Mean-field analysis of Markov models with reward feedback

Anton Stefanek, Richard A. Hayden, Mark Mac Gonagle, Jeremy T. Bradley

Department of Computing, Imperial College London, London SW7 2BZ
{as1005,rh,jb}@doc.ic.ac.uk

Abstract. We extend the *population continuous time Markov chain* formalism so that the state space is augmented with continuous variables accumulated over time as functions of component populations. System feedback can be expressed using accumulations that in turn can influence the Markov chain behaviour via functional transition rates. We show how to obtain mean-field differential equations capturing means and higher-order moments of the discrete populations and continuous accumulation variables. We also provide first- and second-order convergence results and suggest a novel normal moment closure that can greatly improve the accuracy of means and higher moments.
We demonstrate how such a framework is suitable for modelling feedback from globally-accumulated quantities such as energy consumption, cost or temperature. Finally, we present a worked example modelling a hypothetical heterogeneous computing cluster and its interaction with air conditioning units.

1 Introduction

The behaviour of large computing clusters is often controlled by feedback from various accumulated continuous quantities, such as temperature, energy consumption or total cost. For example, an air-conditioning controller in a server farm will react to the ambient temperature. At the same time, sophisticated thermally-aware schedulers [23] can use temperature sensors to regulate server operation and thus indirectly environmental temperature, creating a feedback loop.

Stochastic models of computing clusters will typically be very large and thus, due to state-space explosion, will often lie outside the capabilities of traditional performance analysis. However, the nature of these systems, consisting of many identically-behaving cooperating components is suitable for mean-field type analyses [e.g. 7, 10, 12]. Mean-field techniques have recently been extended to capture certain accumulated rewards [21]. We show how to further adapt this approach to allow modelling of feedback between the system model and the generated accumulated quantities.

In Section 2, we extend the discrete state space of a Markov population model with accumulated variables governed by integral equations. The accumulation functions can involve component populations and the discrete transition rate functions can depend on the accumulated variables, thus allowing feedback loops. In Section 2.2 we extend the mean-field techniques to analyse means and higher moments of component populations and accumulated variables. Section 3 justifies this approach by proving convergence to the solution of the mean field equations as the scale of the system increases. In Section 3.2, based on second-order convergence to a Gaussian process, we introduce a moment closure that improves the accuracy of the approximation when the rates contain occurrences of the minimum or maximum functions, common situations when modelling computer systems, for example, the process algebra *PEPA* [10] or *stochastic Petri nets* [19]. We demonstrate the techniques on a larger example of a heterogeneous computing cluster with controlled temperature in Section 4.

1.1 Related work

Capturing the feedback interaction between process-based agents and continuously varying physical properties of a system falls in the realm of hybrid system modelling. In the field of performance analysis, an initial example of this would have been in FSPNs or fluid stochastic Petri nets [13] where fluid places are used to capture continuously varying quantities. Discrete Petri net behaviour was in turn governed by the level of a given fluid place. FSPNs could be simulated but were restricted to only one or two fluid places in practice.

A detailed comparative study of hybrid process algebras can be found in [15]. A common feature in each of these hybrid process algebras is the expression of continuous evolution via the embedding of ordinary differential equations in the process model itself. In contrast, Bortollussi *et al.* [3] have developed stochastic HYPE, a process formalism that generates both discrete and continuously varying dynamics from the semantics of the process model alone.

In this paper we present a process mechanism that expresses feedback control as a result of accumulated reward variables in Markov population models. Analysis of Markov Reward Models (MRMs) [18, 24] is if anything more computationally demanding than analysis of plain Markov models. In Stefanek *et al.* [21], we showed how a fluid approximation could be constructed for a class of MRMs, but we had no way of providing a feedback mechanism based on those reward values.

In this paper we show how accumulated reward variables can be used to influence transition guards and rates in a large Markov model. We have not endeavoured to express the continuously varying rewards and variables in a process-style language, as in [3]. Instead we have focused on showing convergence between the ODE solution of the resulting population CTMC model with reward accumulations (aPCTMC) and simulations of the underlying stochastic process. Further, we consider higher moments of rewards with feedback, something we believe has not been presented before for hybrid systems of such scale.

2 Markov population models with accumulations

In this section, we define an extension of a continuous-time Markov population process (PCTMC). The state space of a PCTMC consists of vectors $\mathbf{x} \in \mathbb{Z}_+^n$

of integer-valued populations, where $\mathbf{x}_0$ is the initial configuration. Transitions of the Markov chain are defined via a set of transition classes $\mathcal{C}$. Each class $c \in \mathcal{C}$ specifies a difference vector $l_c \in \mathbb{Z}^n$ between the populations before and after such a transition occurs and a rate function $r_c \colon \mathbb{Z}_+^n \to \mathbb{R}_+$ defining the infinitesimal rate of transitions of class c as a function of populations in the given state.

We illustrate the following definitions on a PCTMC representing a simple client/system. We use the PEPA stochastic process algebra [12] to define a PCTMC:

$$\begin{array}{ll} \mathit{Client}_0 \stackrel{\mathit{def}}{=} (\mathit{data}, r_{\mathit{data}}).\mathit{Client}_1 & \mathit{Server}_0 \stackrel{\mathit{def}}{=} (\mathit{data}, r_{\mathit{data}}).\mathit{Server}_1 \\ \mathit{Client}_1 \stackrel{\mathit{def}}{=} (\mathit{task}, r_{\mathit{task}}).\mathit{Client}_0 & \mathit{Server}_1 \stackrel{\mathit{def}}{=} (\mathit{reset}, r_{\mathit{reset}}).\mathit{Server}_0 \end{array}$$

$$\mathbf{Clients}\{\mathit{Client}_0[N_C]\} \underset{\mathit{data}}{\bowtie} \mathbf{Servers}\{\mathit{Server}_0[N_S]\}$$

Here, the discrete state space consists of numerical vectors $\mathbf{x} = (C_0, C_1, S_0, S_1) \in \mathbb{Z}_+^4$ and the initial state is $(N_C, 0, N_S, 0)$, keeping track of the populations of clients and servers in their respective states. There are 3 transition classes in this model – one corresponding to the synchronised event where a client sends its data to a server and two independent events where the client and the server reset to their initial states. According to the PEPA operational semantics, the respective change vectors and rate functions are $l_1 = (-1, -1, 1, 1)$ with $r_1(\mathbf{x}) = \min(C_0, S_0) r_{\mathit{data}}$, $l_2 = (1, -1, 0, 0)$ with $r_2(\mathbf{x}) = C_1 \cdot r_{\mathit{task}}$ and $l_3 = (0, 0, 1, -1)$ with $r_3(\mathbf{x}) = S_1 \cdot r_{\mathit{reset}}$.

We augment the state space with a set of continuous variables governed by an auxiliary system of integral equations whose evolution may additionally depend on the discrete populations. The continuous variables can be used to track the evolution of associated quantities such as energy use or temperature. Furthermore, the rates of the Markovian evolution of the discrete populations may also depend on the value of these variables, thus allowing, for example, energy usage over time to feedback into the control of the system.

2.1 Definition

The state space of a *PCTMC with accumulations* (*aPCTMC*) is a subset of $\mathbb{Z}_+^n \times \mathbb{R}^m$ consisting of states $(\mathbf{x}, \mathbf{y})$, where $\mathbf{x} \in \mathbb{Z}_+^n$ captures the discrete populations and $\mathbf{y} \in \mathbb{R}^m$ captures the continuous accumulation variables. The discrete populations evolve as in traditional PCTMCs, that is, according to a set $\mathcal{C}$ of transition classes. The associated rate functions are extended onto the full state space, that is $r_c : \mathbb{Z}_+^n \times \mathbb{R}^m \to \mathbb{R}_+$ We denote the discrete-state component of the associated stochastic process by $\mathbf{X}(t)$ with its initial state given by $\mathbf{x}_0$.

The evolution of the continuous variables $\mathbf{Y}(t)$ is given by an integral equation of the form:

$$\mathbf{Y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{g}(\mathbf{X}(s), \mathbf{Y}(s))\, \mathrm{d}s \tag{1}$$

where $\mathbf{g} : \mathbb{Z}_+^n \times \mathbb{R}^m \to \mathbb{R}^m$ is an accumulation function and $\mathbf{y}_0$ is the initial state of the accumulation variables.

For example, in the client/server model we might wish to model generation of heat energy by servers when in the active state $Server_1$, resulting in an increase in the total energy in the server room. In order to model the heating–cooling process, we extend the discrete model also with air conditioning units:

$$Aircon_0 \stackrel{def}{=} (on, \lambda_{on}(t)).Aircon_1 \quad Aircon_1 \stackrel{def}{=} (off, \lambda_{off}(t)).Aircon_0$$

$$\Big(\mathbf{Clients}\{Client_0[N_C]\} \underset{data}{\bowtie} \mathbf{Servers}\{Server_0[N_S]\}\Big) \parallel \mathbf{Aircon}\{Aircon_0[N_A]\}$$

where the rates λ_{on} and λ_{off} are defined below. The active air conditioning units contribute to the cooling of the environment, by transferring heat energy out of the room. If we assume that the heat generation and cooling rates (r_{heat} and r_{cool}) are constant over time, the heat energy in the server room can be captured by an accumulated variable:

$$E(t) = E_0 + \int_0^t r_{heat} S_1(u) - r_{cool} A_1(u) \, \mathrm{d}u$$

where E_0 is the initial energy in the room.

We can introduce feedback into the system by making the air conditioning transition rates depend on the current temperature of the room. An approximate physical model for the temperature is:

$$T(t) = \frac{c}{v} E(t) \tag{2}$$

where c is a constant and v is the total volume of air in the room. One possible control policy for the air conditioning units might be: when the temperature is above a given threshold T_{thresh}, units switch on at some rate, otherwise active units switch off:

$$\lambda_{on}(t) = r_{on} \text{ if } T(t) > T_{thresh} \qquad \lambda_{off}(t) = r_{off} \text{ if } T(t) < T_{thresh}$$

and 0 otherwise.

In general, an *a*PCTMC process can be realised as a *piecewise deterministic Markov process (PDMP)* [5]. However, in order for the above construction to result in a uniquely well-defined PDMP on any finite interval of time, some regularity conditions are required. In particular, it is important that the possibility of infinitely many jumps of the discrete component in a finite period of time is prevented and also that the continuous component cannot grow unbounded in a finite period of time, that is, that it cannot explode. The following conditions are sufficient to achieve this, where $\mathcal{X} \subset \mathbb{Z}_+^n$ is defined to be the reachable state space of the discrete component:

1. There exist $A, B \in \mathbb{R}_+$ such that for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{y} \in \mathbb{R}^m$ and $c \in \mathcal{C}$:
$$\|\mathbf{g}(\mathbf{x}, \mathbf{y})\| \leq A(\|\mathbf{y}\| + 1) \quad \text{and} \quad r_c(\mathbf{x}, \mathbf{y}) \leq B(\|\mathbf{y}\| + 1)$$
2. The function $\mathbf{g}(\mathbf{x}, \cdot)\colon \mathbb{R}^m \to \mathbb{R}^m$ satisfies a local Lipschitz condition for each $\mathbf{x} \in \mathcal{X}$.
3. For each $c \in \mathcal{C}$, the function $r_c(\mathbf{x}, \cdot)\colon \mathbb{R}^m \to \mathbb{R}_+$ is measurable for each $\mathbf{x} \in \mathcal{X}$.

Assumption 2 guarantees that, between discrete jumps, the continuous component is defined uniquely and exists as long as it does not explode. In fact, the only way that the above construction will fail is if the continuous component explodes, since, otherwise, the maximal jump rate is bounded by assumption 1. However, if the continuous component does explode, say, at time t^*, then for any $t < t^*$, we have:

$$\|\mathbf{Y}(t)\| \leq \|\mathbf{y}_0\| + \int_0^t \|\mathbf{g}(\mathbf{X}(s), \mathbf{Y}(s))\| \, \mathrm{d}s \leq \|\mathbf{y}_0\| + At^* + \int_0^t A\|\mathbf{Y}(s))\| \, \mathrm{d}s$$

Applying a version of Grönwall's lemma [e.g. 6, Page 498] yields:

$$\|\mathbf{Y}(t)\| \leq [\|\mathbf{y}_0\| + At^*] \exp(tA)$$

This implies that $\mathbf{Y}(t)$ cannot explode at time t^* since it is continuous and bounded by $[\|\mathbf{y}_0\| + At^*] \exp(t^*A)$ for any $t < t^*$. Thus we have a contradiction and have shown that, subject to the assumptions above, our construction is well-defined on finite intervals of time.

Note that transition rates can be defined using a discontinuous indicator function without breaking any of the above assumptions, such as the rates $\lambda_{on}(t)$ and $\lambda_{off}(t)$ above. Therefore the client/server model defines a valid aPCTMC model.

2.2 Mean-field approximations

It is straightforward to extend simulation algorithms for CTMCs to realise traces of the evolution of the discrete and continuous state components of aPCTMC models. However, in the case of large models, simulation suffers from high computational costs.

We show how to extend the efficient mean-field (a.k.a. *fluid-analysis*) approach for the analysis of massive CTMC models [e.g. 10, 12, 25] to the case of aPCTMC models. Specifically, define $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ by $\mathbf{f}(\mathbf{x}, \mathbf{y}) := \sum_{c \in \mathcal{C}} r_c(\mathbf{x}, \mathbf{y}) \mathbf{l}_c$, for suitable real extensions of the functions r_c. Then an intuitive extension of the mean-field approach yields the following systems of integral equations:

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{f}(\mathbf{x}(s), \mathbf{y}(s)) \, \mathrm{d}s \qquad \mathbf{y}(t) = \mathbf{y}_0 + \int_0^t \mathbf{g}(\mathbf{x}(s), \mathbf{y}(s)) \, \mathrm{d}s \tag{3}$$

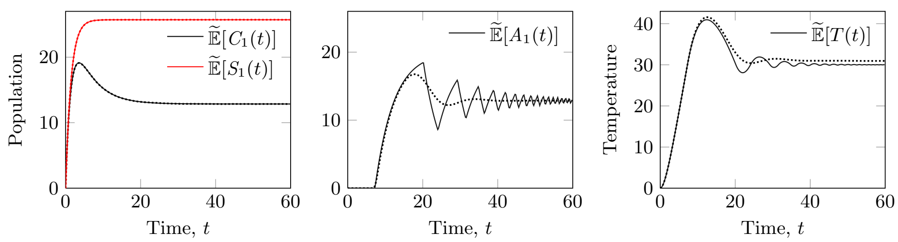
whose solutions can be interpreted as approximations to the means of the stochastic processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, respectively, or for sufficiently large populations, as approximations to individual traces of the stochastic processes.

For example, in the client/server model, we get equations such as:

$$s_0(t) = N_S + \int_0^t r_{reset} s_1(u) - r_{data} \min(c_0(u), s_0(u)) \, \mathrm{d}u$$

where we use lower case letters for the mean-field approximations of the respective population and accumulation processes.

In Section 3 we show that, in the limit of large populations, the traces of the processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ (and in particular the means $\mathbb{E}[\mathbf{X}(t)]$ and $\mathbb{E}[\mathbf{Y}(t)]$) converge to the mean-field solutions $\mathbf{x}(t)$ and $\mathbf{y}(t)$, respectively. Since we will usually be comparing means, we will adopt the notation $\widetilde{\mathbb{E}}[\mathbf{X}(t)]$ for $\mathbf{x}(t)$. For example, Figure 1 shows the numerical solutions to the mean-field model of Equation (3) as applied to the client/server model, compared to the estimates of the exact means sampled from 10^5 simulation runs of the stochastic process. In all figures in this paper, unless noted otherwise, the estimates from simulation are shown as dotted lines. Appendix A shows the specific values of parameters used to produce this figure and all the subsequent figures in this paper.



(a) Active clients/servers. (b) Active aircon. units. (c) Mean temperature.

Fig. 1: Approximation of mean component populations and accumulations in the client/server model.

2.3 Higher-order moments

In addition to approximations of means of populations and the accumulation variables, systems of equations approximating higher-order moments may also be derived by extending existing approaches [e.g. 9, 10, 21] for CTMCs. The joint process $(\mathbf{X}(t), \mathbf{Y}(t))$ is clearly Markovian with infinitesimal generator $\mathcal{A}$ defined on continuous and bounded functions $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ that are differentiable in the last m variables:

$$\begin{aligned}\mathcal{A}h(\mathbf{x}, \mathbf{y}) &:= \lim_{t\to 0} \frac{\mathbb{E}[h(\mathbf{X}(t), \mathbf{Y}(t))|(\mathbf{X}(0), \mathbf{Y}(0)) = (\mathbf{x}, \mathbf{y})] - h(\mathbf{x}, \mathbf{y})}{t} \\ &= \sum_{i=1}^{m} g_i(\mathbf{x}, \mathbf{y}) \frac{\partial h}{\partial y_i}(\mathbf{x}, \mathbf{y}) + \sum_{c\in\mathcal{C}} r_c(\mathbf{x}, \mathbf{y})[h(\mathbf{x} + \mathbf{l}_c, \mathbf{y}) - h(\mathbf{x}, \mathbf{y})]\end{aligned}$$

It thus follows by Dynkin's formula [e.g. 14, Lemma 17.21] that for $t \in \mathbb{R}_+$:

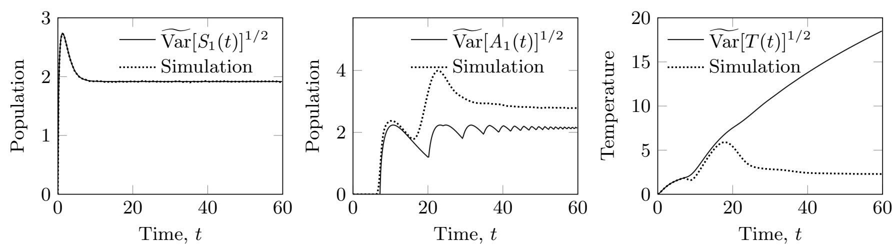
$$\mathbb{E}[h(\mathbf{X}(t), \mathbf{Y}(t))] = h(\mathbf{x}_0, \mathbf{y}_0) + \int_0^t \mathbb{E}[\mathcal{A}h(\mathbf{X}(s), \mathbf{Y}(s))]\,\mathrm{d}s \tag{4}$$

Equations for second-order moments can be obtained by choosing $h(\mathbf{x}, \mathbf{y}) := x_i y_j$, $x_i x_j$ and $y_i y_j$ for each appropriate i and j.[1] In fact, monomial functions

[1] Assuming that $\mathcal{X}$ is finite then the arguments of Section 2 guarantee that, over finite intervals of time, the process $(\mathbf{X}(t), \mathbf{Y}(t))$ is bounded to remain in some compact set, and then the boundedness requirement for the functions h need only be honoured on this set.

of any order can be used to obtain equations for arbitrary order moments. However, if the functions $\mathbf{f}$ and $\mathbf{g}$ are non linear (as is usually the case), the term $\mathbb{E}[\mathcal{A}h(\mathbf{X}(s), \mathbf{Y}(s))]$ will involve expectations of non-linear functions of populations and will thus need to be simplified by applying some form of *moment-closure* approximation.

For example, in the client server model, the right hand side of Equation (4) will contain terms of the form $\mathbb{E}[\min(C_0(t), S_0(t))]$. In the past, and in Equation (3) above, the approximation $\min(\mathbb{E}[C_0(t)], \mathbb{E}[S_0(t)])$ has been used. This has been shown to work quite well in general for a large class of performance models [20]. However, if the process remains close to states where the arguments of the minimum function are equal, so-called *switch points*, for a long period of time, the accuracy of this approximation can decrease significantly for systems with low populations [20]. This is even more visible when the minimum function involves accumulated variables in aPCTMC models. We address this issue with a novel normal moment closure in Section 3.2.



(a) Active servers (b) Active aircon. units (c) Temperature

Fig. 2: Approximation of the evolution of standard deviation of populations and accumulations in the client/server model.

Figure 2 shows approximations of standard deviations in the client/server model (we extend the $\widetilde{\mathbb{E}}[\cdot]$ notation to higher moments and expressions on them, such as variance). As demonstrated in previous work [20], this is quite accurate in the case of the client and server populations, which are not dependent on the accumulated variables, as depicted in Figure 2(a). However, in case of the population of air conditioning units (Figure 2(b)), and the temperature variable (Figure 2(c)), there are large quantitative and qualitative differences accumulated over time. Section 3.2 will discuss ways to improve the accuracy.

3 Convergence properties

In this section of the paper we will prove that, in the limit of large populations, a suitably rescaled aPCTMC model converges to its mean-field approximation. We construct a sequence of aPCTMC models $\{(\mathbf{X}^N(t), \mathbf{Y}^N(t)) \in \mathbb{Z}_+^n \times \mathbb{R}^m\}_{N \in \mathbb{Z}_+}$. We assume that the elements $\mathcal{C}$ and $\mathbf{l}_c$ are fixed, but that the rate functions r_c^N and also $\mathbf{g}^N$ may vary with N. The initial conditions for the Nth model in the sequence are given by $(N\mathbf{x}_0, N\mathbf{y}_0)$ for some $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{Z}_+^n \times \mathbb{R}^m$. For each model in this sequence, we assume that the assumptions of Section 2 are satisfied so that all of the processes are well defined and write $S^N \subseteq \mathbb{Z}_+^n$ for the reachable state space of the discrete component of the Nth process.

We assume further that the functions $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $f_c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ can be defined independently of N as follows:

$$\mathbf{f}(\mathbf{x},\mathbf{y}) := \sum_{c\in\mathcal{C}} \mathbf{l}_c f_c(\mathbf{x},\mathbf{y}) := \sum_{c\in\mathcal{C}} (\mathbf{l}_c/N) r_c^N(N\mathbf{x}, N\mathbf{y}) \quad \mathbf{g}(\mathbf{x},\mathbf{y}) := (1/N)\mathbf{g}^N(N\mathbf{x}, N\mathbf{y})$$

and that $\mathbf{f}$ and $\mathbf{g}$ satisfy local Lipschitz conditions on $\mathbb{R}^n \times \mathbb{R}^m$. Further, we assume that solutions to the mean-field model given by Equation (3) exist globally. Define the rescaled processes $\bar{\mathbf{X}}^N(t) := \mathbf{X}^N(t)/N$ and $\bar{\mathbf{Y}}^N(t) := \mathbf{Y}^N(t)/N$, then we require that there is some compact subset of $\mathbb{R}^n$ that contains all of the state spaces of the rescaled processes $\bar{\mathbf{X}}^N(t)$. Assume also that $\mathbf{g}^N(\mathbf{x},\mathbf{y}) \leq C(\|\mathbf{x}\| + \|\mathbf{y}\| + 1)$ for all $\mathbf{x} \in S^N$ and $\mathbf{y} \in \mathbb{R}^m$ where $C \in \mathbb{R}_+$ is independent of N. Then by an application of Grönwall's lemma similar to that of Section 2, we have that for all $t \in [0,T]$, the rescaled stochastic processes and the mean-field approximations can be contained within a single compact set $S \subset \mathbb{R}^{n+m}$ that is independent of N.[2] Finally, we require that $r_c^N(\mathbf{x},\mathbf{y}) \leq D(\|\mathbf{x}\| + \|\mathbf{y}\| + 1)$ for all $c \in \mathcal{C}$, $\mathbf{x} \in S^N$ and $\mathbf{y} \in \{Ns : s \in S\}$ where $D \in \mathbb{R}_+$ is independent of N.

The following theorem shows that the rescaled processes converge in probability to the mean-field approximation.

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

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

Proof. We begin by representing each process $(\bar{\mathbf{X}}^N(t), \bar{\mathbf{Y}}^N(t))$ in terms of mutually independent rate-1 Poisson processes $\{P_c(t) : c \in \mathcal{C}\}$ by the *random-time change* approach [6]:

$$\bar{\mathbf{X}}^N(t) = \mathbf{x}_0 + \sum_{c\in\mathcal{C}} P_c\left(\int_0^t r_c^N(\mathbf{X}^N(s), \mathbf{Y}^N(s))\,\mathrm{d}s\right)\mathbf{l}_c/N$$

$$\bar{\mathbf{Y}}^N(t) = \mathbf{y}_0 + \int_0^t \mathbf{g}(\bar{\mathbf{X}}^N(s), \bar{\mathbf{Y}}^N(s))\,\mathrm{d}s$$

On S, $\mathbf{f}$ and $\mathbf{g}$ are both Lipschitz continuous; let K be a Lipschitz constant for both functions. Now define:

$$D^N(t) := \sup_{s\in[0,t]} \left\| \bar{\mathbf{X}}^N(s) - \mathbf{x}_0 - \int_0^s \mathbf{f}(\bar{\mathbf{X}}^N(u), \bar{\mathbf{Y}}^N(u))\,\mathrm{d}u \right\|$$

and $\epsilon^N(t) := \|\bar{\mathbf{X}}^N(t) - \mathbf{x}(t)\| + \|\bar{\mathbf{Y}}^N(t) - \mathbf{y}(t)\|$. Then we have for $t \in [0,T]$:

$$\epsilon^N(t) \leq D^N(T) + \int_0^t \|\mathbf{g}(\bar{\mathbf{X}}^N(s), \bar{\mathbf{Y}}^N(s)) - \mathbf{g}(\mathbf{x}(s), \mathbf{y}(s))\|\,\mathrm{d}s$$

[2] Note that it is then only strictly necessary for this theorem that $\mathbf{f}$, f_c and $\mathbf{g}$ are defined on S rather than on the whole of $\mathbb{R}^{n+m}$.

$$+\int_0^t \|\mathbf{f}(\bar{\mathbf{X}}^N(s), \bar{\mathbf{Y}}^N(s)) - \mathbf{f}(\mathbf{x}(s), \mathbf{y}(s))\| \, \mathrm{d}s \;\leq D^N(T) + 2K \int_0^t \epsilon^N(s) \, \mathrm{d}s$$

and by Grönwall's inequality, we obtain $\epsilon^N(t) \leq D^N(T) \exp(2KT)$. Now note that:

$$D^N(T) \leq$$
$$\sup_{s\in[0,T]} \left\| \sum_{c\in\mathcal{C}} \frac{\mathbf{l}_c}{N} \left[P_c\left(\int_0^s r_c^N(\mathbf{X}^N(u), \mathbf{Y}^N(u)) \, \mathrm{d}u \right) - \int_0^s r_c^N(\mathbf{X}^N(u), \mathbf{Y}^N(u)) \, \mathrm{d}u \right] \right\|$$

which can be bounded above by $\sum_{c\in\mathcal{C}} \|\mathbf{l}_c\| \sup_{s\in[0,T]} |P_c(NCs)/N - Cs|$ for some $C \in \mathbb{R}_+$ independent of N. The result then follows by the strong law of large numbers for the Poisson process, which is equivalent to the functional strong law of large numbers [e.g. 26, Section 3.2], that is, for all $S \in \mathbb{R}_+$, $\sup_{s\in[0,S]} \|P_c(Ns)/N - s\| \to 0$ as $N \to \infty$ with probability 1. □

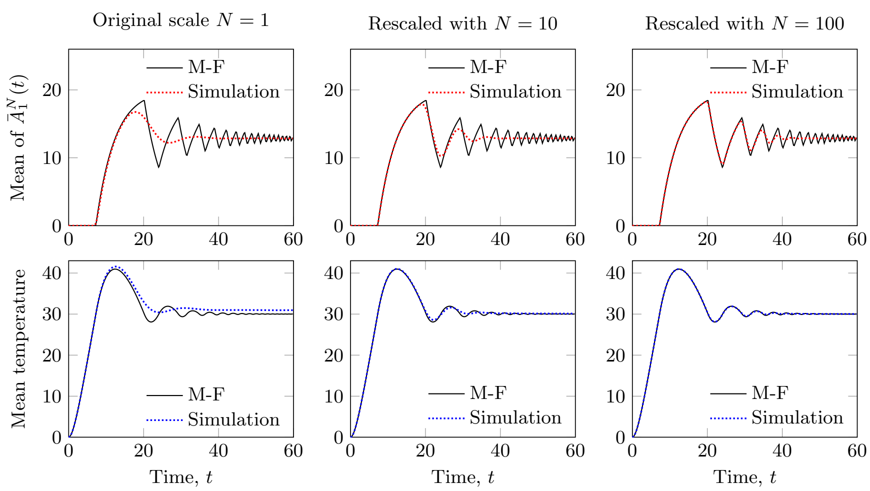


Fig. 3: Effect of rescaling on the first order mean-field approximation.

In terms of the client/server model, scaling the number of components by N, and, in particular, the number of servers, can be assumed to require a room approximately N times larger in volume than that of the original system. Therefore it makes sense if the initial heat energy content of the room E_0 is also scaled by N and the total heat energy content of the room is divided by N in order to obtain a physical model of the temperature as N increases, that is:

$$T^N(t) = \frac{c}{Nv} E^N(t) \quad \text{ and } \quad E_0^N = NE_0$$

Theorem 1 requires a continuity assumption on the transition rate and accumulation functions in a aPCTMC model. The indicator functions λ_{on}^N and λ_{off}^N in the client/server model do not satisfy these requirements. However, Figure 3 does seem to suggest empirically that convergence may still occur. Indeed, extensions of Theorem 1 to discontinuous rate functions may be possible by considering mean-field models in terms of *differential inclusions* [2, 8], but we do not pursue this further in this paper.

Instead, we can replace the 0/1–valued indicator functions in λ^N_{on} and λ^N_{off} with a more smooth proportional control, setting:

$$\lambda^N_{on}(t) = (T^N(t) - T_{thresh})^+ r_{on} \tag{5}$$

where f^+ is the positive part of f, that is $\max(f, 0)$. For simplicity we also set $\lambda^N_{\mathit{off}}(t) = r_{\mathit{off}}$ for all N. With this modification, Theorem 1 then applies, and is illustrated in Figure 6.

3.1 Second-order convergence

In this section, we give a second-order Gaussian convergence result for the sequence of rescaled *a*PCTMC models, which will directly motivate the improved moment closure approach of Section 3.2. We maintain all of the notation of the previous section.

In addition to the assumptions of the previous section, we assume that we can decompose $\mathbf{f}(\mathbf{x},\mathbf{y}) = \sum_i \mathbf{1}_{\{(\mathbf{x},\mathbf{y})\in F_i\}}\mathbf{f}^i(\mathbf{x},\mathbf{y})$ and $\mathbf{g}(\mathbf{x},\mathbf{y}) = \sum_j \mathbf{1}_{\{(\mathbf{x},\mathbf{y})\in G_j\}}\mathbf{g}^j(\mathbf{x},\mathbf{y})$ where $\{F_i\}$ and $\{G_j\}$ are finite collections of disjoint open sets in $\mathbb{R}^n \times \mathbb{R}^m$ such that for each i [resp. j], $\mathbf{f}^i$ [$\mathbf{g}^j$] is totally differentiable on $F^i \cap \mathrm{int}(S)$ [$G^j \cap \mathrm{int}(S)$] with uniformly continuous total derivative there. Then $\mathbf{f}$ [$\mathbf{g}$] has uniformly continuous total derivative on $\cup_i F_i \cap \mathrm{int}(S)$ [$\cup_i G_i \cap \mathrm{int}(S)$], which we write as $D\mathbf{f}$ [$D\mathbf{g}$].

Theorem 2. *Fix $T > 0$. Assume that the set $\{t \in [0,T] : (\mathbf{x}(t), \mathbf{y}(t)) \notin \cup_i F_i \cap \cup_j G_j \cap int(S)\}$ has Lebesgue measure zero. Then for mutually independent standard Brownian motions $\{B_c(t) : c \in \mathcal{C}\}$, the following equations have a unique strong solution [e.g. 16, Theorem 6.30] such that $(\mathbf{E}^X(t), \mathbf{E}^Y(t))$ is jointly-Gaussian:*

$$\mathbf{E}^X(t) := \int_0^t D\mathbf{f}(\mathbf{x}(s), \mathbf{y}(s)) \cdot (\mathbf{E}^X(s), \mathbf{E}^Y(s))^T\, ds + \sum_{c\in\mathcal{C}} B_c\left(\int_0^t f_c(\mathbf{x}(s), \mathbf{y}(s))\, ds\right) \mathbf{l}_c$$

$$\mathbf{E}^Y(t) := \int_0^t D\mathbf{g}(\mathbf{x}(s), \mathbf{y}(s)) \cdot (\mathbf{E}^X(s), \mathbf{E}^Y(s))^T\, ds$$

Furthermore, $\left(\frac{\mathbf{X}^N(t) - N\mathbf{x}(t)}{\sqrt{N}}, \frac{\mathbf{Y}^N(t) - N\mathbf{y}(t)}{\sqrt{N}}\right) \Rightarrow (\mathbf{E}^X(t), \mathbf{E}^Y(t))$ as $N \to \infty$, where the convergence is weak on $D([0,T]; \mathbb{R}^{n+m})$ endowed with the uniform topology [e.g. 1].[3]

Proof. We assume the representation of the processes $\bar{\mathbf{X}}^N(t)$ and $\bar{\mathbf{Y}}^N(t))$ given in Equation (5). Further it is possible [6, Corollary 5.5 and Remark 5.4] to construct, on the same probability space as these processes, mutually independent standard Brownian motions $\{B_c(t) : c \in \mathcal{C}\}$, such that:

$$Z_c := \sup_{t\in\mathbb{R}_+} \frac{|P_c(t) - t - B_c(t)|}{\log(2 \vee t)} < \infty \quad \text{almost surely}$$

[3] Informally, this is 'uniform convergence in distribution over $[0,T]$'.

From this it follows that as $N \to \infty$, almost surely:

$$\sqrt{N} \sup_{t \in [0,T]} \left\| \bar{\mathbf{X}}^N(t) - \mathbf{x}_0 - \int_0^t \mathbf{f}(\bar{\mathbf{X}}^N(s), \bar{\mathbf{Y}}^N(s))\, \mathrm{d}s \right.$$
$$\left. - \sum_{c \in \mathcal{C}} B_c \left(\int_0^t r_c^N(\mathbf{X}^N(s), \mathbf{Y}^N(s))\, \mathrm{d}s \right) (\mathbf{l}_c / N) \right\| \longrightarrow 0 \qquad (6)$$

A direct comparison of $\frac{\mathbf{X}^N(t) - N\mathbf{x}(t)}{\sqrt{N}}$ with $\mathbf{E}^X(t)$ and similarly for $\mathbf{E}^Y(t)$ using Equation (6) yields the result. We omit further details here for the sake of brevity. □

Theorem 2 also demands the continuity assumption on the transition rate and accumulation functions in an aPCTMC model so does not apply to the client/server model with threshold-based control. Indeed, Figure 4 shows that we do not even seem to observe convergence of the standard deviation approximation empirically in this case.

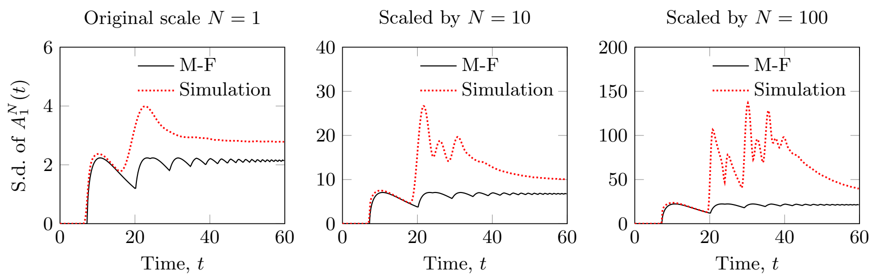


Fig. 4: Effect of scaling on the mean–field approximation of standard deviation of active air conditioning unit population.

In the case of the proportionally-controlled client/server model introduced above, Theorem 2 can be applied, although, in our experiments, for populations that are of similar orders to those considered in Figure 4, the approximation of the standard deviation of the temperature variable can be very inaccurate — convergence occurs very slowly. In the next section, based on the Gaussian assumption justified by Theorem 2, we introduce a technique that can provide significant improvements.

3.2 Normal approximations

Theorem 2 suggests that both the discrete and continuous components $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ of an aPCTMC model can be approximated by a jointly Gaussian process for sufficiently large populations. The proportionally-controlled client/server example considered in this paper and, more generally, a large class of computer performance models, for example, those specified using PEPA, stochastic Petri nets or many-server queueing networks contain rates with occurrences of minimum functions. In such cases, Equation (4), when applied to extract a first moment, contains expectations of the form $\mathbb{E}[\min(\alpha, \beta)]$ where α and β are linear combinations of any of the discrete or continuous components in the model at some time t. The mean field approximation $\min(\mathbb{E}[\alpha], \mathbb{E}[\beta])$ can often be quite accurate, but Theorem 2 suggests an alternative. Because a sequence of aPCTMC processes converges to a Gaussian process, the marginal distributions at each point in time converge to multivariate normal random variables. Using a result for the moments of a minimum of two bivariate normal random variables [4], we can obtain the following approximation (where Φ and ϕ are the CDF and PDF, respectively, of a standard normal random variable):

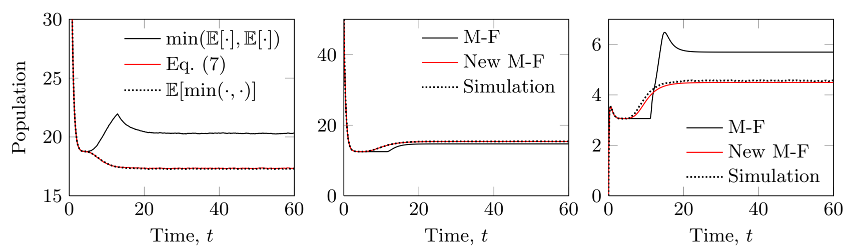
$$\mathbb{E}[\min(\alpha, \beta)] = \mathbb{E}[\alpha]\Phi(\Delta) + \mathbb{E}[\beta]\Phi(-\Delta) - \theta\phi(\Delta) \tag{7}$$

where $\theta := (\mathrm{Var}[\alpha] - 2\mathrm{Cov}[\alpha, \beta] + \mathrm{Var}[\beta])^{1/2}$ and $\Delta = (\mathbb{E}[\beta] - \mathbb{E}[\alpha])/\theta$. This uses only first- and second-order moments for which we can apply Equation (4) in order to extract equations governing their evolution.

For second-order moments, the mean field equations contain expectations of the form $\mathbb{E}[\gamma \min(\alpha, \beta)]$. Experiments suggest that the following approximation results in accurate approximations:

$$\mathbb{E}[\gamma \min(\alpha, \beta)] \approx \mathbb{E}[\gamma\alpha]\Phi(\Delta) + \mathbb{E}[\gamma\beta]\Phi(-\Delta) - \mathbb{E}[\gamma]\theta\phi(\Delta) \tag{8}$$

Figure 5 illustrates the improved accuracy for the discrete components of the model with proportional control.



(a) Expectation of $\min(\cdot, \cdot)$ (b) Mean of $S_0(t)$ (c) S.d. of $S_0(t)$

Fig. 5: Improved accuracy with normal approximations. Figure (a) compares the expected rate $\mathbb{E}[\min(C_0(t), S_0(t))r_{data}]$ with the two approximations, as obtained from simulation. Figures (b) and (c) show the effect of the different approximations when used in the mean-field equations for the mean and standard deviation of the S_0 population.

It is straightforward to adapt the result of [4] to obtain an expression for the maximum of bivariate normal random variables, so that an analogous approximation can be applied also to the proportional control expression (Equation (5)). Figure 6 compares simulation estimates with the numerical solution to the mean-field equations obtained from Equation (4) and with solutions to the new set of mean-field equations obtained by replacing occurrences of the minimum and maximum function according to the methods of this section. We see that this results in significant improvements in accuracy.

Figure 7 shows further that the normal moment closure can result in an accurate approximation of the standard deviation of the temperature even at relatively low scales of the system.

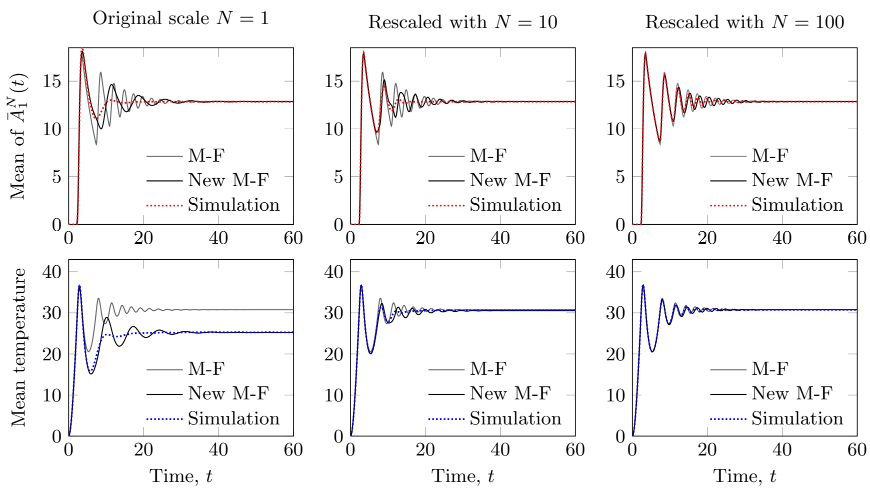


Fig. 6: Effect of scaling on the mean approximation when using the new normal approximations.

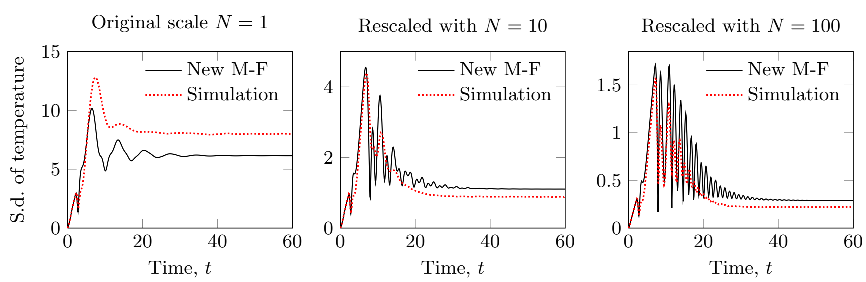


Fig. 7: Effect of scaling on standard deviation of temperature when using the new normal approximations.

4 Worked example

In this section we demonstrate the aPCTMC formalism and the efficient mean-field techniques on a larger example of a heterogeneous computing cluster. Similar to the client/server model, we consider a high level abstraction of the system. We assume that there are two types of servers in the cluster — ones with low (class A) and ones with high power consumption (class B), respectively. Clients in the system submit two types of jobs — with low (type 1) and high loads (type 2) on the servers. As in the client/server model, we include air conditioning units that maintain the ambient temperature in the room. Additionally, servers are capable of entering a sleep mode in the case that the temperature increases above a threshold. Unlike in the case of the client/server model where the client and server components of the discrete state space were unaffected by the accumulated variables, this will result in an aPCTMC with a complete dependence between the discrete components and the accumulated variables.

We use the PEPA process algebra to concisely describe the aPCTMC model ($j \in \{A, B\}$ is a server class and $i \in \{1, 2\}$ is a job type):

$$Client \stackrel{def}{=} \textstyle\sum_{i=1}^{2}(queue_i, r_{q,i}).Job_i \qquad Job_i \stackrel{def}{=} (service_i, r_{service_i}).Client$$
$$Server^j \stackrel{def}{=} \textstyle\sum_{i=1}^{2}(service^j_i, r_{service,i}).Server^j_i + (sleep, \lambda_{sleep}(t)).Server^j_{sleep}$$
$$Server^j_i \stackrel{def}{=} (reset, r_{reset}).Server^j \qquad Server^j_{sleep} \stackrel{def}{=} (wakeup, r_{wakeup}).Server^j$$

$$\big(\mathbf{Servers}\{Server^A[N_{SA}]|Server^B[N_{SB}]\}\big) \parallel \mathbf{Aircon}\{Aircon_0[N_A]\}$$
$$\underset{\{service_i|1\leq i\leq 4\}}{\bowtie} \mathbf{Clients}\{Client[N_C]\}$$

with rates $\lambda_{off}(t) = r_{on}$ and

$$\lambda_{sleep}(t) = (T(t) - T_{sleep})^{+} \cdot r_{j,sleep} \quad \text{and} \quad \lambda_{on}(t) = (T(t) - T_{thresh})^{+} \cdot r_{on}$$

where temperature is defined as in Equation (2) and the energy variable is

$$E(t) = E_0 + \int_0^t \sum_j (S^j(u)c_{j,s} + S^j_{sleep}(u)c_{j,sl} + S^j_1(u)c_{j,1} + S^s(u)c_{j,2}) - A_1(u)c_a \mathrm{d}u$$

for some constants $c_{j,s}, c_{j,sl}, c_{j,1}, c_{j,2}, c_a$.

Additionally, we transform the model so that it is possible to use mean-field techniques to calculate cumulative distribution functions of various passage-time random variables [11]. We will compute the time until an individual client executes its first high load job. Such measures are often used when expressing service level agreements (SLAs). The example will show how the presented framework can be used to study the trade-off between SLA satisfaction and the energy efficiency of the system. An increasingly common metric assessing energy efficiency of data centres is the *Power Usage Efficiency (PUE)* metric [17], calculated as the ratio between the total energy consumption and the energy used by the servers. In the above model, we can model the total energy consumption as an accumulated variable:

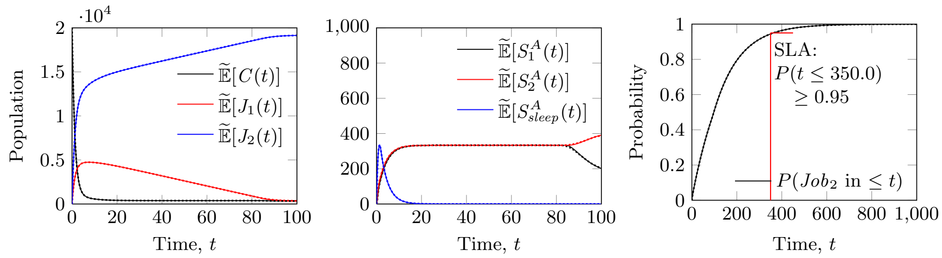
$$P(t) = \int_0^t \sum_j (p_{j,sl}S^j_{sleep}(u) + p_{j,s}S^j(u) + p_{j,1}S^j_1(u) + p_{j,2}S^j_2(u)) + p_a A_1(u)\mathrm{d}u$$

for some constants $p_{j,s}, p_{j,sl}, p_{j,1}, p_{j,2}, p_a$.

The quantity $U(t)$ represents the energy used for computation and is defined as $P(t)$, omitting the contribution of the air conditioning units and the servers in the sleeping state. To obtain an approximation of the mean PUE, we compute $\widetilde{\mathbb{E}}[P(t)]/\widetilde{\mathbb{E}}[U(t)]$ for sufficiently large t (1000 in the examples below).

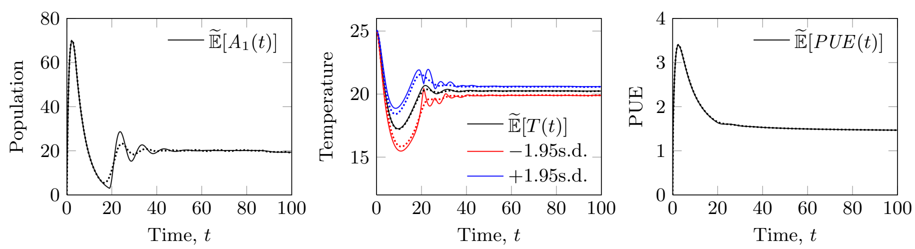
Figure 8 shows the mean populations of client and server-A components and the passage-time CDF as obtained by the mean-field analysis. Figure 9 shows the mean population of air conditioning units, its effect on the mean controlled temperature and the PUE of the system.

One benefit of mean-field analysis is the relatively low computational cost of numerically integrating the mean-field equations. This allows the evaluation of a large number of system configurations in a short time. For example, we can look at the relationship between the two temperature thresholds T_{thresh} and



(a) Mean client populations (b) Mean server populations (c) Passage time CDF

Fig. 8: Means of client/server populations and passage-time CDF in the computing cluster model.



(a) Active aircon. units (b) Temperature (c) PUE

Fig. 9: Mean population of active airconditioning units and mean temperature in the cluster model. The approximation of standard deviation in figure (b) was obtained by applying the normal min closure from Section 3.2.

T_{sleep} that specify when the air conditioning units start contributing to cooling and servers switch to sleep mode, respectively. We fix the server threshold at 23 units and search for the best air conditioning threshold. Our target measure to minimise will be the PUE in steady state of the system and the constraints are given by requiring satisfaction of the above SLA. Figure 10 explores a range of system configurations with the number of servers of each type $N_S = N_{SA} = N_{SB}$ varying between 50 and 1500 and the threshold T_{thresh} varying between 20 and 26 units.

Figure 10(a) shows the mean steady-state PUE for each configuration. For each size of the computing cluster given by a value of N_S, there is an optimal value of T_{thresh} achieving a minimal PUE metric. These thresholds and the corresponding optimal PUE values are shown by the thick solid line.

It can be seen that this is slightly below the server threshold, shown as the red dotted line. For example, for $N_S = 850$, the value of T_{thresh} achieving the optimal PUE is 22.7. The SLA is achieved only when there are sufficiently many servers in the system, shown as the darker region on the surface plot. Figure 10(b) shows that the optimal PUE line minimises the number of sleeping servers, while keeping the air conditioning units as lightly loaded as possible. Figure 10(c)

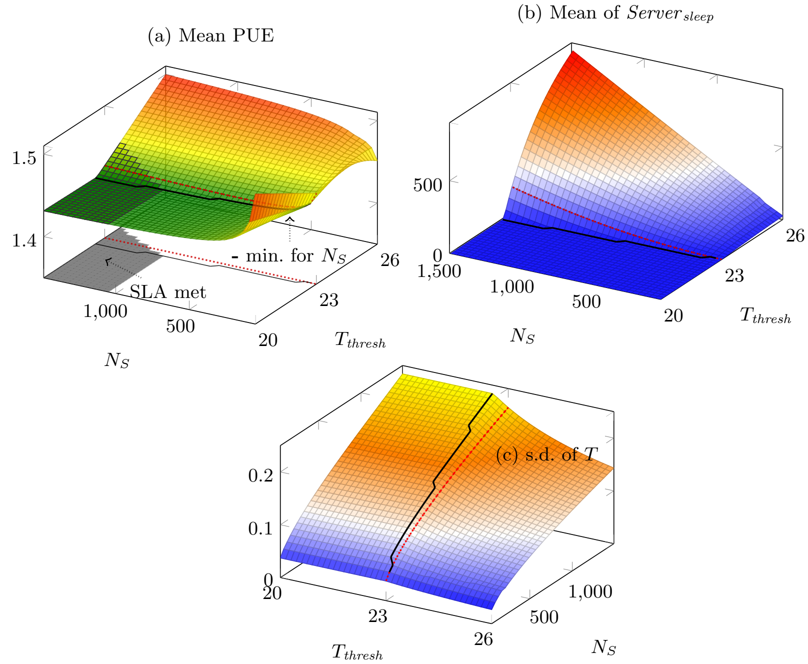


Fig. 10: The effect of cooling threshold and the number of servers on the steady state PUE metric and the number of servers in sleeping state. For each initial server population N_S, the thick black line shows the threshold under which the minimum PUE is achieved.

shows that line of minimum PUE separates the region with maximal standard deviation of the temperature variable.

5 Conclusion and future work

We have introduced the aPCTMC formalism, an extension to Markov population models that allows efficient modelling of feedback from accumulated quantities. We have extended the existing mean-field techniques to provide means and higher moments of populations and accumulations for aPCTMC models. Furthermore, we have provided convergence results justifying the mean-field approximation in the first- and second-order cases. The second-order result shows that sequences of aPCTMC models with increasing component populations converge to a jointly Gaussian process. This justifies the novel use of a normal approximation of minimum and maximum functions in the mean-field equations, resulting in significantly improved accuracy in both first- and second-order cases.

We have demonstrated the new framework on a substantial example of a computing cluster where server behaviour reacts to the ambient temperature controlled by an air conditioning system. An important advantage of the mean-field techniques is the low computational cost that can be used to explore a large number of different system configurations.

All of the numerical results in this paper were produced using a prototype implementation of the techniques in an extension to the *Grouped PEPA Analyser tool* [22].

In future, we plan to more formally investigate the accuracy improvements possible from the normal moment closures detailed in this paper. We also plan to investigate more complex accumulation mechanisms. For example, Gaussian noise might be introduced into the accumulation equations to account for error in sensor measurements of the continuous quantities.

Acknowledgment

Anton Stefanek, Richard A. Hayden and Jeremy T. Bradley are funded by EPSRC on the Analysis of Massively Parallel Stochastic Systems (AMPS) project (reference EP/G011737/1).

References

[1] Billingsley, P.: Convergence of Probability Measures. John Wiley & Sons (1968)
[2] Bortolussi, L.: Hybrid Limits of Continuous Time Markov Chains. In: QEST'11. pp. 3–12. IEEE, Aachen (Sep 2011)
[3] Bortolussi, L., Galpin, V., Hillston, J.: HYPE with stochastic events. EPTCS 57, 120–133 (Jul 2011)
[4] Cain, M.: The moment-generating function of the minimum of bivariate normal random variables. The American Statistician 48(2), 124–125 (1994)
[5] Davis, M.H.A.: Markov models and optimization. Chapman & Hall/CRC (1993)
[6] Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley (2005)
[7] Gast, N., Bruno, G.: A mean field model of work stealing in large-scale systems. In: SIGMETRICS. p. 13. ACM, New York, USA (Jun 2010)
[8] Gast, N., Gaujal, B.: Mean field limit of non-smooth systems and differential inclusions. SIGMETRICS Performance Evaluation Review 38(2), 30–32 (2010)
[9] Gillespie, C.S.: Moment-closure approximations for mass-action models. IET Systems Biology 3(1), 52–58 (2009)
[10] Hayden, R.A., Bradley, J.T.: A fluid analysis framework for a Markovian process algebra. Theoretical Computer Science 411(22-24), 2260–2297 (May 2010)
[11] Hayden, R.A., Stefanek, A., Bradley, J.T.: Fluid computation of passage-time distributions in large Markov models. Theoretical Computer Science 413(1), 106–141 (Jan 2012)
[12] Hillston, J.: Fluid flow approximation of PEPA models. In: QEST. pp. 33–42. IEEE (Sep 2005)
[13] Horton, G., Kulkarni, V.G., Nicol, D.M., Trivedi, K.S.: Fluid stochastic Petri nets: Theory, applications, and solution techniques. European Journal of Operational Research 105(1), 184–201 (1998)
[14] Kallenberg, O.: Foundations of Modern Probability. Springer (2002)
[15] Khadim, U.: A comparative study of process algebras for hybrid systems. Computer Science Report 06–23, Technische Universiteit Eindhoven (2006)

[16] Klebaner, F.C.: Introduction to stochastic calculus with applications. Imperial College Press, second edn. (2006)
[17] Rawson, A., Pfleuger, J., Cader, T.: Data Center Power Efficiency Metrics: PUE and DCiE. The Green Grid (2007)
[18] de Souza e Silva, E., Gail, R.: An algorithm to calculate transient distributions of cummulative rate and impulse-based rewards. Communications in Statistics: Stochastic Models 14(3), 509–536 (1998)
[19] Silva, M., Júlvez, J., Mahulea, C., Vázquez, C.R.: On fluidization of discrete event models: observation and control of continuous Petri nets. Discrete Event Dynamic Systems 21(4), 427–497 (Sep 2011)
[20] Stefanek, A., Hayden, R.A., Bradley, J.T.: A new tool for the performance analysis of massively parallel computer systems. In: QAPL. Electronic Proceedings in Theoretical Computer Science (Mar 2010)
[21] Stefanek, A., Hayden, R.A., Bradley, J.T.: Fluid Analysis of Energy Consumption using Rewards in Massively Parallel Markov Models. In: Computing. p. 121. ACM Press (2011)
[22] Stefanek, A., Hayden, R.A., Bradley, J.T.: GPA – A Tool for Fluid Scalability Analysis of Massively Parallel Systems. In: QEST. pp. 147–148. IEEE (Sep 2011)
[23] Tang, Q., Gupta, S., Varsamopoulos, G.: Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. Parallel and Distributed Systems, IEEE Transactions on 19(11), 1458–1472 (2008)
[24] Telek, M., Rácz, S.: Numerical analysis of large Markovian reward models. Performance Evaluation 36–37, 95–114 (Aug 1999)
[25] Tribastone, M., Gilmore, S., Hillston, J.: Scalable Differential Analysis of Process Algebra Models. IEEE Transactions on Software Engineering 38(1), 205–219 (Jan 2012)
[26] Whitt, W.: Internet supplement to Stochastic-Process Limits (2002), http://www.columbia.edu/~ww2040/supplement.html

A Parameters used in the examples

Values of rate and initial population parameters used in the client/server example, Figures 1, 2, 3, 4, 5:

$r_{data} = 0.6$	$r_{task} = 0.2$	$r_{reset} = 0.1$	$r_{on} = 0.2$	$r_{off} = 0.2$	$r_{heat} = 0.2$	$r_{cool} = 0.4$
$N_C = 40$	$N_S = 30$	$N_A = 20$	$T_{thresh} = 30$	$v = 1$	$c = 1$	

Values of rate and initial population parameters used in the worked example, Figures 8, 9, 10:

$r_{q,1} = 0.2$	$r_{q,2} = 0.5$	$r_{s,1} = 0.2$	$r_{s,2} = 0.2$	$r_{reset} = 0.2$	$r_{wakeup} = 0.3$
$r_{on} = 0.2$	$r_{off} = 0.2$	$T_0 = 25$	$T_{thresh} = 20$	$T_{sleep} = 23$	
$p_{A,s} = 10$	$p_{A,sl} = 1$	$p_{A,1} = 30$	$p_{A,2} = 37.5$	$r_{cool} = 0.026$	
$N_C = 20000$	$N_S = 1000$	$N_A = 100$			

The constants $p_{B,\cdot}$ are set as the respective $p_{A,\cdot}$ constants multiplied by 1.7 and the heat constants $c_{\cdot,\cdot}$ are set as the corresponding $p_{\cdot,\cdot}$ constants multiplied by a conversion factor 7.71×10^{-6}.