Mean field for performance models with generally-distributed timed transitions

Richard A. Hayden[1], Illés Horváth[2], and Miklós Telek[3,4]

[1] Department of Computing, Imperial College London, United Kingdom
[2] MTA-BME Information Systems Research Group
[3] Budapest University of Technology and Economics
Department of Networked Systems and Services
[4] Inter-University Center of Telecommunications and Informatics,Debrecen

Abstract. In this paper we extend the mean-field limit of a class of stochastic models with exponential and deterministic delays to include exponential and generally-distributed delays. Our main focus is the rigorous proof of the mean-field limit.
Keywords: mean-field limit, generally-distributed delay, delayed differential equation.

1 Introduction

In this paper, we further develop the new mean-field methodology introduced in [16] for a class of massively-parallel *generalised semi-Markov processes (GSMPs)* [20, 14, 15]. We focus on population models where individuals can enable both Markovian and generally-timed transitions, which are crucial for the accurate modelling of many real-world computer and networking protocols. We encode such models in a low-level formalism, the *population generalised semi-Markov process (PGSMP)*.

The motivation for the mean-field approach is the same as in the continuous-time Markov chain (CTMC) case — unsurprisingly, GSMP models with many components also become computationally intractable to explicit state techniques [7, 9] rapidly as a result of the familiar state-space explosion problem. Our approach is based on the derivation of *delay differential equations (DDEs)* from PGSMP models and generalises the traditional mean-field approach as applied to CTMC models based on ordinary differential equations (ODEs) [1, 13, 4, 17].

The class of models to which our approach applies is very broad — the only significant restriction we make is that at most one generally-timed transition may be enabled by each individual in any given local state. However, globally, there is no restriction on the concurrent enabling of generally-timed transitions by different individuals.

As in the CTMC case, the size of the system of DDEs is equal to the number of physical local states that components in the model can be in. Therefore this approach represents a significant improvement with respect to both accuracy and efficiency when compared with the traditional CTMC mean-field

approach where generally-timed transitions are approximated using phase-type distributions. The mean-field approach based on DDEs presented here captures generally-timed distributions directly without the need for additional physical states or for approximation of the generally-timed distribution itself.

This paper focuses on the non-racing case, that is, under the assumption that generally-timed transitions do not compete locally with exponential transitions (hence the term *delay-only* PGSMPs). The main contribution is to show how systems of coupled DDEs can be derived directly from PGSMP models with generally-timed transitions, and to give a full proof of transient mean-field convergence. The paper is quite proof-heavy; for a worked example, see Section 4, and for more examples, we refer to [16] and [5].

1.1 Related work

Related work can be found in the biology and chemistry literature. Systems of DDEs have been derived to approximate stochastic models of reaction networks where deterministic delays are possible after reactions occur [3, 6, 21]. However, these models differ from those considered here in a number of critical ways; most importantly, the presentation in this paper lacks the severe rigidity of models encountered in biology and chemistry, making it suitable for a much larger class of population models.

Closest related work is due to [16] and [5] which both deal with deterministic delay-only PGSMPs in different ways; our presentation is closest in spirit to [16], but the upgrade from deterministic delays to generally-timed delays calls for a careful and involved analysis.

The approach in [5] highlights the connection to ODE approximations of DDEs [19] which is directly analogous to the Erlang approximation of the delay in the PGSMP. The approach in the present paper, however, avoids any Erlang approximations whatsoever, proving the mean-field limit directly via probability concentration theorems.

2 Population generalised semi-Markov processes

2.1 Definition of PGSMPs

A PGSMP model consists of many interacting components each inhabiting one of a finite set of *local states* $\mathcal{S}$. The global state space, say $\mathcal{X}$, of a PGSMP model then consists of elements $\mathbf{x} = (x_s)_{s \in \mathcal{S}}$ where each $x_s \in \mathbf{Z}^+$ tracks the number of components currently in the local state s.

Exponential transitions are specified by a finite set of *Markovian transitions* $\mathcal{C}$. Each $c \in \mathcal{C}$ specifies a finite *change multiset* L_c, which consists of tuples $(s, s') \in \mathcal{S} \times \mathcal{S}$ each of which specifies that as part of a c-transition, a distinct, uniformly randomly selected component currently in local state s moves to local state s'. Write also $\mathbf{l}^c = (l_s^c)_{s \in \mathcal{S}}$ where each $l_s^c := |\{(s', s) \in L_c\}| - |\{(s, s') \in L_c\}|$, which represents the total change in components that are in local state s when a

c-transition occurs. The aggregate rate of c-transitions is given by a *rate function* $r_c : \mathcal{X} \to \mathbf{R}^+$. We assume that the rate function is defined such that it is zero whenever a transition is not possible due to there not being enough distinct components in the required local states.

Generally-timed transitions are specified by *event clocks* in a similar fashion to standard GSMPs [14]. Specifically, we assume a finite set of event clocks $\mathcal{E}$. Each event clock $e \in \mathcal{E}$ is specified by its *set of active states* $\mathcal{A}_e \subseteq \mathcal{S}$, its *event transition probability function* $p_e : \mathcal{S} \times \mathcal{S} \to [0,1]$ and the *clock time distribution* given by a cumulative distribution function (CDF) F_e used to the set the clock.

When a component enters a state s in $\mathcal{A}_e$ for the first time, the clock is initialized according to the CDF F_e. After the clock time has elapsed, it moves immediately to a new local state by sampling from the discrete probability distribution $p_e(s, \cdot)$. The clock is disabled when the component leaves s and is reset by resampling from the distribution if it later returns to the set of active states.

As mentioned above, the key restriction we make for all PGSMP models considered in this paper is that at most one event clock may be active in any local state. That is, for each $s \in \mathcal{S}$, $|\{e \in \mathcal{E} : s \in \mathcal{A}_e\}| \leq 1$. We will see that this restriction is necessary for the mean-field analyses presented in the sequel. This restriction also means that, with probability one, it is not possible for two transitions (Markovian or generally-timed) to occur simultaneously within a single component. Finally, we write $\mathbf{x}^0$ for the initial state of the model.

2.2 Delay-only PGSMPs

We will focus on a class of PGSMPs with the structural restriction that, within a given component, generally-timed transitions may not be enabled concurrently with Markovian ones. We refer to such models as *delay only* since the general transitions in the constituent components then serve only to introduce generally-distributed delays between periods of otherwise Markovian behaviour.

Formally, the class of delay-only PGSMPs is specified by two restrictions: for all $e \in \mathcal{E}$; if $s \in \mathcal{A}_e$ then there can be no $c \in \mathcal{C}$ with $(s, s') \in L_c$; and for each $s \in \mathcal{A}_e$, there must exist some $s' \in \mathcal{S}$ such that $s' \notin \mathcal{A}_{e'}$ for any $e' \in \mathcal{E}$ and $p_e(s, s') = 1$. The first restriction guarantees that no Markovian transitions are enabled concurrently with general transitions, as above. The second restriction guarantees, firstly, that after any general transition completes, the component jumps into a unique state.[5] Secondly, it also ensures that the completion of a general transition cannot immediately enable another.

2.3 Construction of delay-only PGSMPs in terms of Poisson processes

In this section we give a construction of the population processes of a delay-only PGSMP in terms of Poisson processes. Write $\mathbf{x}(t) \in \mathbf{Z}^{+|\mathcal{S}|}$ for the underlying

---

[5] This is a technical but not, in fact, a modelling restriction, as the state space may be reconfigured so the general transition is followed by a Markovian transition sampling from any discrete probability distribution p_e.

population process of a delay-only PGSMP, where $x_s(t) \in \mathbf{Z}^+$ tracks the number of components currently in the local state s.

Now let $\{P_c \,:\, c \in \mathcal{C}\}$ be a set of mutually-independent rate-1 Poisson processes and, for each $e \in \mathcal{E}$, let $\{T_i^e\}_{i=1}^{\infty}$ be mutually independent sequences of identically-distributed random variables distributed according to F_e, all also mutually independent of the Poisson processes. Then we may write:

$$
\begin{aligned}
x_s(t) = & x_s^0 + \sum_{c \in \mathcal{C}} l_s^c P_c \left(\int_0^t r_c(\mathbf{x}(u)) \, \mathrm{d}u \right) \\
& + \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}} \left(\sum_{s' \in \mathcal{A}_e} p_e(s', s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right) \\
& \times \int_{z=0}^t \mathbf{1}_{\left\{ T^e_{P_c\left(\int_0^z r_c(\mathbf{x}(u))\,\mathrm{d}u\right)} \leq t - z \right\}} \, \mathrm{d}P_c \left(\int_0^z r_c(\mathbf{x}(u)) \, \mathrm{d}u \right) \quad (1)
\end{aligned}
$$

This is similar to the direct definition of [16]; the extra variables T_k^e are necessary due to the more complicated nature of the process.

3 Mean-field approximation of delay-only PGSMPs

For each local state $s \in \mathcal{S}$, we write $v_s(t)$ for the mean-field approximation to the number of components in state s at time $t \in \mathbf{R}^+$ and we also let $\mathbf{v}(t) = (v_s(t))_{s \in \mathcal{S}}$. The mean-field approximations satisfy the following system of integral equations:

$$
\begin{aligned}
v_s(t) = & v_s^0 + \sum_{c \in \mathcal{C}} l_s^c \int_0^t r_c(\mathbf{v}(u)) \, \mathrm{d}u \\
& + \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}} \left(\sum_{s' \in \mathcal{A}_e} p_e(s', s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right) \times \int_0^t F_e(t - u) r_c(\mathbf{v}(u)) \, \mathrm{d}u \quad (2)
\end{aligned}
$$

3.1 Transient mean-field convergence

In this section of the paper, we prove transient mean-field convergence for delay-only PGSMPs. We begin by constructing a sequence of delay-only PGSMP models indexed by $N \in \mathbf{Z}^+$ with increasing total component population size. Their underlying stochastic processes are denoted $\{\mathbf{x}^N(t) \in \mathbf{R}^{+|\mathcal{S}|}\}_{N \in \mathbf{Z}^+}$, where $\mathbf{x}^N(t) = (x_s^N(t))_{s \in \mathcal{S}}$ and $x_s^N(t) \in \mathbf{Z}^+$ tracks the number of components currently in the local state s for the Nth model.

We assume that the set of local states $\mathcal{S}$; the set of transitions $\mathcal{C}$ and the change multisets L_c; the sets of event clocks $\mathcal{E}$, the sets of active states $\mathcal{A}_e$, the transition probability functions p_e and the delay CDFs F_e are all fixed for all elements of the sequence. The rate functions r_c^N are allowed to vary with N and the initial conditions for the Nth model in the sequence are given by $N\mathbf{x}^0$ for

some $\mathbf{x}^0 \in \mathbf{Z}^{+|\mathcal{S}|}$. Write $\mathcal{X}^N \subseteq \mathbf{Z}^{+|\mathcal{S}|}$ for the reachable state space of the Nth model. Note that following Section 2.3, we may write, in terms of a single set of Poisson processes $\{P_c \,:\, c \in \mathcal{C}\}$ and delay variables $\{T_i^e \,:\, e \in \mathcal{E}\}_{i=1}^\infty$:

$$x_s^N(t) = x_s^N(0) + \sum_{c \in \mathcal{C}} l_s^c P_c \left(\int_0^t r_c^N(\mathbf{x}^N(u))\,\mathrm{d}u \right)$$

$$+ \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}} \left(\sum_{s' \in \mathcal{A}_e} p_e(s',s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right)$$

$$\times \int_0^t \mathbf{1}_{\left\{ T^e_{P_c\left(\int_0^z r_c^N(\mathbf{x}^N(u))\,\mathrm{d}u\right)} \leq t - z \right\}} \,\mathrm{d}P_c \left(\int_0^z r_c^N(\mathbf{x}^N(u))\,\mathrm{d}u \right) \qquad (3)$$

Similarly to the case of *density-dependent Markov chains* [12, 18], we assume that we may define $r_c(\mathbf{x}) := (1/N) r_c^N(N\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R}^{+|\mathcal{S}|}$ independently of N. Furthermore, we assume that r_c satisfies a local Lipschitz condition on $\mathbf{R}^{+|\mathcal{S}|}$ and that for all $c \in \mathcal{C}$, $r_c^N(\mathbf{x}) \leq R\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathcal{X}^N$ where $R \in \mathbf{R}^+$ is independent of N. Define the rescaled processes $\bar{\mathbf{x}}^N(t) := (1/N)\mathbf{x}^N(t)$ that thus satisfy $r_c(\bar{\mathbf{x}}^N(t)) = r_c^N(\mathbf{x}^N(t))$.

We assume that initially, the system is concentrated on the non-active states $\mathcal{C} \setminus \cup_{e \in \mathcal{E}} \mathcal{A}_e$, in which case no initialization is necessary for the non-Markovian clocks. (For a discussion of the issue of initialization in the deterministic delay case, see [5]). We also assume

$$\|\mathbf{v}^0 - \bar{\mathbf{x}}^N(0)\| \to 0.$$

Note that in most applications, it is perfectly natural to set a deterministic initial condition, but we may also allow $\bar{\mathbf{x}}^N(0)$ to be random; in this case, assume

$$\mathbf{P}(\|\mathbf{v}^0 - \bar{\mathbf{x}}^N(0)\| > \epsilon) \to 0.$$

Theorem 1. *Under the assumptions and setup given above, we have, for any $T > 0$ and $\epsilon > 0$:*

$$\lim_{N \to \infty} \mathbf{P} \left\{ \sup_{t \in [0,T]} \|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| > \epsilon \right\} = 0$$

Remark. Actually, assuming the initial condition converges almost surely, we have almost sure convergence in the theorem, which is stronger than convergence in probability, but, since the probabilistic coupling presented in (3) is a technical issue with no underlying deeper connection, there is not much of a difference.

Proof. Define the auxiliary process

$$y_s^N(t) := v_s^0 + \sum_{c \in \mathcal{C}} l_s^c \int_0^t r_c(\bar{\mathbf{x}}^N(u))\,\mathrm{d}u$$

$$+ \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}} \left(\sum_{s' \in \mathcal{A}_e} p_e(s',s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right) \times \int_0^t F_e(t - u) r_c(\bar{\mathbf{x}}^N(u))\,\mathrm{d}u$$

Then
$$|\bar{x}_s^N(t) - v_s(t)| \le |\bar{x}_s^N(t) - y_s^N(t)| + |y_s^N(t) - v_s(t)|.$$

Denote
$$D_s^N(T) = \sup_{t \in [0,T]} |\bar{x}_s^N(t) - y_s^N(t)|$$

We estimate $\mathbf{y}^N(t) - \mathbf{v}(t)$ by

$$|y_s^N(t) - v_s(t)| \le \sum_{c \in \mathcal{C}} |l_s^c| \int_0^t |r_c^N(\mathbf{x}^N(u)) - r_c(\mathbf{v}(u))| \, \mathrm{d}u$$

$$+ \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}} \left| \sum_{s' \in \mathcal{A}_e} p_e(s',s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right| \times \int_0^t F_e(t-u) |r_c(\bar{\mathbf{x}}^N(u)) - r_c(\mathbf{v}(u))| \, \mathrm{d}u$$

$$\le ZR \int_0^t \|\mathbf{x}^N(u) - \mathbf{v}(u)\| \mathrm{d}u$$

where

$$Z := \sum_{c \in \mathcal{C}} |l_{k,s}^c| + \sum_{e \in \mathcal{E}_k} \sum_{c \in \mathcal{C}} \left| \sum_{s' \in \mathcal{A}_e} p_e(s',s) l_{k,s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_{k,s}^c \right|$$

and $\|.\|$ is the maximum norm on $\mathbb{R}^{|\mathcal{S}|}$. We aim to show that $D_s^N(T) \to 0$ in probability for each $s \in \mathcal{S}$; once we have that, we have

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \le \max_{s \in \mathcal{S}} D_s^N(t) + ZR \int_0^t \|\bar{\mathbf{x}}^N(u) - \mathbf{v}(u)\| \mathrm{d}u \qquad (4)$$

and an application of Grönwall's lemma ([12], page 498) readily yields

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \le \max_{s \in \mathcal{S}} D_s^N(T) \exp(ZRT),$$

proving the theorem.

It now remains to show that for each $s \in \mathcal{S}$, $D_s^N(T) \to 0$ in probability as $N \to \infty$. To see this note that:

$$D_s^N(T) \le |v_s^0 - \bar{x}_s^N(0)| + \sum_{c \in \mathcal{C}} \frac{|l_s^c|}{N} \sup_{t \in [0,T]} \left| P_c \left(\int_0^t r_c(\bar{\mathbf{x}}^N(u)) \, \mathrm{d}u \right) - \int_0^t r_c(\bar{\mathbf{x}}(u)) \, \mathrm{d}u \right|$$

$$+ \sum_{c \in \mathcal{C}} \sum_{e \in \mathcal{E}} \frac{Y_{c,e}}{N} \sup_{t \in [0,T]} \left| \int_0^t \mathbf{1}_{\left\{ T^e_{J_c^N(u)} \le t-u \right\}} \, \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u) r_c^N(\bar{\mathbf{x}}(u)) \, \mathrm{d}u \right|$$

where $Y_{c,e} := \left| \sum_{s' \in \mathcal{A}_e} p_e(s',s) l_{s'}^c - \mathbf{1}_{\{s \in \mathcal{A}_e\}} l_s^c \right|$, using the shorthand $J_c^N(u) := P_c \left(\int_0^u r_c^N(\mathbf{x}^N(z)) \, \mathrm{d}z \right)$.

The first term converges per our assumptions; we argue that the second and third terms on the right-hand side converge almost surely. The second term is handled in the following lemma.

Lemma 1. *For any $c \in \mathcal{C}$*

$$\frac{1}{N} \sup_{t \in [0,T]} \left| P_c \left(\int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right) - \int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right| \to 0$$

almost surely as $N \to \infty$.

Proof. By the Lipschitz-condition, $0 \le \int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \le RTN$ and thus

$$\frac{1}{N} \sup_{t \in [0,T]} \left| P_c \left(\int_0^t r_c^N(\mathbf{x}^N(u)) \, \mathrm{d}u \right) - \int_0^t r_c^N(\mathbf{x}^N(u)) \, \mathrm{d}u \right| \le$$

$$\le \frac{1}{N} \sup_{s \in [0,RT]} |P_c(Ns) - Ns| ,$$

which goes to 0 almost surely by the functional strong law of large numbers (FSLLN) for the Poisson process ([22], Section 3.2).

To handle the third term we note that:

$$\left| \int_0^t \mathbf{1}_{\left\{ T^e_{J_c^N(u)} \le t-u \right\}} \, \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u) r_c^N(\mathbf{x}^N(u)) \, \mathrm{d}u \right| \le$$

$$\left| \int_0^t \mathbf{1}_{\left\{ T^e_{J_c^N(u)} \le t-u \right\}} \, \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u) \, \mathrm{d}J_c^N(u) \right|$$

$$+ \left| \int_0^t F_e(t-u) \, \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u) r_c^N(\mathbf{x}^N(u)) \, \mathrm{d}u \right|$$

($Y_{c,e}$'s are constants not depending on N and t, and as such, there is no need to carry them around.) The two terms on the right hand side require tools different enough to separate them into Lemmas 2 and 3. The proof of Lemma 2 is essentially a consequence of the FSLLN for the Poisson process, while the heart of the proof of Lemma 3 is a probability concentration (or large deviation) theorem (*Azuma's inequality*).

We have some more preparations first. We already have that

$$\sup_{t \in [0,T]} \frac{1}{N} \left| P_c \left(\int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right) - \int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right| \to 0$$

almost surely as $N \to \infty$. As a direct consequence of this, we also have

$$\sup_{s,t \in [0,T]} \frac{1}{N} \left| P_c \left(\int_s^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right) - \int_s^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \right| \to 0$$

almost surely since

$$\sup_{s,t \in [0,T]} \left| \int_s^t \cdot \right| = \sup_{s,t \in [0,T]} \left| \int_0^t \cdot - \int_0^s \cdot \right| \le 2 \sup_{t \in [0,T]} \left| \int_0^t \cdot \right|$$

Also as a preparation, we have

$$\sup_{t\in[0,T]} \frac{1}{N} \int_0^t r_c^N(\mathbf{x}^N(u)) \mathrm{d}u \le \sup_{t\in[0,T]} \frac{1}{N} \int_0^t R\|\mathbf{x}^N(u)\| \mathrm{d}u \le \frac{1}{N} \sup_{t\in[0,T]} NRt = RT$$

independent of N, again using $\|\mathbf{x}^N\| \le N$ and $r_c^N(\mathbf{x}) \le R\|\mathbf{x}\| \ \forall \mathbf{x} \in \mathcal{X}^N$. Lemma 1 then also implies $\frac{1}{N} \int_0^t \mathrm{d}J_c^N(u) \le RT + \varepsilon_N$, where $\varepsilon_N \to 0$ almost surely as $N \to \infty$.

Lemma 2.

$$\sup_{t\in[0,T]} \frac{1}{N} \left| \int_0^t F_e(t-u)\mathrm{d}J_c^N(u) - \int_0^t F_e(t-u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u \right| \to 0$$

almost surely as $N \to \infty$.

Proof. Let $\varepsilon > 0$ be fixed. Write

$$F_e(t-u) = g_{e,t,\varepsilon}(u) + h_{e,t,\varepsilon}(u),$$

where $g = g_{e,t,\varepsilon}$ is a piecewise constant function with $0 \le g(u) \le 1$ and $\|h\|_\infty \le \varepsilon$. Their exact definition is as follows. Take the $\varepsilon, 2\varepsilon, \dots$ quantiles of $F_e(t-u)$ (recall $F_e(t-u)$ is a nonincreasing function between 0 and 1); that is, let $u_k = \inf\{u : F(t-u) \le k\varepsilon\}$. Some of these u_k's may be equal if F has discontinuities. The number of distinct quantiles is certainly no more than $\lceil \varepsilon^{-1} \rceil$, independent of N and t.

Let g be the piecewise constant function

$$g(u) = F_e(t-u_k) \quad \text{if} \quad u \in (u_{k-1}, u_k],$$

so $g(u) \le F_e(t-u)$. The choice of u_k's guarantees that $h(u) = F_e(t-u) - g(u) \le \varepsilon$.

Then we can write

$$\frac{1}{N} \left| \int_0^t F_e(t-u)\mathrm{d}J_c^N(u) - \int_0^t F_e(t-u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u \right| \le$$
$$\frac{1}{N} \left| \int_0^t g(u)\mathrm{d}J_c^N(u) - \int_0^t g(u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u \right| +$$
$$\frac{1}{N} \left| \int_0^t h(u)\mathrm{d}J_c^N(u) - \int_0^t h(u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u \right|$$

Since g is piecewise constant,

$$\frac{1}{N}\left|\int_0^t g\,\mathrm{d}J_c^N(u) - \int_0^t g(u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| =$$

$$\frac{1}{N}\left|\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil} g(u_k)\left(J_c^N(u_k) - J_c^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right)\right| \leq$$

$$\frac{1}{N}\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil} \left|g(u_k)\left(J_c^N(u_k) - J_c^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right)\right| \leq$$

$$\frac{1}{N}\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil} \left|J_c^N(u_k) - J_c^N(u_{k-1}) - \int_{u_{k-1}}^{u_k} r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \leq$$

$$\sum_{k=1}^{\lceil\varepsilon^{-1}\rceil} \sup_{s,t\in[0,T]} \frac{1}{N}\left|P_c\left(\int_s^t r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right) - \int_s^t r_c^N(\mathbf{x}^N(u))\right| =$$

$$\lceil\varepsilon^{-1}\rceil \cdot \sup_{s,t\in[0,T]} \frac{1}{N}\left|P_c\left(\int_s^t r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right) - \int_s^t r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \to 0$$

almost surely as $N \to \infty$ since ε is independent of N.

Since $\|h\|_\infty \leq \varepsilon$, we have

$$\frac{1}{N}\left|\int_0^t h(u)\mathrm{d}J_c^N(u) - \int_0^t h(u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \leq$$

$$\frac{1}{N}\left|\int_0^t h(u)\mathrm{d}J_c^N(u)\right| + \frac{1}{N}\left|\int_0^t h(u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \leq$$

$$\frac{\varepsilon}{N}\left|\int_0^t \mathrm{d}J_c^N(u)\right| + \frac{\varepsilon}{N}\left|\int_0^t r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \leq \varepsilon(2RT + \varepsilon_N),$$

independent of t (with $\varepsilon_N \to 0$ almost surely as $N \to \infty$).

Letting $\varepsilon \to 0$ proves

$$\sup_{t\in[0,T]}\left|\int_0^t F_e(t-u)\mathrm{d}J_c^N(u) - \int_0^t F_e(t-u)r_c^N(\mathbf{x}^N(u))\mathrm{d}u\right| \to 0$$

almost surely as $N \to \infty$.

Lemma 3.

$$\sup_{t\in[0,T]} \frac{1}{N}\left|\int_0^t \mathbf{1}_{\left\{T_{J_c^N(u)}^e \leq t-u\right\}}\mathrm{d}J_c^N(u) - \int_0^t F_e(t-u)\mathrm{d}J_c^N(u)\right| \to 0$$

almost surely as $N \to \infty$.

Proof. Let ε be fixed. Also fix t for now. We want to prove

$$\mathbf{P}\left(\left|\int_0^t \mathbf{1}_{\left\{T^e_{J^N_c(u)}\leq t-u\right\}}\mathrm{d}J^N_c(u) - \int_0^t F_e(t-u)\mathrm{d}J^N_c(u)\right| > \varepsilon\right)$$

is exponentially small in N via Azuma's inequality [8, 2]. Once we have that, we can apply Borel–Cantelli lemma (see e.g. [10] Chapter 2.3) to conclude that for any fixed ϵ, the above event happens only finitely many times, which is equivalent to almost sure convergence to 0. To apply Azuma, we need to write the above integral as a martingale with bounded increments. The measure $\mathrm{d}J^N_c(u)$ is concentrated on points u where P_c has an arrival at $\int_0^u r^N_c(\mathbf{x}^N(z)\mathrm{d}z$. Let we denote these points by $u_1, u_2, \dots$. The integral only has contributions from these points; it is natural to write (using a slightly different notation)

$$S_l := \left(\mathbf{1}_{\{T^e_1\leq t-u_1\}} - F_e(t-u_1)\right) + \cdots + \left(\mathbf{1}_{\{T^e_l\leq t-u_l\}} - F_e(t-u_l)\right)$$
$$M_N := P_c\left(\int_0^t r^N_c(\mathbf{x}^N(z))\mathrm{d}z\right)$$

so that

$$\int_0^t \mathbf{1}_{\left\{T^e_{J^N_c(u)}\leq t-u\right\}}\mathrm{d}J^N_c(u) - \int_0^t F_e(t-u)\mathrm{d}J^N_c(u) = S_{M_N}.$$

We first resolve the difficulty that M_N is in fact random.

$$\mathbf{P}\left(\frac{1}{N}\left|\int_0^t \mathbf{1}_{\left\{T^e_{J^N_c(u)}\leq t-u\right\}}\mathrm{d}J^N_c(u) - \int_0^t F_e(t-u)\mathrm{d}J^N_c(u)\right| > \varepsilon\right) =$$
$$\mathbf{P}\left(\left|\frac{S_{M_N}}{N}\right| > \varepsilon\right) = \sum_{l=0}^{\infty}\mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon, M_N = l\right) \leq$$
$$\sum_{l=0}^{2RTN}\mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right) + \sum_{2RTN+1}^{\infty}\mathbf{P}(M_N = l) =$$
$$\sum_{l=0}^{2RTN}\mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right) + \mathbf{P}(M_N > 2RTN).$$

The sum was cut at $2RTN$ because M_N is stochastically dominated by a Poisson distribution with parameter RTN, so $\mathbf{P}(M_N > 2RTN)$ is exponentially small due to Cramér's large deviation theorem (see e.g. Theorem II.4.1 in [11]):

$$\mathbf{P}(M_N > 2RTN) \leq e^{-RTN(2\ln 2 - 1)}.$$

(The Cramér rate function of the Poisson-distribution with parameter λ is $I(x) = x\ln(x/\lambda) - x + \lambda$.)

To apply Azuma to each of the terms $\mathbf{P}\left(\left|\frac{S_l}{N}\right| > \varepsilon\right)$, we also need to check that S_l is indeed a martingale with bounded increments. To set it up properly

as a martingale, note that $\{u_l\}$ is an increasing sequence of stopping times, so the filtration $\{\mathcal{F}_l\}$ is well-defined; $\mathcal{F}_l$ contains all the information known up to time u_l, including the values of all of the non-Markovian clocks that started by the time u_l.

S_l has bounded increments, since

$$|\mathbf{1}_{\{T_l^e \leq t-u_l\}} - F_e(t-u_l)| \leq 1.$$

The last step to apply Azuma is that we need to check that S_l is a martingale with respect to $\mathcal{F}_l$. It is clearly adapted, and

$$\mathbf{E}(\mathbf{1}_{\{T_{l+1}^e \leq t-u_{l+1}\}}|\mathcal{F}_l) = \mathbf{E}(\mathbf{E}(\mathbf{1}_{\{T_{l+1}^e \leq t-u_{l+1}\}}|\mathcal{F}_l, u_{l+1})|\mathcal{F}_l) =$$

$$\mathbf{E}(\mathbf{P}(T_{l+1}^e \leq t-u_{l+1})|\mathcal{F}_l, u_{l+1})|\mathcal{F}_l) = \mathbf{E}(F_e(t-u_{l+1})|\mathcal{F}_l)$$

shows that it is a martingale as well. (In the last step, we used the fact that u_{l+1} is measurable with respect to $\sigma\{\mathcal{F}_l \cup \{u_{l+1}\}\}$ while T_{l+1}^e is independent from it.)

We have everything assembled to apply Azuma's inequality:

$$\sum_{l=0}^{2RTN} \mathbf{P}\left(\left|\frac{S_l - \mathbf{E}(S_l)}{N}\right| > \varepsilon\right) \leq \sum_{l=0}^{2RTN} 2e^{-\frac{2\varepsilon^2 N^2}{l}} \leq$$

$$\leq 2RTN \cdot 2e^{-\frac{2\varepsilon^2 N^2}{2RTN}} = 4RTNe^{-\frac{\varepsilon^2 N}{RT}}.$$

In the last inequality, we estimated each term in the sum by the largest one, which is for $l = 2RN$.

The estimate obtained is

$$\mathbf{P}\left(\frac{1}{N}\left|\int_0^t \mathbf{1}_{\left\{T_{J_c^N(u)}^e \leq t-u\right\}} \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u)\mathrm{d}J_c^N(u)\right| > \varepsilon\right) \leq$$

$$4RTNe^{-\frac{\varepsilon^2 N}{RT}} + e^{-RTN(2\ln 2-1)}.$$

Remember that t was fixed; we need to upgrade this estimate into an estimate that is valid for $\sup_{t \in [0,T]}(.)$ before applying Borel–Cantelli lemma. We do this by partitioning the interval $[0,T]$ into N subintervals uniformly, and then controlling what happens *at* the partition points and *between* the partition points separately. For the former, we apply the previous estimate. Let

$$t_i := \frac{iT}{N}, \; i = 0, 1, \ldots N,$$

then

$$\mathbf{P}\left(\max_{0 \leq i \leq N} \frac{1}{N}\left|\int_0^{t_i} \mathbf{1}_{\left\{T_{J_c^N(u)}^e \leq t-u\right\}} \mathrm{d}J_c^N(u) - \int_0^{t_i} F_e(t-u)\mathrm{d}J_c^N(u)\right| > \varepsilon\right) \leq$$

$$(N+1)\left(4RTNe^{-\frac{\varepsilon^2 N}{RT}} + e^{-RTN(2\ln 2-1)}\right),$$

which is still summable.

Now we turn our attention to the intervals $[t_i, t_{i+1}]$. Since

$$\int_0^t \mathbf{1}_{\left\{T^e_{J_c^N(u)} \leq t-u\right\}} \mathrm{d}J_c^N(u) \quad \text{and} \quad \int_0^t F_e(t-u)\mathrm{d}J_c^N(u)$$

are both increasing in t, we only have to check that neither of them increases by more than εN over an interval $[t_i, t_{i+1}]$.

Let i be fixed. We handle the two integrals separately. First, for

$$\int_0^t F_e(t-u)\mathrm{d}J_c^N(u),$$

we have

$$\int_0^{t_{i+1}} F_e(t_{i+1}-u)\mathrm{d}J_c^N(u) - \int_0^{t_i} F_e(t_i-u)\mathrm{d}J_c^N(u) =$$

$$\int_0^{t_i} F_e(t_{i+1}-u) - F_e(t_i-u)\mathrm{d}J_c^N(u) + \int_{t_i}^{t_{i+1}} F_e(t_{i+1}-u)\mathrm{d}J_c^N(u) \leq$$

$$\int_0^{t_i} F_e(t_{i+1}-u) - F_e(t_i-u)\mathrm{d}J_c^N(u) + \int_{t_i}^{t_{i+1}} 1\mathrm{d}J_c^N(u).$$

The second term is equal to $J_c^N(t_{i+1}) - J_c^N(t_i)$, e.g. the number of arrivals of P_c in the interval $[t_i, t_{i+1}]$. By the Lipschitz-condition, this is stochastically dominated from above by $Z \sim \text{Poisson}(RT)$ given that the length of the interval is T/N, and thus

$$\mathbf{P}\left(\frac{1}{N}\int_{t_i}^{t_{i+1}} F_e(t_{i+1}-u)\mathrm{d}J_c^N(u) > \varepsilon\right) \leq \mathbf{P}\left(\frac{Z}{N} > \varepsilon\right) = \mathbf{P}\left(\frac{Z}{\varepsilon} > N\right).$$

Note that the right hand side is summable in N, its sum being equal to the expectation of $\left\lceil \frac{Z}{\varepsilon} \right\rceil$.

To estimate the other term, note that

$$u \in [t_{l-1}, t_l] \implies F_e(t_{i+1}-u) - F_e(t_i-u) \leq F_e(t_{i+1}-t_{l-1}) - F_e(t_i - t_l) =$$
$$F_e(t_{i+1}-t_{l-1}) - F_e(t_{i+1}-t_l) + F_e(t_{i+1}-t_l) - F_e(t_i-t_l),$$

which gives

$$\int_0^{t_i} F_e(t_{i+1}-u) - F_e(t_i-u)\mathrm{d}J_c^N(u) =$$

$$\sum_{l=1}^i \int_{t_{l-1}}^{t_l} F_e(t_{i+1}-u) - F_e(t_i-u)\mathrm{d}J_c^N(u) \leq$$

$$\sum_{l=1}^i \int_{t_{l-1}}^{t_l} F_e(t_{i+1}-t_{l-1}) - F_e(t_i-t_l)\mathrm{d}J_c^N(u) =$$

$$\sum_{l=1}^i (F_e(t_{i+1}-t_{l-1}) - F_e(t_i-t_l))(J_c^N(t_l) - J_c^N(t_{l-1})).$$

We use two things here: the fact that $(J_c^N(t_l) - J_c^N(t_{l-1}))$ is stochastically dominated by $\mathrm{Poisson}(RT)$ and the fact that the sum

$$\sum_{l=1}^{i}(F_e(t_{i+1} - t_{l-1}) - F_e(t_i - t_l)) = \sum_{l=1}^{i} F_e(t_{i-l+2}) - F_e(t_{i-l}) =$$
$$F_e(t_{i+1}) + F_e(t_i) - F_e(1) - F_e(0) \le 2$$

is telescopic. This means that the whole sum can be stochastically dominated by $\mathrm{Poisson}(2RT)$ (note that the number of clocks starting at each interval is not independent, but because of the Lipschitz-condition, we may still use independent Poisson variables when stochastically dominating the sum). Using the notation $Z \sim \mathrm{Poisson}(RT)$ again, we get that

$$\sum_{N=1}^{\infty} \mathbf{P}\left(\frac{2Z}{N} > \varepsilon\right) = \sum_{N=1}^{\infty} \mathbf{P}\left(\frac{2Z}{\varepsilon} > N\right) \le \frac{2RT}{\varepsilon} + 1.$$

(In fact, $\mathbf{P}\left(\frac{2Z}{\varepsilon} > N\right)$ goes to 0 superexponentially in N.)

The last term to estimate is the increment of

$$\int_0^t \mathbf{1}_{\left\{T_{J_c^N(u)}^e \le t-u\right\}} \mathrm{d}J_c^N(u).$$

between t_i and t_{i+1}, e.g. the number of clocks expiring between t_i and t_{i+1}.

Partition the clocks according to which interval $[t_{l-1}, t_l]$ they started in. The number of clocks starting in $[t_{l-1}, t_l]$ is stochastically dominated by $Z \sim \mathrm{Poisson}(RT)$ by the Lipschitz-condition, and for each such clock, the probability that it goes off in $[t_i, t_{i+1}]$ is less than or equal to $F_e(t_{i+1}) - F_e(t_{l-1})$. This implies that the number of the clocks starting in $[t_{l-1}, t_l]$ and going off in $[t_i, t_{i+1}]$ is stochastically dominated by $W_{i,l} \sim \mathrm{Poisson}(RT(F_e(t_{i+1}) - F_e(t_{l-1})))$. The total number of clocks going off in $[t_i, t_{i+1}]$ is stochastically dominated by $\mathrm{Poisson}(RT\sum_{l=1}^{i}(F_e(t_{i+1}) - F_e(t_{l-1})))$, where the familiar telescopic sum appears in the parameter. (Once again, the Lipschitz-condition was used implicitly.) So the total number of clocks going off in $[t_i, t_{i+1}]$ is stochastically dominated by $\mathrm{Poisson}(2RT)$, which means we arrive at the also familiar $\mathbf{P}\left(\frac{2Z}{\varepsilon} > N\right)$ value, which we already examined and proved to be summable in N.

Putting it altogether, we get that

$$\mathbf{P}\left(\sup_{t \in [0,T]} \frac{1}{N} \left| \int_0^t \mathbf{1}_{\left\{T_{J_c^N(u)}^e \le t-u\right\}} \mathrm{d}J_c^N(u) - \int_0^t F_e(t-u) \mathrm{d}J_c^N(u) \right| > \varepsilon\right) \le C_{N,\varepsilon}$$

where

$$\sum_{N=1}^{\infty} C_{N,\varepsilon} < \infty,$$

so the Borel–Cantelli lemma gives almost sure convergence as $N \to \infty$.

With Lemmas 1-3 finished, the proof of Theorem 1 is complete.

Theorem 1 proves mean-field convergence in the transient case. The question of stationary regime is quite different; for some remarks on the stationary regime, we refer to Section V of [16].

4 Example

In this section, we derive the system of DDEs as defined in the previous section for a simple example model of a peer-to-peer software update process. For a more detailed discussion of a peer-to-peer update example, we refer to [16], where essentially the same model was introduced, albeit with deterministic delays instead of general ones.

We consider two general types of nodes in this model which we term *old* and *updated*. Old nodes are those running an old software version and new nodes are those which have been updated to a new version. Nodes alternate between being *on* and *off*; when an old node turns on, it searches for an update in peer-to-peer fashion, with the probability of successfully finding an update being proportional to the number of nodes already updated. If it does not find an update, it gives up after a timeout. After that, it stays on for some time and then eventually turns off. New nodes do not search for updates, just alternate between on and off. We assume that the off time of a node is random and Pareto-distributed. So, nodes have five possible local states: updated nodes can be on and off, which we denote by a and b, respectively. Old nodes can be on (c), off (e) or in a state representing an old node which is on but has given up seeking updates (d). In the notation of Section 2.1, the set of local states is thus $\mathcal{S} := \{a, b, c, d, e\}$. The local behaviour of a node is depicted in Figure 1.

In this example, we consider all transitions to be Markovian except for the transitions bringing nodes from their off state into their on state, which have density function $f(s)$. Formally, there are two event clocks t_0 and t_1 with $\mathcal{A}_{t_0} := \{e\}$, $d_{t_0} := \eta$, $p_{t_0}(e, c) = 1$, $\mathcal{A}_{t_1} := \{b\}$. $d_{t_1} := \eta$ and $p_{t_1}(b, a) := 1$.

The DDEs corresponding to this model are:

$$\dot{v}_a(t) = -\rho v_a(t) + \beta v_c(t) v_a(t) + \rho \int_0^t v_a(t - s) f(s) \, \mathrm{d}s$$

$$\dot{v}_b(t) = -\rho \int_0^t v_a(t - s) f(s) \, \mathrm{d}s + \rho v_a(t)$$

$$\dot{v}_c(t) = -\rho v_c(t) - \beta v_c(t) v_a(t) - \kappa v_c(t)$$
$$\qquad + \rho \int_0^t (v_d(t - s) + v_c(t - s)) f(s) \, \mathrm{d}s$$

$$\dot{v}_d(t) = -\rho v_d(t) + \kappa v_c(t)$$

$$\dot{v}_e(t) = -\rho \int_0^t (v_d(t - s) + v_c(t - s)) f(s) \, \mathrm{d}s + \rho v_d(t) + \rho v_c(t)$$
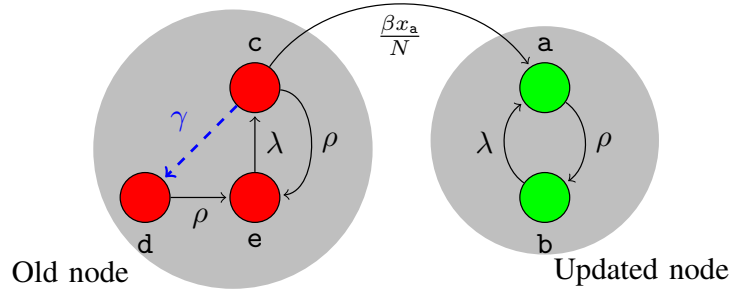
Fig. 1. Representation of the behaviour of a single node in the delay-only software update model.

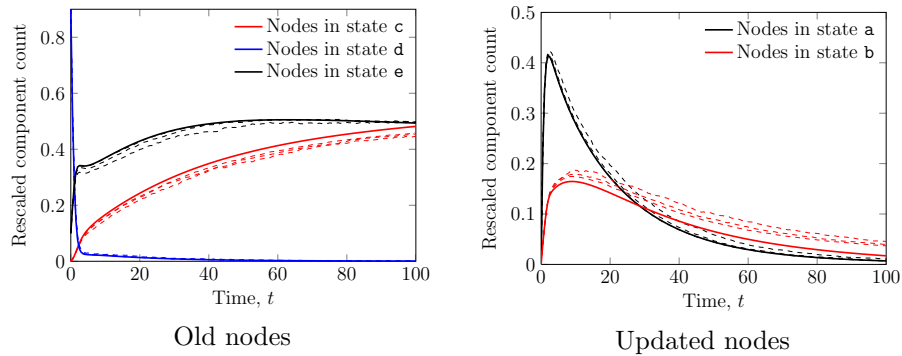Old nodes   Updated nodes

Fig. 2. Delay-only software update model rescaled DDE approximation (solid lines) compared with rescaled actual means for $N = 20$, 50 and 200 (dashed lines). Initial component proportions are $(0.1, 0, 0.9, 0, 0)$ and parameters are $\beta = 2.0$, $\rho = 0.1$, $\kappa = 0.67$ with $f(s)$ a Pareto density with scale parameter 1.5 and shape parameter 0.9.

This system of DDEs can be integrated numerically by adapting existing ODE solvers or specialised DDE routines such as the `dde23` routine in `MATLAB`®. The solution of these DDEs for one set of parameters is shown in Figure 2 compared with the corresponding rescaled component-count expectations as computed by many stochastic simulation replications. We observe that the means do appear to converge to the mean-field solutions in line with Theorem 1.

Acknowledgment

References

1. J. Anselmi and M. Verloop. Energy-aware capacity scaling in virtualized environments with performance guarantees. *Perf. Eval.*, 68(11):1207–1221, 2011.

2. K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

3. Roberto Barbuti, Giulio Caravagna, Andrea Maggiolo-Schettini, and Paolo Milazzo. Delay stochastic simulation of biological systems: a purely delayed approach. *Transactions on Computational Systems Biology XIII*, 6575:61–84, January 2011.

4. Michel Benaïm and Jean-Yves Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, November 2008.

5. Luca Bortolussi and Jane Hillston. Fluid approximation of ctmc with deterministic delays. *Int. Conf. on Quantitative Evaluation of Systems*, pages 53–62, 2012.

6. Giulio Caravagna and Jane Hillston. Bio-PEPAd: A non-Markovian extension of Bio-PEPA. *Theoretical Computer Science*, 419:26–49, 2012.

7. Hoon Choi, Vidyadhar Kulkarni, Kishor Trivedi, and Marco Ajmone Marsan. Transient analysis of deterministic and stochastic petri nets. In *Application and Theory of Petri Nets*, volume 691, pages 166–185. Springer Berlin / Heidelberg, 1993.

8. Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127, 2006.

9. David R. Cox. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(03):433–441, July 1955.

10. R. Durrett. *Probability: Theory and Examples*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2010.

11. R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Classics in Mathematics. Springer, 2005.

12. Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 2005.

13. Nicolas Gast and Gaujal Bruno. A mean field model of work stealing in large-scale systems. *SIGMETRICS Perform. Eval. Rev.*, 38(1):13–24, 2010.

14. P. W. Glynn. A GSMP formalism for discrete event systems. *Proceedings of the IEEE*, 77(1):14–23, 1989.

15. Peter G. Harrison and Ben Strulo. SPADES - a process algebra for discrete event simulation. *Journal of Logic and Computation*, 10(1):3–42, 2000.

16. R.A. Hayden. Mean field for performance models with deterministically-timed transitions. In *Quantitative Evaluation of Systems (QEST), 2012 Ninth International Conference on*, pages 63–73, Sept 2012.

17. Richard A. Hayden and Jeremy T. Bradley. A fluid analysis framework for a markovian process algebra. *Theoretical Computer Science*, 411:2260–2297, 2010.

18. Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978.

19. S. Maset. Numerical solution of retarded functional differential equations as abstract cauchy problems. *J. Comput. Appl. Math.*, 161(2):259–282, 2003.

20. K. Matthes. Zur theorie der bedienungsprozesse. In *3rd Prague Conf. on Inf. Theory, Statistical Decision Functions and Random Processes*, pages 512–528, 1962.

21. Robert Schlicht and Gerhard Winkler. A delay stochastic process with applications in molecular biology. *Journal of Mathematical Biology*, 57(5):613–48, 2008.

22. Ward Whitt. Internet supplement to Stochastic-Process Limits, 2002.