. $\tilde{M}^n_{l,j}(L)$) contains the σ-algebra generated by terms in $\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L})$ (resp. $\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L})$)—this is the reason for the last term in Lemma 10.

The next result characterizes $\breve{E}^n_{l,j}(L)$. In order to state it, we define a filtration $\breve{\mathbf{E}}^n_{l,j}(L) := \{\breve{\mathscr{E}}^n_{t,l,j}(L), t\ge 0\}$ associated with the following σ-algebra:

$$\begin{aligned}\breve{\mathscr{E}}^n_{t,l,j}(L) &:= \sigma\{c^j_{i,j+1:j+l} + s^j_{i,j+l+1}1_{\{L=H\}}, i=1,2,\ldots,\breve{A}^n_j(t)\}\\ &\quad\vee\sigma\{1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}, s\le t, i\colon \breve{\gamma}_{i,l,j}(L)\le t\}\vee\sigma\{\breve{A}^n_j(s), s\le t\}\vee\mathcal{N}, \quad t\ge 0;\end{aligned} \tag{B.5}$$

under this definition, $\breve{\mathbf{E}}^n_{l,j}(L)$ satisfies the usual conditions (Jacod and Shiryaev 2003, p. 2).

Remark 6. Note there is more than one way to define $\breve{\gamma}_{i,l,j}(L)$ and $v_{i,l,j}(L)$, such that $\breve{\gamma}_{i,l,j}(L) + v_{i,l,j}(L) = \breve{\eta}_{i,j} + c^j_{i,j+1:j+l} + s^j_{i,j+l+1} + 1_{\{L=H\}}a^j_{i,j+l+1}$ (which is the point that defines one of one-point point processes $\breve{\vartheta}_{m,j,l}(L)$, $1\le m\le n$). There are two reasons for the particular choice of $\breve{\gamma}_{i,l,j}(L)$ and $v_{i,l,j}(L)$ (see (B.1) and (B.2)). First, such a choice leads to orthogonal martingale terms $\breve{E}^n_{i,l,j}(L)$ (see (B.3)) that correspond to different machines (different i). Second, the last term in Lemma 10 can be expressed in a recursive form. In view of Section 4.1, $\breve{\gamma}_{i,l,j}(L)$ is an analogue of an arrival in an open system. The components of $\breve{\mathscr{E}}^n_{t,l,j}(L)$ in (B.5) describe $\breve{\gamma}_{i,l,j}(L)$ and $v_{i,l,j}(L)$. In particular, the first two sigma algebras in (B.5) contain $\sigma\{\breve{\gamma}_{i,l,j}(L)\le t, i=1,2,\ldots,n\}$, while the third sigma algebra contains information on $1_{\{\breve{\gamma}_{i,l,j}(L)+v_{i,l,j}(L)\le t\}}$.

Lemma 11. *The process $\breve{E}^n_{l,j}(L)$ is an $\breve{\mathbf{E}}^n_{l,j}(L)$-square integrable martingale with predictable quadratic-variation process $\langle\breve{E}^n_{l,j}(L)\rangle := \{\langle\breve{E}^n_{l,j}(L)\rangle(t), t\ge 0\}$, where*

$$\langle\breve{E}^n_{l,j}(L)\rangle(t) = \frac{1}{n}\sum_{i=1}^n\int_0^{v_{i,l,j}(L)\wedge(t-\breve{\gamma}_{i,l,j}(L))^+}\frac{\bar{G}_L(u)}{(\bar{G}_L(u-))^2}\,\mathrm{d}G_L(u).$$

Proof. We follow the three steps of the proof of Lemma 3.5 in Krichagina and Puhalskii (1997). In the first step, we argue that the process $\breve{E}^n_{i,l,j}(L)$, $1\le i\le n$, is an $\breve{\mathbf{E}}^n_{l,j}(L)$-square-integrable martingale. The second step involves finding the predictable quadratic-variation process of $\breve{E}^n_{i,l,j}(L)$. The third step of the proof involves showing that the martingales $\breve{E}^n_{i,l,j}(L)$ and $\breve{E}^n_{q,l,j}(L)$ are orthogonal, for $i\ne q$, i.e., that $\breve{E}^n_{i,l,j}(L)\breve{E}^n_{q,l,j}(L)$ is an $\breve{\mathbf{E}}^n_{l,j}(L)$-square-integrable martingale.

Step I. Note that $\breve{E}^n_{i,l,j}(L)$ is $\breve{\mathbf{E}}^n_{l,j}(L)$-adapted and

$$\sup_{t\ge 0}\mathbb{E}(\breve{E}^n_{i,l,j}(L;t))^2 \le \infty.$$

To complete this step of the proof, it is sufficient to show that, for $s<t$,

$$1_{\{\breve{\gamma}_{i,l,j}(L)>s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)] = 0 \tag{B.6}$$

and

$$1_{\{\breve{\gamma}_{i,l,j}(L)\le s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]=\breve{E}^n_{i,l,j}(L;s). \tag{B.7}$$

Observe that $\breve{\gamma}_{i,l,j}(L)$ is an $\breve{\mathbf{E}}^n_{l,j}(L)$-stopping time, and, thus, the σ-field $\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)$ is well defined (Brémaud 1981, p. 297) (see (B.5)). Then, one has

$$1_{\{\breve{\gamma}_{i,l,j}(L)>s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]=1_{\{\breve{\gamma}_{i,l,j}(L)>s\}}\mathbb{E}[\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)]\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]. \tag{B.8}$$

In view of $v_{i,l,j}(L)$ being a positive random variable, $v_{i,l,j}(L)$ and $\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)$ are independent, which implies

$$\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)]=\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\gamma}_{i,l,j}(L)]=0,$$

where the second equality follows from (B.3). Combining the preceding equality with (B.8) results in (B.6).

Now, we consider the right-hand side of (B.7):

$$1_{\{\breve{\gamma}_{i,l,j}(L)\le s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]=1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]+1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)].$$

Since $1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}$ and $1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}$ are both $\breve{\mathscr{E}}^n_{s,l,j}(L)$-measurable, from (B.3) it follows that

$$1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}\breve{E}^n_{i,l,j}(L;t)=1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}-1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}\int_0^{v_{i,l,j}(L)\wedge(t-\breve{\gamma}_{i,l,j}(L))^+}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)},$$

where the last term is $\breve{\mathscr{E}}^n_{s,l,j}(L)$-measurable. Hence, the preceding implies

$$\begin{aligned}&1_{\{\breve{\gamma}_{i,l,j}(L)\le s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\&\quad=1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}-1_{\{v_{i,l,j}(L)\le s-\breve{\gamma}_{i,l,j}(L)\}}\int_0^{v_{i,l,j}(L)}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)}+1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)].\end{aligned} \tag{B.9}$$

Next, we evaluate the last term in (B.9). On the event $\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}$, $v_{i,l,j}(L)$ and $\{\breve{A}^n_j(u),u\le s\}$ are independent. To show this, note that there exists a Borel function f such that $\breve{A}^n_j=f(\xi^n,s_{q,m},a_{q,m},q\ge 1,m\ge 1)$. Now, if we replace $v_{i,l,j}(L)$ with ∞ in the function f, we obtain another process $\dot{A}^n_j$, which is independent from $v_{i,l,j}(L)$. Observe that, on the event $\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}$, $\dot{A}^n_j(u)$ coincides $\breve{A}^n_j(u)$ for $u\le s$. Recall that $v_{i,l,j}(L)$ is independent of $c^j_{q,j+1:j+l}+s_{q,j+l+1}1_{\{L=H\}}$ for $q=1,2,\ldots,n$. Thus, on the event $\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}$, $\breve{E}^n_{i,l,j}(L;t)$ depends on $\breve{\mathscr{E}}^n_{s,l,j}(L)$ only through $\breve{\gamma}_{i,l,j}(L)$ and $v_{i,l,j}(L)$. This, together with (B.5), implies

$$\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}\subset(\sigma(\{\xi^n,s_{q,m},a_{q,m},q=1,2,\ldots,n,m\ge 1\}\setminus\{v_{i,l,j}(L)\})\vee\mathcal{N})\cap\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}.$$

The preceding relation and Krichagina and Puhalskii (1997, Lemma 3.6) result in

$$\begin{aligned}&1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\&\quad=1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}\frac{\mathbb{E}[1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\}}\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\gamma}_{i,l,j}(L)]}{\mathbb{P}[v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\,|\,\breve{\gamma}_{i,l,j}(L)]}=-1_{\{v_{i,l,j}(L)>s-\breve{\gamma}_{i,l,j}(L)\ge 0\}}\int_0^{s-\breve{\gamma}_{i,l,j}(L)}\frac{\mathrm{d}G_L(s)}{\bar{G}_L(s-)}.\end{aligned}$$

Combining the preceding equality with (B.9) yields (B.7). This completes the first step of the proof.

Step II. Given that the second term on the right-hand side of (B.3) is $\breve{\mathbf{E}}^n_{l,j}(L)$-predictable, the $\breve{\mathbf{E}}^n_{l,j}(L)$-predictable measure of jumps of the process $\{1_{\{v_{i,l,j}(L)\le t-\breve{\gamma}_{i,l,j}(L)\}},t\ge 0\}$ is given by Krichagina and Puhalskii (1997, p. 259)

$$\breve{e}^n_{l,j}([0,t],A)=1_{\{1\in A\}}\int_0^t 1_{\{\breve{\gamma}_{i,l,j}(L)<u\le v_{i,l,j}(L)+\breve{\gamma}_{i,l,j}(L)\}}\frac{\mathrm{d}G_L(u-\breve{\gamma}_{i,l,j}(L))}{\bar{G}_L((u-\breve{\gamma}_{i,l,j}(L))-)}.$$

Hence, the predictable quadratic-variation process of $\breve{E}^n_{i,l,j}(L)$ satisfies (Liptser and Shiryaev 1989, pp. 192–193)

$$\begin{aligned}\langle\breve{E}^n_{i,l,j}(L)\rangle(t)&=\int_0^t\int_{\mathbb{R}}x^2\,\breve{e}^n_{l,j}(du,dx)-\sum_{0<u\le t}\left(\int_{\mathbb{R}}x\,\breve{e}^n_{l,j}(u,\mathrm{d}x)\right)^2\\&=\int_0^t 1_{\{\breve{\gamma}_{i,l,j}(L)<u\le v_{i,l,j}(L)+\breve{\gamma}_{i,l,j}(L)\}}\frac{\mathrm{d}G_L(u-\breve{\gamma}_{i,l,j}(L))}{\bar{G}_L((u-\breve{\gamma}_{i,l,j}(L))-)}-\sum_{0<u\le t}1_{\{\breve{\gamma}_{i,l,j}(L)<u\le v_{i,l,j}(L)+\breve{\gamma}_{i,l,j}(L)\}}\left(\frac{\Delta G_L(u-\breve{\gamma}_{i,l,j}(L))}{\bar{G}_L((u-\breve{\gamma}_{i,l,j}(L))-)}\right)^2\\&=\int_0^{v_{i,l,j}(L)\wedge(t-\breve{\gamma}_{i,l,j}(L))^+}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)}-\sum_{0<u\le v_{i,l,j}(L)\wedge(t-\breve{\gamma}_{i,l,j}(L))^+}\left(\frac{\Delta G_L(u)}{\bar{G}_L(u-)}\right)^2=\int_0^{v_{i,l,j}(L)\wedge(t-\breve{\gamma}_{i,l,j}(L))^+}\frac{\bar{G}_L(u)}{(\bar{G}_L(u-))^2}\,\mathrm{d}G_L(u);\end{aligned}$$

here, $\Delta G_L(u)$ denotes the size of a jump of $G_L(\cdot)$ at u, and the sums are over all the jumps. This completes the second step of the proof.

Step III. It is enough to show that, for $s<t$,

$$1_{\{\breve{\gamma}_{i,l,j}(L)\vee\breve{\gamma}_{q,l,j}(L)>s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]=0 \tag{B.10}$$

and

$$1_{\{\breve{\gamma}_{i,l,j}(L)\vee\breve{\gamma}_{q,l,j}(L)\le s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]=\breve{E}^n_{i,l,j}(L;s)\breve{E}^n_{q,l,j}(L,s); \tag{B.11}$$

note that

$$\begin{aligned}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]&=\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\Gamma_{i,q}]\mathbb{P}[\Gamma_{i,q}\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\&\quad+\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\underline{\Gamma}_{i,q}]\mathbb{P}[\underline{\Gamma}_{i,q}\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)],\end{aligned}$$

where $\Gamma_{i,q}:=\{\breve{\gamma}_{i,l,j}(L)\ge\breve{\gamma}_{q,l,j}(L)\}$ and $\underline{\Gamma}_{i,q}:=\{\breve{\gamma}_{i,l,j}(L)<\breve{\gamma}_{q,l,j}(L)\}$. To demonstrate (B.10), we write

$$\begin{aligned}&1_{\{\breve{\gamma}_{i,l,j}(L)\vee\breve{\gamma}_{q,l,j}(L)>s\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\Gamma_{i,q}]\\&\quad=1_{\{\breve{\gamma}_{i,l,j}(L)\vee\breve{\gamma}_{q,l,j}(L)>s\}}\times\mathbb{E}[\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\Gamma_{i,q}],\end{aligned} \tag{B.12}$$

where

$$\begin{aligned}&\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\\&\quad=1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\\&\qquad+1_{\{v_{q,l,j}(L)>\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}].\end{aligned} \tag{B.13}$$

In view of both $1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}$ and

$$1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\breve{E}^n_{q,l,j}(L;t)=1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\breve{E}^n_{q,l,j}(L;t\wedge\breve{\gamma}_{i,l,j}(L))$$

being $\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}$-measurable, one has

$$\begin{aligned}&1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\\&\quad=\mathbb{E}[1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\\&\quad=1_{\{v_{q,l,j}(L)\le\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\}}\breve{E}^n_{q,l,j}(L;t)\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}].\end{aligned}$$

Since $v_{i,l,j}(L)$ is a positive random variable, independent of $\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}$, one has

$$\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]=\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\,|\,\breve{\gamma}_{i,l,j}(L)]=0;$$

thus, the first term on the right-hand side of (B.13) is equal to 0.

Next, we consider the second term on the right-hand side of (B.13). There exists a random variable $\check{\gamma}_{i,l,j}(L)$ which is a Borel function of ξ^n and all the service times and working times except $v_{q,l,j}(L)$ and $v_{i,l,j}(L)$, such that $\{v_{q,l,j}(L)>\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}=\{v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}$ and $\breve{\gamma}_{i,l,j}(L)=\check{\gamma}_{i,l,j}(L)$ on either event. Then, Krichagina and Puhalskii (1997, Lemma 3.6) and the fact that $v_{i,l,j}(L)$ and $v_{q,l,j}(L)$ are independent from $\check{\gamma}_{i,l,j}(L)$ and $\breve{\gamma}_{q,l,j}(L)$ yield

$$\begin{aligned}&1_{\{v_{q,l,j}(L)>\breve{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\mathbb{E}[\breve{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{\breve{\gamma}_{i,l,j}(L),l,j}(L)\cap\Gamma_{i,q}]\\&\quad=1_{\{v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\times\frac{\mathbb{E}[1_{\{v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\check{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\check{\gamma}_{i,l,j}(L),\breve{\gamma}_{q,l,j}(L)]}{\mathbb{P}[v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\,|\,\check{\gamma}_{i,l,j}(L),\breve{\gamma}_{q,l,j}(L)]},\end{aligned}$$

where $\check{E}^n_{i,l,j}(L)$ denotes $\breve{E}^n_{i,l,j}(L)$ with $\check{\gamma}_{i,l,j}(L)$ substituted for $\breve{\gamma}_{i,l,j}(L)$. Due to $v_{i,l,j}(L)$ is independent of $v_{q,l,j}(L)$, $\check{\gamma}_{i,l,j}(L)$ and $\breve{\gamma}_{i,l,j}(L)$, one obtains

$$\begin{aligned}&\mathbb{E}[1_{\{v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\check{E}^n_{i,l,j}(L;t)\breve{E}^n_{q,l,j}(L;t)\,|\,\check{\gamma}_{i,l,j}(L),\breve{\gamma}_{q,l,j}(L)]\\&\quad=\mathbb{E}[1_{\{v_{q,l,j}(L)>\check{\gamma}_{i,l,j}(L)-\breve{\gamma}_{q,l,j}(L)\ge0\}}\breve{E}^n_{q,l,j}(L;t)\,|\,\check{\gamma}_{i,l,j}(L),\breve{\gamma}_{q,l,j}(L)]\times\mathbb{E}[\check{E}^n_{q,l,j}(L;t)\,|\,\check{\gamma}_{i,l,j}(L)],\end{aligned}$$

where the last multiplier is 0 by definition of $\check{E}^n_{q,l,j}(L)$ and the fact that $v_{i,l,j}(L)$ is independent of $\check{\gamma}_{i,l,j}(L)$. In view of the preceding, the right-hand side of (B.13) is zero, as well as the right-hand side of (B.12). A similar reasoning can be used to show that (B.12) remains true when $\Gamma_{i,q}$ is replaced with $\underline{\Gamma}_{i,q}$. Consequently, (B.10) holds.

For the proof of (B.11), we use the same approach as for proof of (B.10); hence, some details are omitted. One has

$$\begin{aligned} &1_{\{\breve\gamma_{i,l,j}(L)\vee\breve\gamma_{q,l,j}(L)\le s\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\qquad+1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\qquad+1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\qquad+1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]. \end{aligned}\tag{B.14}$$

The terms in the preceding equality can be estimated as follows:

$$\begin{aligned} &1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=\mathbb{E}[1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}\breve E^n_{i,l,j}(L;t)1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\breve E^n_{i,l,j}(L;s)\breve E^n_{q,l,j}(L,s), \end{aligned}\tag{B.15}$$

$$\begin{aligned} &1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\breve E^n_{i,l,j}(L;s)\mathbb{E}[\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)\le s-\breve\gamma_{i,l,j}(L)\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\breve E^n_{i,l,j}(L;s)\breve E^n_{q,l,j}(L,s) \end{aligned}\tag{B.16}$$

and

$$\begin{aligned} &1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\breve E^n_{q,l,j}(L,s)\mathbb{E}[\breve E^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)\le s-\breve\gamma_{q,l,j}(L)\}}\breve E^n_{i,l,j}(L;s)\breve E^n_{q,l,j}(L,s), \end{aligned}\tag{B.17}$$

where the martingale property of $\breve E^n_{i,l,j}(L)$ and $\breve E^n_{q,l,j}(L)$ was used in (B.16) and (B.17).

Now, we consider the last term on the right-hand-side of (B.14). There exists random variables $\check\gamma_{i,l,j}(L)$ and $\check\gamma_{q,l,j}(L)$ which are Borel function of ξ^n and all the service times and working times except $v_{q,l,j}(L)$ and $v_{i,l,j}(L)$, such that $\{v_{i,l,j}(L)>s-\breve\gamma_{i,l,j}(L)\ge 0, v_{q,l,j}(L)>s-\breve\gamma_{q,l,j}(L)\ge 0\}=\{v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0, v_{q,l,j}(L)>s-\check\gamma_{q,l,j}(L)\ge 0\}$, and $\breve\gamma_{i,l,j}(L)=\check\gamma_{i,l,j}(L)$ and $\breve\gamma_{q,l,j}(L)=\check\gamma_{q,l,j}(L)$ on either event. Hence, it follows that

$$\begin{aligned} &1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)> s-\check\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\check\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)], \end{aligned}$$

where $v_{i,l,j}(L)$ and $v_{q,l,j}(L)$ are independent of $\check\gamma_{q,l,j}(L)$ and $\check\gamma_{i,l,j}(L)$. In Krichagina and Puhalskii (1997, Lemma 3.6) and

$$\begin{aligned} &\breve{\mathscr{E}}^n_{s,l,j}(L)\cap\{v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0\}\cap\{v_{q,l,j}(L)>s-\check\gamma_{q,l,j}(L)\ge 0\}\subset(\sigma(\{\xi^n,s_{o,m},a_{o,m},o=1,2,\dots,n,m\ge 1\}\backslash\{v_{i,l,j}(L),v_{q,l,j}(L)\})\vee\mathcal{N})\\ &\qquad\cap\{v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0\}\cap\{v_{q,l,j}(L)>s-\check\gamma_{q,l,j}(L)\ge 0\} \end{aligned}$$

yield

$$\begin{aligned} &1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)> s-\check\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\check\gamma_{q,l,j}(L)\ge 0\}}\frac{\mathbb{E}[1_{\{v_{i,l,j}(L)> s-\check\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\check\gamma_{q,l,j}(L)\ge 0\}}\breve E^n_{i,l,j}(L;t)\breve E^n_{q,l,j}(L,t)\,|\,\check\gamma_{i,l,j}(L),\check\gamma_{q,l,j}(L)]}{\mathbb{P}[v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0, v_{q,l,j}(L)>s-\check\gamma_{q,l,j}(L)\ge 0\,|\,\check\gamma_{i,l,j}(L),\check\gamma_{q,l,j}(L)]}\\ &\quad=1_{\{v_{i,l,j}(L)> s-\check\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\check\gamma_{q,l,j}(L)\ge 0\}}\frac{\mathbb{E}[1_{\{v_{i,l,j}(L)> s-\check\gamma_{i,l,j}(L)\ge 0\}}\breve E^n_{i,l,j}(L;t)\,|\,\check\gamma_{i,l,j}(L)]}{\mathbb{P}[v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0\,|\,\check\gamma_{i,l,j}(L)]}\times\frac{\mathbb{E}[1_{\{v_{q,l,j}(L)> s-\check\gamma_{q,l,j}(L)\ge 0\}}\breve E^n_{q,l,j}(L,t)\,|\,\check\gamma_{q,l,j}(L)]}{\mathbb{P}[v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0\,|\,\check\gamma_{i,l,j}(L)]}\\ &\quad=1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{i,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\times 1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\mathbb{E}[\breve E^n_{q,l,j}(L;t)\,|\,\breve{\mathscr{E}}^n_{s,l,j}(L)]\\ &\quad=1_{\{v_{i,l,j}(L)> s-\breve\gamma_{i,l,j}(L)\ge 0\}}1_{\{v_{q,l,j}(L)> s-\breve\gamma_{q,l,j}(L)\ge 0\}}\breve E^n_{i,l,j}(L;s)\breve E^n_{q,l,j}(L,s), \end{aligned}\tag{B.18}$$

where the second equality follows from the mutual independence of $v_{i,l,j}(L)$ and $v_{q,l,j}(L)$, as well as their independence of $\check\gamma_{q,l,j}(L)$ and $\check\gamma_{i,l,j}(L)$; the last equality is due to the fact that $\breve\gamma_{i,l,j}(L)=\check\gamma_{i,l,j}(L)$ and $\breve\gamma_{q,l,j}(L)=\check\gamma_{q,l,j}(L)$ on $\{v_{i,l,j}(L)>s-\breve\gamma_{i,l,j}(L)\ge 0, v_{q,l,j}(L)>s-\breve\gamma_{q,l,j}(L)\ge 0\}=\{v_{i,l,j}(L)>s-\check\gamma_{i,l,j}(L)\ge 0, v_{q,l,j}(L)>s-\check\gamma_{q,l,j}(L)\ge 0\}$, and the martingale property of $\breve E^n_{i,l,j}(L)$ and $\breve E^n_{q,l,j}(L)$.

Substituting (B.15)–(B.18) into (B.14), results in (B.11). This completes the proof. □

We conclude this section with the proofs of Lemmas 3 and 4.

Proof of Lemma 3. The proof is based on a decomposition of $\breve{M}^n_{l,j}(L)$ (an analogous decomposition holds for $\tilde{M}^n_{l,j}(L)$). In particular, Lemma 10 yields

$$\mathbb{E}\|\breve{M}^n_{l,j}(L)\|_t \le \mathbb{E}\|\breve{E}^n_{l,j}(L)\|_t + \mathbb{E}\sup_{s\le t}\left|\int_0^s \breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L};s-u)\,\mathrm{d}G_L(u)\right| + \mathbb{E}\sup_{s\le t}\left|\int_0^s \frac{\breve{B}^n_{l,j}(L;\breve{J}^n_{l,j}(L;s-u)/n, G_L(u-))}{\bar{G}(u-)}\,\mathrm{d}G_L(u)\right|. \tag{B.19}$$

Next, we bound the three terms on the right-hand side of (B.19). First, in Liptser and Shiryaev (1989, Theorem 1.9.5), Jensen's inequality for concave functions and Lemma 11 result in

$$\mathbb{E}\|\breve{E}^n_{l,j}(L)\|_t \le 3\mathbb{E}(\langle \breve{E}^n_{l,j}(L)\rangle(t))^{1/2} \le 3(\mathbb{E}\langle \breve{E}^n_{l,j}(L)\rangle(t))^{1/2} \le 3\left(\mathbb{E}\left[\frac{1}{n}\sum_{i:\breve{\gamma}_{i,l,j}(L)\le t}\int_0^{v_{i,l,j}(L)}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)}\right]\right)^{1/2}$$
$$\le 3(G^{(1_{\{L=H\}})} * \ddot{H} * H^{(l+j-1)}(t))^{1/2} \le 3(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2}, \tag{B.20}$$

where in the second to last inequality is based on

$$\mathbb{E}\int_0^{v_{i,l,j}(L)}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)} = 1; \tag{B.21}$$

everywhere in the paper $G * H^{(-1)}$ is interpreted as 1.

Second, by Krichagina and Puhalskii (1997, Lemma 3.1) (see also p. 252), $\{\breve{B}^n_{l,j}(L;t,G_L(u-)), t\ge 0\}$ is a locally square-integrable martingale with respect to an appropriately defined filtration; its predictable quadratic-variation process is given by

$$\langle \breve{B}^n_{l,j}(L;\cdot,G_L(u-))\rangle(t) = \frac{\lfloor nt\rfloor}{n} G_L(u-)\bar{G}_L(u-).$$

This, together with Liptser and Shiryaev (1989, Theorem 1.9.5), Jensen's inequality for concave functions and (B.4), yields

$$\mathbb{E}\sup_{s\le \breve{J}^n_{l,j}(L;t)/n}|\breve{B}^n_{l,j}(L;s,G_L(u-))|$$
$$\le 3\mathbb{E}(\langle \breve{B}^n_{l,j}(L;\cdot,G_L(u-))\rangle(\breve{J}^n_{l,j}(L;t)/n))^{1/2} \le 3(\mathbb{E}\langle \breve{B}^n_{l,j}(L;\cdot,G_L(u-))\rangle(\breve{J}^n_{l,j}(L;t)/n))^{1/2} \le 3(G^{(1_{\{L=H\}})} * \ddot{H} * H^{(l+j-1)}(t))^{1/2}(\bar{G}_L(x-))^{1/2}$$
$$\le 3(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2}(\bar{G}_L(x-))^{1/2}.$$

The preceding inequality allows one to bound the second term in (B.19):

$$\mathbb{E}\sup_{s\le t}\left|\int_0^s \frac{\breve{B}^n_{l,j}(L;\breve{J}^n_{l,j}(L;s-u)/n,G_L(u-))}{\bar{G}(u-)}\,\mathrm{d}G_L(u)\right|$$
$$\le \int_0^t \frac{\mathbb{E}\sup_{s\le \breve{J}^n_{l,j}(L;t)/n}|\breve{B}^n_{l,j}(L;s,G_L(u-))|}{\bar{G}(u-)}\,\mathrm{d}G_L(u)$$
$$\le 3(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2}\int_0^t \frac{\mathrm{d}G_L(u)}{(\bar{G}(u-))^{1/2}} \le 6(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2}, \tag{B.22}$$

where last inequality follows from the fact that the integral is upper bounded by 2. A bound for the third term in (B.19) is straightforward:

$$\mathbb{E}\sup_{s\le t}\left|\int_0^s \breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L};s-u)\,\mathrm{d}G_L(u)\right| \le \mathbb{E}\sup_{s\le t}\int_0^s \|\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L})\|_{s-u}\,\mathrm{d}G_L(u) \le \int_0^t \mathbb{E}\|\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L})\|_{t-u}\,\mathrm{d}G_L(u). \tag{B.23}$$

Next, we use induction to prove (the same bound holds for $\tilde{M}^n_{l,j}(L)$)

$$\mathbb{E}\|\breve{M}^n_{l,j}(L)\|_t \le 9(2l+1+1_{\{L=H\}})(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2}. \tag{B.24}$$

To this end, for $2l+1+1_{\{L=H\}}=1$ (i.e., $l=0$ and $L=G$), the result follows from (B.19), (B.20), (B.22) and $\breve{M}^n_{-1,j}(L)\equiv 0$. Now, suppose that (B.24) holds for $2l+1+1_{\{L=H\}}=q-1\ge 1$. Then, this inductive assumption, (B.19), (B.20), (B.22) and (B.23) imply that (B.24) holds for $2l+1+1_{\{L=H\}}=q$:

$$\mathbb{E}\|\breve{M}^n_{l,j}(L)\|_t \le 9(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2} + 9(2(l-1_{\{L=G\}})+1+1_{\{\underline{L}=H\}})\int_0^t (G^{(1_{\{\underline{L}=G\}})} * H^{(l-1_{\{L=G\}}+j-1-1_{\{\underline{L}=G\}})}(t-u))^{1/2}\,\mathrm{d}G_L(u)$$
$$\le 9(2l+1+1_{\{L=H\}})(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2},$$

where the last inequality follows from Jensen's inequality and

$$G^{(1_{\{L=G\}})} * H^{(l-1_{\{L=G\}}+j-1-1_{\{L=G\}})} * G_L(u) = G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}.$$

Finally, the union bound, Markov inequality and (B.24) yiled

$$\begin{aligned}\mathbb{P}\Bigg[\sum_{\substack{l+j>m\\ j\neq 0}} \|\breve{M}^n_{l,j}(L)\|_t + \|\tilde{M}^n_{l,j}(L)\|_t) > \epsilon\Bigg] &\le \sum_{\substack{l+j>m\\ j\neq 0}} (\mathbb{P}[\|\breve{M}^n_{l,j}(L)\|_t > \epsilon_{l,j}] + \mathbb{P}[\|\tilde{M}^n_{l,j}(L)\|_t > \epsilon_{l,j}])\\ &\le \sum_{\substack{l+j>m\\ j\neq 0}} 18\epsilon^{-1}_{l,j}(2l+1+1_{\{L=H\}})(G^{(1_{\{L=G\}})} * H^{(l+j-1-1_{\{L=G\}})}(t))^{1/2},\end{aligned} \tag{B.25}$$

where positive $\epsilon_{l,j}$'s satisfy

$$\epsilon = 2\sum_{\substack{l+j>0\\ j\neq 0}} \epsilon_{l,j}.$$

Since $H^{(l+j)}(t)$ decays exponentially fast in $(l+j)$, one can select $\{\epsilon_{l,j}\}$ in such a way that the sum in (B.25) is convergent for $m=1$ (e.g., $\{\epsilon_{l,j}\}$ polynomially decreasing in $(l+j)$). The statement of the lemma follows. □

Proof of Lemma 4. Straightforward algebra and (25) yield

$$\breve{M}^n_{l,0}(L;t) - \tilde{M}^n_{l,0}(L;t) = \breve{N}^n_l(L;t) - \tilde{N}^n_l(L;t), \tag{B.26}$$

where $\breve{N}^n_l(L) := \{\breve{N}^n_l(L;t), t\ge 0\}$, $\tilde{N}^n_l(L) := \{\tilde{N}^n_l(L;t), t\ge 0\}$,

$$\breve{N}^n_l(L;t) = \frac{1}{\sqrt{n}} \sum_{i=\xi^n\wedge k^n+1}^{\xi^n} (1_{\{c_{i,1:l}+s_{i,l+1}+a_{i,l+1}1_{\{L=H\}}\le t\}} - L * H^{(l)}(t))$$

and

$$\tilde{N}^n_l(L;t) = \frac{1}{\sqrt{n}} \sum_{i=\xi^n\wedge k^n+1}^{\xi^n} (1_{\{c_{i,1:l}+s_{i,l+1}+a_{i,l+1}1_{\{L=H\}}\le t-w_{i,0}\}} - L * H^{(l)}(t-w_{i,0})).$$

Applying Lemma 6 result in, as $n\to\infty$,

$$\sum_{L\in\{G,H\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}} \breve{N}^n_l(L) \Rightarrow 0. \tag{B.27}$$

Next, we consider $\tilde{N}^n_l(L)$. Based on an analogue of Lemma 10, $\tilde{N}^n_l(L)$ can be represented as a sum of three components, where the first one is a square-integrable martingale with an appropriate filtration (an analogue of Lemma 11). Then, the same argument as in the proof of Lemma 3 yields

$$\mathbb{E}\|\tilde{N}^n_l(L)\|_t \le 9\mathbb{E}\left[\frac{(\xi^n-k^n)^+}{n}\right] G^{(1_{\{L=G\}})} * H^{(l-1_{\{L=G\}})}(t); \tag{B.28}$$

in view of $0\le(\xi^n-k^n)^+/n\le 1$, the expected value of this random variable is well defined. This, together with (6) and the dominated convergence theorem, results in, as $n\to\infty$,

$$\frac{1}{n}\mathbb{E}(\xi^n-k^n)^+ \to 0. \tag{B.29}$$

Now, by Markov's inequality and (B.28), one has

$$\begin{aligned}&\limsup_{n\to\infty} \mathbb{P}\left[\left\|\sum_{l=0}^{\infty}\sum_{L\in\{G,H\}} (-1)^{1_{\{L=H\}}} \tilde{N}^n_l(L)\right\|_t > \epsilon\right]\\ &\le \limsup_{n\to\infty} \sum_{l=0}^{\infty}\sum_{L\in\{G,H\}} \mathbb{P}[\|\tilde{N}^n_l(L)\|_t > \epsilon_l] \le \limsup_{n\to\infty} \mathbb{E}\left[\frac{(\xi^n-k^n)^+}{n}\right] \sum_{l=0}^{\infty}\sum_{L\in\{G,H\}} 9\epsilon^{-1}_l G^{(1_{\{L=G\}})} * H^{(l-1_{\{L=G\}})}(t),\end{aligned}$$

where ϵ_l are positive numbers such that $2\sum_{l=0}^{\infty}\epsilon_l = \epsilon$; they can be chosen in a manner that the double summation is finite. This fact and (B.29) imply, as $n\to\infty$,

$$\sum_{L\in\{G,H\}} \sum_{l=0}^{\infty} (-1)^{1_{\{L=H\}}} \tilde{N}^n_l(L) \quad\Rightarrow\quad 0.$$

Equality (B.26), the last limit and (B.27) yield the statement of the lemma. □

B.2. Tightness

In this subsection, we state a technical result.

Lemma 12. *If the sequence* $\{\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L}), n\geq 1\}$ *(resp.* $\{\tilde{M}^n_{l-1_{\{L=G\}},j}(\underline{L}), n\geq 1\}$*) is C-tight, then* $\{\breve{M}^n_{l,j}(L), n\geq 1\}$ *(resp.* $\{\tilde{M}^n_{l,j}(L), n\geq 1\}$*) is tight.*

Remark 7. The sequences $\{\breve{M}^n_{-1,j}(H), n\geq 1\}$ and $\{\tilde{M}^n_{-1,j}(H), n\geq 1\}$ are C-tight, since $\breve{M}^n_{-1,j}(H)\equiv \tilde{M}^n_{-1,j}(H)\equiv 0$. The proof of Lemma 2 is based on the fact that the sequence $\{\breve{M}^n_{l,j}(L)\}_n$ ($\{\tilde{M}^n_{l,j}(L)\}_n$) is tight. Under this condition, Lemma 2 states that the limit of $\{\breve{M}^n_{l,j}(L)\}_n$ has a.s. continuous paths. Hence, the tightness of $\{\breve{M}^n_{l,j}(L)\}_n$ implies the C-tightness of the same sequence.

Proof. First, by Aldous' sufficient criterion (Liptser and Shiryaev 1989, p. 515), we argue that $\{\breve{E}^n_{l,j}(L;t), n\geq 1\}$ is tight. In particular, for all $T>0$ and $\epsilon>0$, we show that

$$\lim_{\kappa\to\infty}\limsup_{n\to\infty}\mathbb{P}[\|\breve{E}^n_{l,j}(L;t)\|_T>\kappa]=0 \tag{B.30}$$

and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\tau\in\mathcal{T}^n_{l,j}(L)}\mathbb{P}\Big[\sup_{0\leq t\leq\delta}|\breve{E}^n_{l,j}(L;\tau+t)-\breve{E}^n_{l,j}(L;\tau)|>\epsilon\Big]=0, \tag{B.31}$$

where $\mathcal{T}^n_{l,j}(L)$ is is the set of all $\breve{\mathbf{E}}^n_{l,j}(L)$-stopping times no greater than T. The proof of (B.30) is straightforward—(B.20) and Markov inequality yield

$$\mathbb{P}[\|\breve{E}^n_{l,j}(L;t)\|_T>\kappa]\leq 3/\kappa.$$

Next, observe that

$$\sum_{i=1}^{\breve{A}^n_{j+l}(t)}1_{\{s^j_{i,j+l+1}1_{\{L=H\}}\leq t\}}\leq \breve{J}^n_{l,j}(L;t)\leq\sum_{i=1}^{n}1_{\{c^j_{i,1:j+l}+s^j_{i,j+l+1}1_{\{L=H\}}\leq t\}}. \tag{B.32}$$

Applying the argument used in Lemma 9 results in, as $n\to\infty$,

$$\left\{\frac{1}{n}\sum_{i=1}^{\breve{A}^n_{j+l}(t)}1_{\{s^j_{i,j+l+1}1_{\{L=H\}}\leq t\}}, t\geq 0\right\}\Rightarrow G^{(1_{\{L=H\}})}*\ddot{H}*H^{(l+j-1)}; \tag{B.33}$$

a SLLN (e.g., see Daffer and Taylor 1979) can be invoked to show, as $n\to\infty$,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}1_{\{c^j_{i,1:j+l}+s^j_{i,j+l+1}1_{\{L=H\}}\leq t\}}, t\geq 0\right\}\Rightarrow G^{(1_{\{L=H\}})}*\ddot{H}*H^{(l+j-1)}. \tag{B.34}$$

Therefore, from (B.32)–(B.34) and the fact that $G^{(1_{\{L=H\}})}*\ddot{H}*H^{(l+j-1)}$ is a continuous function, it follows that $\{\breve{J}^n_{l,j}(L)/n, n\geq 1\}$ is C-tight.

Now, Lemma 11 and the Lenglart-Rebolledo inequality (e.g., see Liptser and Shiryaev 1989, p. 66) imply, for any $\kappa>0$ and any $\tau\in\mathcal{T}^n_{l,j}(L)$,

$$\begin{aligned}\mathbb{P}\Big[\sup_{0\leq t\leq\delta}|\breve{E}^n_{l,j}(L;\tau+t)-\breve{E}^n_{l,j}(L;\tau)|>\epsilon\Big]&\leq\kappa/\epsilon^2+\mathbb{P}[\langle\breve{E}^n_{l,j}(L)\rangle(\tau+\delta)-\langle\breve{E}^n_{l,j}(L)\rangle(\tau)>\kappa]\\ &\leq\kappa/\epsilon^2+\mathbb{P}\Bigg[\frac{1}{n}\sup_{\substack{s\leq T\\0<t-s\leq\delta}}\sum_{i:s<\breve{\gamma}_{i,l,j}(L)\leq t}\int_0^{v_{i,l,j}(L)}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)}>\kappa\Bigg].\end{aligned} \tag{B.35}$$

Due to the fact that $\#\{i\colon s<\breve{\gamma}_{i,l,j}(L)\leq t\}=\breve{J}^n_{l,j}(t)-\breve{J}^n_{l,j}(s)$, the fact that $\{\breve{J}^n_{l,j}(L)/n, n\geq 1\}$ is C-tight, (B.21) and the SLLN, one has

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\Bigg[\frac{1}{n}\sup_{\substack{s\leq T\\0<t-s\leq\delta}}\sum_{i:s<\breve{\gamma}_{i,l,j}(L)\leq t}\int_0^{v_{i,l,j}(L)}\frac{\mathrm{d}G_L(u)}{\bar{G}_L(u-)}>\kappa\Bigg]=0. \tag{B.36}$$

Thus, (B.31) follows from (B.35) and (B.36).

Second, combining C-tightness of $\{\breve{J}^n_{l,j}(L)/n, n\geq 1\}$ and the argument used in the proof of Lemma 3.4 in Krichagina and Puhalskii (1997) yields that the sequence

$$\left\{\left\{\int_0^t\frac{\breve{B}^n_{l,j}(L;\breve{J}^n_{l,j}(L;t-u)/n, G_L(u-))}{\bar{G}_L(u-)}\,\mathrm{d}G_L(u), t\geq 0\right\}, n\geq 1\right\}$$

is C-tight.

Third, introduce a map f_L: $D[0,\infty)\to D[0,\infty)$:

$$f_L(x)(t)=\int_0^t x(t-s)\,\mathrm{d}G_L(s),$$

for $t\geq 0$ and $x\in D[0,\infty)$. Then, f_L is continuous in the topology of uniform convergence. Moreover, if $x\in C[0,\infty)$, then $f_L(u)\in C[0,\infty)$. The preceding, the C-tightness of $\{\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L}), n\geq 1\}$ and the fact that convergence to a continuous function is equivalent in J_1 and uniform topologies imply that the sequence $\{f_L(\breve{M}^n_{l-1_{\{L=G\}},j}(\underline{L})), n\geq 1\}$ is C-tight.

The proof of lemma follows from the above, Lemma 10 and Corollary VI.3.33 in Jacod and Shiryaev (2003). □

Appendix C. Dependency Graph

$$\begin{tikzcd}
\text{Lemma 7} \arrow[d] & & & & \\
\text{Proposition 1} \arrow[r] \arrow[dd] & \text{Corollary 1} \arrow[d] & & & \text{Lemma 8} \arrow[llll, shorten <=0pt] \\
& \text{Lemma 9} \arrow[r] & \text{Lemma 2} \arrow[d] & & \text{Lemma 12} \arrow[u] \arrow[ll] \\
\text{Theorem 1} & & \text{Lemma 5} \arrow[ll] & \text{Lemma 3} \arrow[l] & \text{Lemma 10/11} \arrow[l] \arrow[u] \\
\text{Lemma 1} \arrow[u, dashed] & \text{Lemma 6} \arrow[l] \arrow[r] & \text{Lemma 4} \arrow[u] & &
\end{tikzcd}$$

References

Armony M, Maglaras C (2004) Contact centers with a call-back option and real-time delay information. *Oper. Res.* 52(4):527–545.
Armony M, Ward A (2010) Fair dynamic routing policies in large-scale systems with heterogeneous servers. *Oper. Res.* 58(3):624–637.
Asmussen S (2003) *Applied Probability and Queues*, 2nd ed. (Springer, New York).
Ata B, Gurvich I (2012) On optimality gaps in the Halfin-Whitt regime. *Ann. Appl. Probab.* 22(1):407–455.
Atar R (2005a) A diffusion model of scheduling control in queueing systems with many servers. *Ann. Appl. Probab.* 15(1B):820–852.
Atar R (2005b) Scheduling control for queueing systems with many servers: Asymptotic optimality in heavy traffic. *Ann. Appl. Probab.* 15(4):2606–2650.
Atar R, Mandelbaum A, Reiman M (2004) Scheduling a multi class queue with many exponential servers: Asymptotic optimality in heavy traffic. *Ann. Appl. Probab.* 14(3):1084–1134.
Atar R, Shaki Y, Shwartz A (2011) A blind policy for equalizing cumulative idleness. *Queueing Syst. Theory Appl.* 67(4):275–293.
Bassamboo A, Kumar S, Randhawa R (2009) Dynamics of new product introduction in closed rental systems. *Oper. Res.* 57(6):1347–1359.
Billingsley P (1995) *Probability and Measure*, 3rd ed. (John Wiley & Sons, New York).
Billingsley P (1999) *Convergence of Probability Measures*, 2nd ed. (John Wiley & Sons, New York).
Brémaud P (1981) *Point Processes and Queues: Martingale Dynamics* (Springer, New York).
Browne S, Whitt W (1995) Piecewise-linear diffusion processes. Dshalalow J, ed. *Probability and Stochastic Series: Advances in Queueing: Theory, Methods and Open Problems* (CRC Press, Boca Raton, FL), 463–480.
Çinlar E (1975) *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ).
Csörgő M, Révész P (1981) *Strong Approximations in Probability and Statistics* (Academic Press, New York).
Daffer P, Taylor R (1979) Laws of large numbers for $D[0,1]$. *Ann. Probab.* 7(1):85–95.
Dai J, He S (2010) Customer abandonment in many-server queues. *Math. Oper. Res.* 35(2):347–362.
Dębicki K, Palmowski Z (1999) On-off fluid models in heavy traffic environment. *Queueing Syst. Theory Appl.* 33(4):327–338.
de Véricourt F, Jennings O (2008) Dimensioning large-scale membership services. *Oper. Res.* 56(1):173–187.
de Véricourt F, Jennings O (2011) Nurse staffing in medical units: A queueing perspective. *Oper. Res.* 59(6):1320–1331.
Dudley R (2002) *Real Analysis and Probability* (Cambridge University Press, Cambridge, UK).
Erlang A (1948) On the rational determination of the number of circuits. Brockmeyer E, Halstrom H, Jensen A, eds. *The Life and Works of A. K. Erlang* (The Copenhagen Telephone Company, Copenhagen), 216–221.
Gamarnik D, Goldberg D (2013a) On the rate of convergence to stationarity of the $M/M/N$ queue in the Halfin-Whitt regime. *Ann. Appl. Probab.* 23(5):1879–1912.
Gamarnik D, Goldberg D (2013b) Steady-state $GI/GI/N$ queue in the Halfin-Whitt regime. *Ann. Appl. Probab.* 23(6):2382–2419.
Gamarnik D, Momčilović P (2008) Steady-state analysis of a multi-server queue in the Halfin-Whitt regime. *Adv. Appl. Probab.* 40(2):548–577.
Green L, Savin S, Murray M (2007) Providing timely access to care: What is the right patient panel size? *Jt. Comm. J. Qual. Patient Saf.* 33(4): 211–218.
Gurvich I, Whitt W (2009) Queue-and-idleness-ratio controls in many-server service systems. *Math. Oper. Res.* 34(2):363–396.
Hahn M (1978) Central limit theorems in $D[0,1]$. *Z. Wahrsch. Verw. Gebiete* 44(2):89–101.
Halfin S, Whitt W (1981) Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* 29(3):567–588.
Haque L, Armstrong M (2007) A survey of the machine interference problem. *Eur. J. Oper. Res.* 179(2):469–482.
Harrison J, Williams R (1996) A multiclass closed queueing network with unconventional heavy traffic behavior. *Ann. Appl. Probab.* 6(1):1–47.

Harrison J, Zeevi A (2004) Dynamic scheduling of a multiclass queue in the Halfin-Whitt heavy traffic regime. *Oper. Res.* 52(2):243–257.
Heath D, Resnick S, Samorodnitsky G (1998) Heavy tails and long range dependence in on/off processes and associated fluid models. *Math. Oper. Res.* 23(1):145–165.
Iglehart D (1965) Limit diffusion approximations for many server queue and the repairman problem. *J. Appl. Probab.* 2(2):429–441.
Jacod J, Shiryaev A (2003) *Limit Theorems for Stochastic Processes*, 2nd ed. (Springer, Berlin).
Jagerman D (1974) Some properties of the Erlang loss function. *Bell System Techn. J.* 53(3):525–551.
Janssen A, van Leeuwaarden J, Zwart B (2011) Refining square-root safety staffing by expanding Erlang C. *Oper. Res.* 59(6):1512–1522.
Jelenković P, Mandelbaum A, Momčilović P (2004) Heavy traffic limits for queues with many deterministic servers. *Queueing Syst. Theory Appl.* 47(1–2):53–69.
Jennings O, Mandelbaum A, Massey W, Whitt W (1996) Server staffing to meet time-varying demand. *Management Sci.* 42(10):1383–1394.
Kaspi H, Ramanan K (2013) SPDE limits of many-server queues. *Ann. Appl. Probab.* 23(1):145–229.
Kiefer J (1972) Skorohod embedding of multivariate rv's and the sample df. *Z. Wahrsch. Verw. Gebiete* 24(1):1–35.
Kleinrock L (1975) *Queueing Systems*, Vol. 1 (John Wiley & Sons, New York).
Kogan Y, Liptser R, Smorodinskii A (1986) Gaussian diffusion approximation of closed Markov models of computer networks. *Problems Inform. Transmission* 22(1):38–51.
Krichagina E, Puhalskii A (1997) A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. *Queueing Syst. Theory Appl.* 25(1–4):235–280.
Kumar S (2000) Two-server closed networks in heavy traffic: Diffusion limits and asymptotic optimality. *Ann. Appl. Probab.* 10(3):930–961.
Liptser R, Shiryaev A (1989) *Theory of Martingales* (Kluwer, Dordrecht, Netherlands).
Mandelbaum A, Momčilović P (2008) Queues with many servers: The virtual waiting-time process in the QED regime. *Math. Oper. Res.* 33(3):561–586.
Mandelbaum A, Momčilović P (2012) Queues with many servers and impatient customers. *Math. Oper. Res.* 37(1):41–64.
Mandelbaum A, Pats G (1995) State-dependent queues: Approximations and applications. Kelly F, Williams R, eds. *Stochastic Networks*. IMA Volumes in Mathematics, Vol. 71 (Springer, New York), 239–282.
Mandelbaum A, Momčilović P, Tseytlin Y (2012) On fair routing from emergency departments to hospital wards: QED queues with heterogeneous servers. *Management Sci.* 58(7):1273–1291.
Pang G, Talreja R, Whitt W (2007) Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surveys* 4:193–267.
Puhalskii A, Reed J (2010) On many-server queues in heavy traffic. *Ann. Appl. Probab.* 20(1):129–195.
Puhalskii A, Reiman M (2000) The multiclass $GI/PH/N$ queue in the Halfin-Whitt regime. *Adv. Appl. Probab.* 32(3):564–595.
Randhawa R, Kumar S (2009) Multiserver loss systems with subscribers. *Math. Oper. Res.* 34(1):142–179.
Rao R (1963) The law of large numbers for $D[0,1]$-valued random variables. *Theory Probab. Appl.* 8(1):70–74.
Reed J (2009) The $G/GI/N$ queue in the Halfin-Whitt regime. *Ann. Appl. Probab.* 19(6):2211–2269.
Resnick S (1992) *Adventures in Stochastic Processes* (Birkhäuser, Boston).
Ross S (1997) *Stochastic Processes*, 2nd ed. (John Wiley & Sons, New York).
Rudin W (1991) *Functional Analysis*, 2nd ed. (McGraw-Hill, Boston).
Sigman K, Whitt W (2011) Heavy-traffic limits for nearly deterministic queues: Stationary distributions. *Queueing Syst. Theory Appl.* 69(2): 145–173.
Skorokhod A (1956) Limit theorems for stochastic processes. *Theory Probab. Appl.* 1(3):261–290.
Stone C (1963) Limit theorems for random walks, birth and death processes, and diffusion processes. *Illinois J. Math.* 7(4):638–660.
Szczotka W (1980) Central limit theorem in $D[0,\infty)$ for breakdown processes. *Probab. Math. Stat.* 1(1):49–57.
Tezcan T (2008) Optimal control of distributed parallel server systems under the Halfin and Whitt regime. *Math. Oper. Res.* 33(1):51–90.
van Leeuwaarden J, Knessl C (2012) Spectral gap of the Erlang A model in the Halfin-Whitt regime. *Stochastic Syst.* 2(1):149–207.
Whitt W (1980) Some useful functions for functional limit theorems. *Math. Oper. Res.* 5(1):67–85.
Whitt W (1985) Queues with superposition arrival processes in heavy traffic. *Stochastic Process. Appl.* 21(1):81–91.
Whitt W (2002) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (Springer, New York).
Whitt W (2007) Proofs of the martingale FCLT. *Probab. Surveys* 4:268–302.
Yom-Tov G, Mandelbaum A (2014) Erlang-R: A time-varying queue with ReEntrant customers, in support of healthcare staffing. *Manufacturing Service Oper. Management* 16(2):283–299.
Zhang B, van Leeuwaarden J, Zwart B (2011) Staffing call centers with impatient customers: Refinements to many-server asymptotics. *Oper. Res.* 60(2):461–474.