Exploiting non-deterministic analysis in the integration of transient solution techniques for Markov Regenerative Processes

Marco Biagi[1], Laura Carnevali[1], Marco Paolieri,[2]
Tommaso Papini[1], and Enrico Vicario[1]

[1] Department of Information Engineering, University of Florence, Italy
{marco.biagi,laura.carnevali,tommaso.papini,enrico.vicario}@unifi.it
[2] Department of Computer Science, University of Southern California
paolieri@usc.edu

Abstract. Transient analysis of Markov Regenerative Processes (MRPs) can be performed through the solution of Markov renewal equations defined by *global* and *local kernels*, which respectively characterize the occurrence of regenerations and transient probabilities between them. To derive kernels from stochastic models (e.g., stochastic Petri nets), existing methods exclusively address the case where at most one generally-distributed timer is enabled in each state, or where regenerations occur in a bounded number of events. In this work, we analyze the state space of the underlying timed model to identify epochs between regenerations and apply distinct methods to each epoch depending on the satisfied conditions. For epochs not amenable to existing methods, we propose an adaptive approximation of kernel entries based on partial exploration of the state space, leveraging heuristics that permit to reduce the error on transient probabilities. The case study of a polling system with generally-distributed service times illustrates the effect of these heuristics and how the approach extends the class of models that can be analyzed.

Keywords: Non-Markovian Petri Nets, Markov Regenerative Process, enabling restriction, stochastic state class, non-deterministic analysis.

1 Introduction

In quantitative evaluation of concurrent models, generally distributed (GEN) durations support modeling validity but break the Markov property and rule out efficient solution techniques for Continuous Time Markov Chains (CTMCs). If the model guarantees that, always, with probability 1 (w.p.1), the Markov property will be eventually satisfied at some *regeneration* point, then the underlying stochastic process belongs to the class of Markov Regenerative Processes (MRPs) [12].

MRPs attain a fortunate trade-off between expressivity of models and feasibility of numerical solution, which is reduced to the evaluation of a *global kernel* and a *local kernel* that characterize behavior in the sequencing of regeneration points and in the *epochs* between them. However, numerical derivation of the kernels has been solved only for some isolated sub-classes of MRP models [7].

Most works address the subclass where at most a single GEN timer is enabled in each state (*enabling restriction*), so that each kernel component can be computed by analyzing the CTMC subordinated to the activity interval of the active GEN [6, 9, 1]. The method of supplementary variables [8, 17] might in principle encompass the case of multiple concurrently enabled GEN timers, but practical feasibility restrains applicability under the enabling restriction. Sampling at equidistant time points [15, 19] permits evaluation for models where all timers have either deterministic (DET) or exponentially distributed (EXP) durations.

The method of stochastic state classes [18] enables quantitative evaluation of stochastic processes with multiple concurrent GEN timers, possibly with bounded support; in particular, for models that always reach a regeneration within a bounded number of discrete events, which we call the *bounded regeneration restriction*, exact evaluation of kernels is performed enumerating stochastic transient trees that cover the states between two subsequent regenerations [10].

For models that break both the enabling and the bounded regeneration restriction, kernel components may be still defectively approximated by truncation of stochastic transient trees [10], which may also serve to reduce complexity for models under bounded regeneration. However, this faces an inherent contrast. On the one hand, state space truncation has a different impact on the final evaluation, depending on the probability of reaching truncation points. On the other hand, when the analysis exploits regenerations to decompose state space coverage, each epoch starts from a memoryless condition, which is not able to distinguish whether the probability mass under analysis is relevant or negligible.

In this paper, we exploit non-deterministic analysis to drive integration of different solution techniques, exact and approximate, that are applicable to different regenerative epochs. To this end, we characterize the structure of the state space through terminating and efficient non-deterministic analysis based on the representation of timing domains with Difference Bounds Matrices (DBMs), identifying regenerative epochs and solution techniques that can be applied for kernel components corresponding to each regeneration (Sect. 3.1). This permits integration of the consolidated technique of enabling restriction with exact and approximate solution based on stochastic state classes (Sect. 3.2). Moreover, we also introduce a novel technique that iteratively adapts the approximation of each kernel component so as to optimize the impact of truncation on the defect in the evaluation of transient probabilities (Sects. 3.3 and 3.4). The approach permits to accurately evaluate transient probabilities of markings, and it is open to further adaptation strategies and to integration of other solution techniques, both numerical and simulative. Application is illustrated with reference to an instance of the polling system problem [11, 13] with generally distributed service times and exhaustive service subordinated to a deterministic timeout (Sect. 4).

To make the paper self-contained, we recall the formalism of Stochastic Time Petri Nets (STPNs) and transient analysis of MRPs (Sect. 2). Finally, we draw our conclusions and discuss future steps enabled (Sect. 5).

2 Preliminaries

2.1 Stochastic Time Petri Nets

Definition 1. *An STPN is a tuple $\langle P, T, A^-, A^+, A^\bullet, m_0, U, EFT, LFT, F, W\rangle$: P is the set of places; T is the set of transitions; $A^- \subseteq P \times T$, $A^+ \subseteq T \times P$, $A^\bullet \subseteq P \times T$ are the sets of precondition, postcondition, inhibitor arcs, respectively; $m_0 \in \mathbb{N}^P$ is the initial marking; U associates each transition t with an update function $U(t) : \mathbb{N}^P \to \mathbb{N}^P$ which, in turn, associates each marking with a new marking; $EFT : T \to \mathbb{Q}_{\geq 0}$ and $LFT : T \to \mathbb{Q}_{\geq 0} \cup \{\infty\}$ associate each transition with an earliest and a latest firing time, respectively; F associates each transition t with a Cumulative Distribution Function (CDF) $F(t)$ over $[EFT(t), LFT(t)]$; and, $W : T \to \mathbb{R}_{>0}$ associates each transition with a weight.*

A place p is an *input*, *output*, *inhibitor* place for a transition t if $\langle p, t\rangle \in A^-$, $\langle t, p\rangle \in A^+$, $\langle p, t\rangle \in A^\bullet$, respectively; precondition and postcondition arcs are represented by arrows, while inhibitor arcs by dotted arrows. A transition t is *immediate* (IMM) if $EFT_t = LFT_t = 0$ and *timed* otherwise; a timed transition t is *exponential* (EXP) if $F_t(x) = 1 - e^{-\lambda x}$ over $[0, \infty]$ with $\lambda \in \mathbb{R}_{>0}$, and *general* (GEN) if it has a non-exponential CDF; a GEN transition t is *deterministic* (DET) if $EFT_t = LFT_t > 0$ and *distributed* otherwise; for each distributed transition t, we assume that F_t is absolutely continuous and thus expressed as the integral function of a Probability Density Function (PDF) f_t, ruling out mixed (continuous and discrete) distributions. IMM, EXP, GEN, DET transitions are represented by thin black, thick white, thick black, thick gray bars, respectively. Update functions and weights are annotated next to transitions as "`place` ← *expression*" and "*weight* = *value*", respectively.

The state of an STPN is a pair $\langle m, \phi\rangle$, where m is a marking and $\phi : T \to \mathbb{R}_{\geq 0}$ associates each transition with a time-to-fire. A transition is *enabled* by a marking if each of its input places contains at least one token and none of its inhibitor places contains any token; an enabled transition t is *firable* in a state if its time-to-fire is equal to zero. The next transition t to fire in a state $s = \langle m, \phi\rangle$ is selected among the set $T_{\mathrm{f},s}$ of firable transitions in s with probability equal to $W(t)/\sum_{t_i \in T_{\mathrm{f},s}} W(t_i)$. When t fires, s is replaced by $s' = \langle m', \phi'\rangle$, where: m' is derived from m by *i*) removing a token from each input place of t (yielding marking $\tilde{m}$), *ii*) adding a token to each output place of t (yielding marking $\hat{m}$), and *iii*) applying the update function $U(t)$ to $\hat{m}$; ϕ' is derived from ϕ by *i*) reducing the time-to-fire of *persistent* transitions (i.e., enabled by m, $\tilde{m}$, $\hat{m}$, m') by the time elapsed in s; *ii*) sampling the time-to-fire of each *newly-enabled* transition t_n (i.e., enabled by m' but not by $\tilde{m}$) according to F_{t_n}; and, *iii*) removing the time-to-fire of *disabled* transitions (i.e., enabled by m but not by m').

Given an initial marking m_0 and an initial PDF $f_{\boldsymbol{\tau}_0}$ for the vector $\boldsymbol{\tau}$ of the times-to-fire of the enabled transitions, the STPN semantics induces a probability space $\langle \Omega_{m_0}, \mathbb{F}_{\boldsymbol{\tau}_0}, \mathbb{P}_{m_0, f_{\boldsymbol{\tau}_0}}\rangle$, where Ω_{m_0} is the set of outcomes (i.e., feasible timed firing sequences of the model) and $\mathbb{P}_{m_0, f_{\boldsymbol{\tau}_0}}$ is a probability measure over them [16]. Note that $\mathbb{P}_{m_0, f_{\boldsymbol{\tau}_0}}$ is zero for outcomes that are not feasible under $f_{\boldsymbol{\tau}_0}$.

Fig. 1 shows a running example. The firing of `restart` enables `gen1` and makes `reg`, `enab`, and `approx` firable: the firing of `reg` enables `gen2`, which fires w.p.1; the firing of `enab` enables the cycle `exp1`–`exp2`, which can fire an unbounded number of times; the firing of `approx` enables the cycle `gen3`–`gen4`, which can fire an unbounded number of times. In all three cases, `gen1` is persistent and will eventually fire w.p.1, bringing the STPN to the initial marking `Restart` (note that the update function of `gen1` flushes places `E1`, `E2`, `G3`, `G4`).

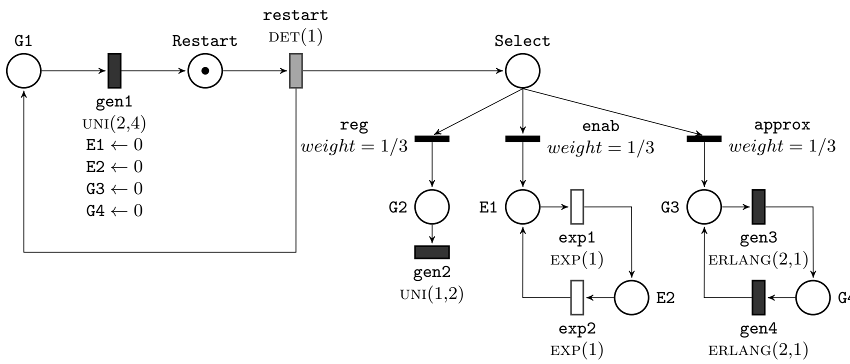


Fig. 1. A simple STPN with multiple concurrent GEN, DET, and EXP transitions: `gen1` and `gen2` have a uniform distribution over $[2, 4]$ and $[1, 2]$, respectively; `gen3` and `gen4` have an Erlang distribution with shape 2 and rate 1; `restart` has firing time equal to 1; `exp1` and `exp2` have an EXP distribution with rate 1.

2.2 Transient analysis of Markov Regenerative Processes

The marking process $\{M(t), t \geq 0\}$, where $M(t)$ is the marking at time t, specifies the logic state of an STPN at each time instant. If the marking process is an MRP [7], its transient evolution is completely characterized by: *i*) the initial probabilities of markings; *ii*) a local kernel $L_{ij}(t) := P\{M(t) = j, T_1 > t \,|\, M(0) = i\}$, where T_1 is the time of the first regeneration after regeneration i, characterizing the evolution between two subsequent regenerations (i.e., $L_{ij}(t)$ is the probability that, starting from regeneration i at time 0, no regeneration is reached within time t and the marking at time t is j); and, *iii*) a global kernel $G_{ik}(t) := P\{M(T_1) = k, T_1 \leq t \,|\, M(0) = i\}$ characterizing the occurrence of regenerations (i.e., $G_{ik}(t)$ is the probability that, starting from regeneration i at time 0, the first regeneration is reached on marking k within time t).

Transient probabilities of markings $\pi_{ij}(t) := P\{M(t) = j \,|\, M(0) = i\}$ are the solution of a set of Markov renewal equations defined by the kernels [5, 12]:

$$\pi_{ij}(t) := P\{M(t) = j \,|\, M(0) = i\} = L_{ij}(t) + \sum_{k \in \Theta} \int_0^t g_{ik}(x)\, \pi_{kj}(t - x)\, dx \quad (1)$$

where $g_{ik}(x) := d\,G_{ik}(x)/dx$. While Eq. 1 can be solved numerically by discretization, kernels can be computed only for some sub-classes of MRP models.

The marking process of the STPN of Fig. 1 is an MRP since the firing of `gen1`, which always occurs w.p.1. (possibly after an unbounded number of firings), brings the process to the initial regeneration where `restart` is newly-enabled.

Analysis under the enabling restriction. The *enabling restriction* [6, 8] assumes that at most a single GEN time-to-fire is enabled in each state, which in turn implies that it is never the case that a GEN transition *continues* (persists) at the firing of another GEN transition. If an MRP complies with the enabling restriction, then in each regenerative epoch the process behaves either as a CTMC, if only EXP transitions are enabled in the initial regeneration, or as a CTMC subordinated to the activity interval of a GEN transition (i.e., the time interval during which the transition is enabled), if a GEN transition is enabled in the initial regeneration. In this case, the kernels can be computed from CTMC transient probabilities through the method of [9, 6].

The marking process of the STPN of Fig. 1 does not satisfy the enabling restriction, since `gen2`, `gen3`, and `gen4` may be enabled concurrently with `gen1`.

Analysis under the bounded regeneration restriction. The method of stochastic state classes [10] permits computation of kernels for models with multiple GEN times-to-fire concurrently enabled, also with overlapping activity intervals, but for exact evaluation requires that: always, a regeneration is eventually reached within a bounded number of discrete events. We term this case as the *bounded regeneration restriction*. The marking process of the STPN of Fig. 1 does not satisfy the bounded regeneration restriction, since both cycles `exp1`–`exp2` and `gen3`–`gen4` may fire an unbounded number of times while `gen1` is persistent, without reaching a regeneration.

A *stochastic state class* samples the state of the MRP immediately after a firing, encoding a marking and a joint domain and PDF for the absolute time and for the times-to-fire of the enabled transitions.

Definition 2. *A stochastic state class is a tuple $\Sigma = \langle m, D, f \rangle$ where: m is a marking; D is the support of the random vector $\langle \tau_{\text{age}}, \boldsymbol{\tau} \rangle$, where τ_{age} is the absolute time and $\boldsymbol{\tau}$ is the vector of the remaining times-to-fire of the enabled transitions; and, f is the PDF of $\langle \tau_{\text{age}}, \boldsymbol{\tau} \rangle$, which we term state density function.*

Starting from an initial stochastic state class with $\tau_{\text{age}} = 0$ and independently distributed times-to-fire for the enabled transitions, enumeration of a reachability relation among stochastic state classes yields a *stochastic transient tree*, where the support of the vector $\boldsymbol{\tau}$ in each class is a Difference Bounds Matrix (DBM), i.e., a linear convex polyhedron that represents the solution of a set of linear inequalities constraining the difference between pairs of times-to-fire.

Definition 3. *A stochastic state class $\Sigma' = \langle m', D', f' \rangle$ is the successor of a stochastic state class $\Sigma = \langle m, D, f \rangle$ through a transition t with probability μ,* *which we write* $\Sigma \overset{t,\mu}{\Rightarrow} \Sigma'$*, iff, given that the marking is* m *and the random vector* $\langle \tau_{\text{age}}, \boldsymbol{\tau} \rangle$ *is distributed over* D *according to* f*,* t *fires with probability* μ*, yielding a marking* m' *and a random vector* $\langle \tau'_{\text{age}}, \boldsymbol{\tau}' \rangle$ *distributed over* D' *according to* f'*.*

A stochastic state class is said to be *regenerative* if the Markov property is satisfied immediately after the class is entered, which occurs iff all active GEN times-to-fires have been enabled for a deterministic time [16]:

Definition 4. *A stochastic state class* Σ *is termed regenerative if the time elapsed from the enabling of each enabled GEN transition* t_i *until the firing that led to* Σ *is a deterministic value* $d_i \in \mathbb{R}_{\geq 0}$*, termed the enabling time of* t_i *in* Σ*.*

In *exact regenerative transient analysis* [10], stochastic state classes are enumerated from each regeneration until any regeneration is reached, yielding a set of *stochastic transient trees* that are rooted in a regenerative stochastic state class and contain non-regenerative successors reached before any regeneration. Under the bounded regeneration restriction, each tree is finite, collecting all stochastic state classes that capture the MRP behavior during a regenerative epoch, with (regenerative) leaf nodes characterizing the global kernel and (non-regenerative) inner nodes characterizing the local kernel. For any regenerative stochastic state classes i, integration of the PDF of $\langle \tau_{\text{age}}, \boldsymbol{\tau} \rangle$ in the stochastic state classes belonging to the tree rooted in i permits to compute the kernel entries $L_{ij}(t)$ and $g_{ik}(t)$ by summing up the measure of probability of states in the classes of the transient stochastic tree rooted in i, for any non-regenerative stochastic state class j, for any regenerative stochastic state class k, and for any time t.

2.3 Non-deterministic analysis

An STPN identifies a Time Petri Net (TPN) [14, 3] with same outcomes Ω_{m_0}.

Definition 5. *A state class* $S = \langle m, D \rangle$ *is made of a marking* m *and a support* D *for the vector* $\boldsymbol{\tau}$ *of the remaining times-to-fire of the enabled transitions.*

Starting from an initial marking m_0 and an initial domain D_0 for $\boldsymbol{\tau}$, enumeration of the reachability relation among state classes yields a *State Class Graph* (SCG), which represents the continuous set of executions Ω_{m_0} and supports correctness verification of the TPN model (*non-deterministic analysis*).

Definition 6. $S' = \langle m', D' \rangle$ *is the successor of* $S = \langle m, D \rangle$ *through transition* t*, i.e.,* $S \overset{t}{\rightarrow} S'$*, iff, given that the marking is* m *and* $\boldsymbol{\tau}$ *is supported over* D*,* t *fires in* S*, yielding marking* m' *and a new vector* $\boldsymbol{\tau}'$ *supported over* D'*.*

If $EFT(t) \in \mathbb{Q}_{\geq 0}$ and $LFT(t) \in \mathbb{Q}_{\geq 0} \cup \{\infty\}$ for every transition t, then the SCG is finite provided that the model generates a finite number of markings [10], which does not comprise a modeling limitation for most applicative scenarios.

3 Integration of transient solution techniques for MRPs

Non-deterministic state space analysis of the underlying TPN of an STPN model permits identification of regeneration epochs and verification of whether each of them satisfies the enabling or bounded regeneration restrictions (Sect. 3.1), driving integration of different solution techniques for the evaluation of kernels (Sect. 3.2). For epochs that satisfy neither of the two restrictions, partial enumeration of stochastic state classes supports approximated evaluation of the kernels, resulting in a safe defective approximation of transient probabilities (Sect. 3.3).

3.1 Analysis of regenerative epochs

The set of states collected in a *stochastic* state class identifies a unique underlying *non-deterministic* state class [18] that represents the marking and the support of the vector of the remaining times-to-fire of the enabled transitions when the class is entered. The association between non-deterministic and stochastic state classes is one-to-many (possibly one-to-infinite) and preserves qualitative properties referred to the set of feasible outcomes Ω_{m_0}, while abstracting from quantitative properties depending on the probability measure $\mathcal{P}_{m_0,f_{\tau_0}}$. Given that a stochastic state class is regenerative if it satisfies Definition 4, which depends on Ω_{m_0} but not on $\mathcal{P}_{m_0,f_{\tau_0}}$, state classes can be used to identify regenerations.

To this end, the state space of the underlying TPN is covered by a set of SCGs, which we call *First-Epoch State Class Graphs* (FESCGs), each rooted in a regenerative state class and containing all non-regenerative successors reached before any regeneration (which is also included in the graph). Enumeration of FESCGs can suppress successor relations that correspond to null probability events, i.e., firings that in any associated stochastic state class would be possible in a null measure subset of the support.

Lemma 1. *Let u be an STPN, v be its underlying TPN, R be the set of successor relations $\Sigma = \langle m, D, f\rangle \xrightarrow{t,\mu} \Sigma' = \langle m', D', f'\rangle$ in the stochastic transient tree of u enumerated from a regenerative stochastic state class $\Sigma_0 = \langle m_0, D_0\rangle$, and $S = \langle m, \bar{D}\rangle \xrightarrow{t} S' = \langle m', \bar{D}'\rangle$ be a succession relation in the SCG of v enumerated from a regenerative state class $S_0 = \langle m_0, \bar{D}_0\rangle$, such that $\bar{D}$, $\bar{D}'$, and $\bar{D}_0$ are the projections of D, D', and D_0 that eliminate τ_{age}, respectively. A succession relation $\Sigma \xrightarrow{t,\mu} \Sigma' \in R$ has probability $\mu = 0$ iff the projection of D that eliminates DET and IMM timers, conditioned to the firing of transition t, has a null measure in $\mathbb{R}^N$, where N is the number of distributed times-to-fire in Σ and S.*

Proof. Let D_t be D conditioned to the firing of t, i.e., $D_t = D \cap \{\tau_t \leq \tau_{t_i}\, \forall\, t_i \in E(m)\}$, where τ_t is the time-to-fire of t and $E(m)$ is the set of transitions enabled by m. Let $\hat{D}_t$ be the projection of D_t that eliminates DET and IMM timers.

(If) If $\hat{D}_t$ has null measure in $\mathbb{R}^N$, either *i)* the STPN includes some transition associated with a mixed distribution, or *ii)* $\mu = 0$. By Definition 1, the CDF of each GEN transition is absolutely continuous over its support, thus $\mu = 0$.

(Only if) If, ab absurdo, $\hat{D}_t$ had non-null measure in $\mathbb{R}^N$, then the integral over $\hat{D}_t$ of the marginal distribution of distributed times-to-fire in Σ conditioned to the firing of t would not be zero, yielding $\mu \neq 0$. □

It is straightforward to show that a regenerative epoch complies with the enabling restriction iff at most one GEN transition is enabled in each state class of its FESCG. Conversely, compliance with the bounded regeneration restriction depends on the presence of cycles in the FESCG.

Lemma 2. *A regenerative epoch complies with the bounded regeneration restriction iff its FESCG does not include any cycle.*

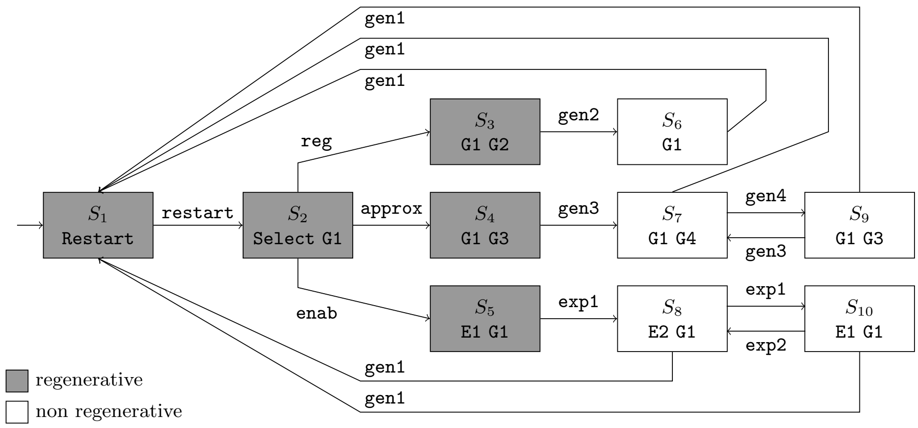
Proof. (If) If, ab absurdo, a regenerative epoch did not satisfy the bounded regeneration restriction, the STPN would allow a timed firing sequence made of an unbounded number of firings that never visits a regeneration; given that an STPN and its underlying TPN have the same set of timed firing sequences Ω_{m_0}, also the TPN would allow that behavior. Given that each state class is associated with one or more stochastic state classes having the same marking and time domain, there would exist a state class associated with an unbounded number of stochastic state classes. As a consequence, the FESCG would include a cycle.

(Only if) If, ab absurdo, the FESCG of a regeneration included a cycle, then, by construction, that cycle would not visit any regenerative state class. Hence, there would exist a timed firing sequence that would allow an unbounded number of firings without visiting a regeneration, and the corresponding regenerative epoch would not comply with the bounded regeneration restriction. □

Fig. 2 shows the SCG of the TPN underlying the STPN of Fig. 1, consisting of 5 regenerative and 5 non-regenerative state classes. In particular: the FESCG rooted in S_3 includes S_6 and S_1, satisfying the bounded regeneration restriction (it is cycle free) but not the enabling restriction (two GEN transitions are enabled in S_6); the FESCG rooted in S_5 includes S_8, S_{10}, and S_1, complying with the enabling restriction but not with the bounded regeneration restriction (due to the cycle S_8–S_{10}); and, the FESCG rooted in S_4 includes S_7, S_9, and S_1, satisfying neither the bounded regeneration restriction (due to the cycle S_7–S_9) nor the enabling restriction (two GEN transitions are enabled in S_4, S_7, and S_9). Note that the firing of transition `gen1` in state class S_3 would have probability zero in any associated stochastic state class and thus it is suppressed.

3.2 An algorithm for transient analysis of MRPs

Given an STPN with underlying MRP, the kernel entries of each regenerative epoch can be derived through a different solution technique depending on whether the epoch satisfies the bounded regeneration restriction, or the enabling restriction, or neither of the two conditions. The applicable solution strategy can be efficiently selected through non-deterministic analysis of the underlying TPN of the model, by enumerating the SCG so as to identify the set Θ of regenerative state classes, the set Ψ of reachable markings, and the FESCG of each regenerative state class $i \in \Theta$:



(b)

Fig. 2. The SCG of the TPN underlying the STPN of Fig. 1: state classes are represented by rectangles labeled with the marking; successor relations between state classes are represented by arrows labeled with the fired transition.

- if the FESCG of i complies with the bounded regeneration restriction (e.g., the FESCG rooted in S_3 in Fig. 2), $L_{ij}(t)$ and $g_{ik}(t)$ are computed through the exact regenerative transient analysis of [10], for any marking $j \in \Psi$, for any regenerative state class $k \in \Theta$, and for any time point t;
- if the FESCG of i satisfies the enabling restriction (e.g., the FESCG rooted in S_5 in Fig. 2), $L_{ij}(t)$ and $g_{ik}(t)$ are derived through the method of [9, 6];
- if the FESCG of i breaks both the enabling and the bounded regeneration restrictions (e.g., the FESCG rooted in S_4 in Fig. 2), $L_{ij}(t)$ and $g_{ik}(t)$ can still be estimated by stochastic simulation of the STPN model or they can be approximated by numerical solution as developed in Sect.3.3.

Note that in so doing the derivation of kernel entries always terminates (even for models with an underlying marking process beyond the class of MRPs), provided that the FESCG of each regenerative state class is finite, which in turn is guaranteed under the fairly general conditions mentioned in Sect. 2.3. Also note that, in the present implementation, regenerative epochs that satisfy both the restrictions are analyzed through exact regenerative transient analysis, but analysis under the enabling restriction could be applied as well; moreover, approximated analysis or simulation might be applied also to regenerative epochs that satisfy one or both the restrictions, as a way to reduce complexity of solution. In-depth comparison and experimentation of the impact of different choices on accuracy and complexity deserves further study.

When kernel entries have been evaluated, transient probabilities of reachable markings are finally derived by numerical integration of the Markov renewal equations of Eq. 1.

3.3 Approximate evaluation of the kernels of an MRP

In general, and in particular for regenerative epochs that do not satisfy either the bounded regeneration or the enabling restrictions, an approximation of kernel entries can be derived by truncating the enumeration of the stochastic transient tree computed in the exact regenerative transient analysis [10]. In this case, following the steps of Sect. 2.2, the approximated kernel entries $\tilde{L}_{hj}(t)$ and $\tilde{g}_{ik}(x)$ are computed on a subset of the classes in the stochastic transient tree of the regenerative state class i, and they thus comprise an under-approximation of the exact values $L_{hj}(t)$ and $g_{ik}(x)$. Specifically, denoting $\Delta_{ij} := L_{hj}(t) - \tilde{L}_{hj}(t)$ and $\delta_{ik} := g_{ik}(x) - \tilde{g}_{ik}(x)$, we have $\Delta_{ij} \geq 0$ and $\delta_{ik} \geq 0$.

To characterize the impact of the approximation, the following Lemma provides a bound on $\epsilon_{ij}(t) := \pi_{ij}(t) - \tilde{\pi}_{ij}(t)$, with $\tilde{\pi}_{ij}(t)$ denoting the solution of Eq. 1 obtained with approximated kernel entries:

$$\tilde{\pi}_{ij}(t) = \tilde{L}_{ij}(t) + \sum_{k\in\Theta} \int_0^t \tilde{g}_{ik}(x)\, \tilde{\pi}_{kj}(t-x)\, dx \tag{2}$$

Lemma 3. *For each regenerative state class $i \in \Theta$, marking $j \in \Psi$, and time t, the error $\epsilon_{ij}(t)$ is non-negative and upper-bounded:*

$$0 \leq \epsilon_{ij}(t) \leq \phi_i(t) + \sum_{k\in\Theta} \int_0^t \left(\tilde{g}_{ik}(x)\epsilon_{kj}(t-x) + \phi_i(x)(\epsilon_{kj}(t-x) + \tilde{\pi}_{kj}(t-x))\right) dx \tag{3}$$

where $\phi_i(t) := \sum_{j\in\Psi}(L_{ij}(t) - \tilde{L}_{ij}(t)) + \sum_{k\in\Theta}(g_{ik}(t) - \tilde{g}_{ik}(t))$.

Proof. By combining Eqs. 1 and 2, we obtain: $\epsilon_{ij}(t) = \Delta_{ij}(t) + \sum_{k\in\Theta}\int_0^t (\tilde{g}_{ik}(t) + \delta_{ik}(x)) \cdot \epsilon_{kj}(t-x) + \delta_{ik}(x)) \cdot \tilde{\pi}_{kj}(t-x)dx$ Since $\Delta_{ij}(t) \geq 0$ and $\delta_{ik}(t) \geq 0$, $\phi_i(t) \geq \Delta_{ij}(t)\ \forall\, j \in \Psi$ and $\phi_i(t) \geq \delta_{ik}(t)\ \forall\, k \in \Theta$. The upper bound of Eq. 3 can thus be obtained by replacing $\Delta_{ij}(t)$ and $\delta_{ik}(t)$ with $\phi_i(t)$.

To prove that $\epsilon_{ij}(t) \geq 0$, $\epsilon_{ij}(t)$ is rewritten as $\epsilon_{ij}(t) = A_{ij}(t) + \sum_{k\in\Theta}\int_0^t (\tilde{g}_{ik}(x) \cdot \epsilon_{kj}(t-x)dx$ where $A_{ij}(t) := \Delta_{ij}(t) + \sum_{k\in\Theta}\int_0^t \delta_{ik}(x)\pi_{kj}(t-x)dx$ Note that $A_{ij} \geq 0$, being $\Delta_{ij}(t) \geq 0$, $\delta_{ik}(x) \geq 0$, and being $\pi_{kj}(t-x)$ a probability. For any discretization step $\tau \in \mathbb{R}_{>0}$, the expression of $\epsilon_{ij}(t)$ can be rewritten by replacing $t = M \cdot \tau$ and $x = m \cdot \tau$, with $m \in [0, M]$. By induction on M, it is easily proven that $\epsilon(t)$ is monotonic non-decreasing with t. Moreover $\epsilon_{ij}(0) = A_{ij}(0) \geq 0$, which proves that $\epsilon(t) \geq 0$. □

Note that, since $0 \leq \tilde{\pi}_{ij}(t)$ for every markings i, j and time t, summation of probabilities over all reachable markings provides a defective (i.e., lower than 1) evaluation of the total probability mass properly allocated; the complement to 1 of this quantity thus comprises a safe upper bound on the maximum value of each computed probability or summation over them.

3.4 Heuristic driven approximation

The quantity $\phi_i(t)$ in Eq. 3 can be safely estimated as the sum of probabilities to reach a truncation point in the partial enumeration of the stochastic transient tree of regenerative class i. According to this, the bounds of Eq. 3 can be used to define a truncation policy in the partial enumeration of regenerative epochs that break both the enabling and the bounded regeneration restrictions (*unrestricted epochs*) with a twofold aim: adapt the error accumulated on kernel entries of each regeneration i to the impact that this epoch takes on the final error $\epsilon_{ij}(t)$; and drive the selection of truncation points within each stochastic transient tree so as to control the trade-off between complexity of enumeration and accuracy of approximation. However, exact implementation of this policy would require repeated evaluation of approximated probabilities $\tilde{\pi}_{ij}(t)$, which in turn implies a major numerical complexity for the solution of Volterra integral equations. Lemma 3 can thus be more conveniently exploited as a ground for the definition of efficient heuristics driving truncation within each regenerative epoch. Note that, while this work emphasizes the use of approximation as a way to make feasible the evaluation of kernel entries, approximation driven by efficient heuristics may be applied also to reduce complexity in epochs that fit the bounded regeneration or the enabling restrictions.

Partial exploration of unrestricted epochs is performed by initially enumerating at most ν_{start} nodes in each tree, and then by iteratively identifying a non-regenerative leaf node and by enumerating at most ν_{iter} of its successors, until the number of classes enumerated in unrestricted epochs is larger than a threshold $\nu_{\max}$ (*heuristic-based approximate analysis*). Given that the upper-bound of Eq. 3 suggests that the approximation error affects more those regenerative epochs that are visited more often, at each iteration we enumerate the successors of the non-regenerative leaf node with the largest *estimated* probability to be reached. Such estimate is evaluated by analyzing a Discrete Time Markov Chain (DTMC) $\mathbb{D}$ specified as follows:

- $\mathbb{D}$ has a state for each regenerative state class $i \in \Theta$ and for each leaf node j (either regenerative or non-regenerative) belonging to any tree $\mathcal{T}_i \in \mathcal{T}$ (regenerative and non-regenerative leaf nodes are absorbing in every tree);
- $\mathbb{D}$ has an arc from each state representing a regenerative state class $i \in \Theta$ to each state representing a leaf node j in $\mathcal{T}_i$, associated with probability μ_{ij};
- if the epoch rooted in i is analyzed exactly, μ_{ij} is equal to $G_{ij}(\infty)$ under the bounded regeneration restriction and to $G_{ij}(t_n)$ under the enabling restriction; otherwise, if the epoch rooted in i is analyzed in approximate manner, μ_{ij} is equal to $\tilde{G}_{ij}(\infty)$ or to $\tilde{L}_{ij}(\infty)$ depending on whether j corresponds to a regenerative or non-regenerative stochastic state class, respectively.

Steady-state analysis of $\mathbb{D}$ yields the vector of state probabilities P: solution relies on a basic implementation of the evaluation of absorption probabilities, which is not optimized with reference to either general techniques [2] or special techniques that might exploit warm restart in the repeated solution of DTMCs that are each a minor perturbation of the one solved at the previous iteration. Then, the steady-state probability of the states that correspond to non-regenerative leaf nodes are normalized, obtaining the vector of state probabilities $\bar{P}$, i.e., for each state l of the DTMC $\mathbb{D}$ that corresponds to a non-regenerative leaf node in a tree $\mathcal{T}_i \in \mathcal{T}$, $\bar{P}_l = P_l / \sum_{h \in \mathcal{S}_L} P_h$, where $\mathcal{S}_L$ is the set of states that correspond to non-regenerative leaf nodes in any tree $\mathcal{T}_i \in \mathcal{T}$. Finally, the non-regenerative leaf node that corresponds to the state w with the largest probability $\bar{P}_w$ is selected as the node to be expanded.

4 A case study

The approach was implemented on top of the Sirio API of the ORIS Tool [4]. Due to the minimal state space, with a single epoch requiring approximation of kernel entries, the STPN of Fig. 1 does not permit to best illustrate the potential of the approach. Hence, experiments were performed on the STPN of Fig. 3, a variant of a 3-station exhaustive-service polling system [11], where service sojourn is bounded by a DET timeout, polling times have a GEN distribution, and service times have an EXP or GEN distribution. For each station $s \in \{1, 2, 3\}$: place `Waiting`s encodes the number of pending service requests; places `AtService`s and `Vacant`s encode whether the station is being served or not, respectively; and, place `Polling`s encodes the state where the server is polling station s. In Fig. 3, all stations have no pending requests and the server is polling station 1.

The service at station s begins with the firing of transition `startService`s, with uniform distribution over $[1, 2]$, and it may terminate either when the queue of pending requests (`Waiting`s) is empty or when `timeout`s fires after a DET maximum duration of value 3. During the service interval, place `Vacant`s is

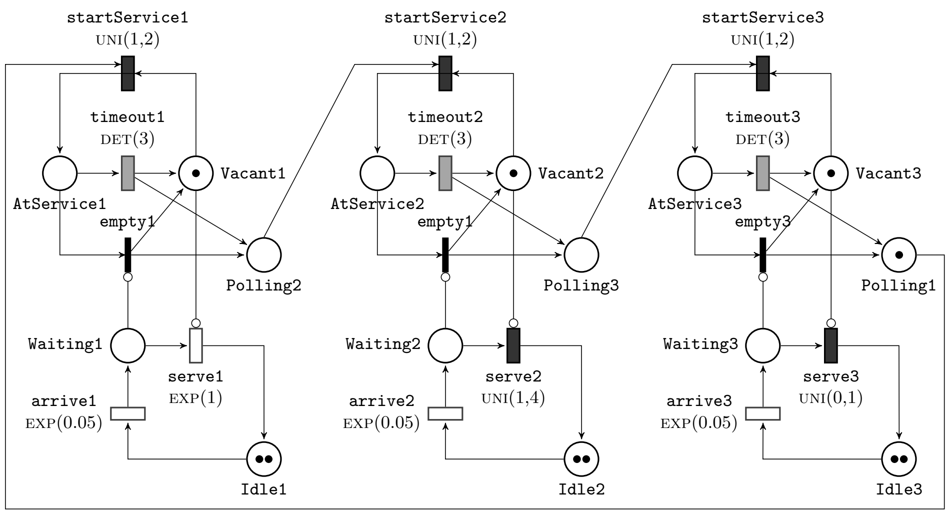


Fig. 3. STPN of a 3-station exhaustive-service polling system with server timeout.

empty and transition `serve`s is enabled, so that any number of requests can be served. Transition `arrive`s models the arrival of a new request as an EXP distribution with mean 20. Since the EXP distribution has a null minimum value, the maximum number of requests served during a service interval is limited only by the relation between the timeout value and the minimum duration of each service. Specifically, the number of requests served during a service interval sojourn is unbounded for stations 1 and 3, and it is bounded to 3 for station 2 where each service requires at least 1 time unit.

The underlying marking process regenerates whenever the server arrives to any station (i.e., at firing of `empty`s or `startService`s) or leaves it (i.e., at firing of `empty`s or `timeout`s), which directly implies that starting from any reachable state, w.p.1, a regeneration will be eventually reached, i.e. the process is an MRP. The process behavior falls in different subclasses of MRP during service sojourns at different stations. When the server is at station 1: the process satisfies the *enabling restriction*, given that `timeout1` is the only non-EXP transition enabled in each state; but it does not satisfy the bounded regeneration restriction, as for any natural number n, there exists a non-null probability that `serve1` and `arrive1` are fired more than n times before the expiration of `timeout1`. When the server is at station 2: the process satisfies the *bounded regeneration restriction*, given that `serve2` cannot be fired more than 3 times before the firing of `timeout2`; but the enabling restriction is not satisfied as `timeout2` and `serve2` can be concurrently enabled. When the server is at station 3: the process falls in the *unrestricted case* as `timeout3` and `serve3` are concurrently enabled, and `serve3` may fire an unbounded number of times before the firing of `timeout3`.

Transient analysis is performed through the approach of Sect. 3 with the following parameters: time limit $t_n = 30$ (each station is served at least twice), time step 0.1, $\nu_{\text{start}} = 20$ (number of stochastic state classes initially enumerated in each unrestricted epoch), $\nu_{\text{iter}} = 20$ (number of stochastic state classes enumerated in each unrestricted epoch at each iteration), and $\nu_{\text{max}} = 500$ (threshold on the total number of stochastic state classes enumerated in unrestricted epochs). Overall, the analysis evaluates the kernel entries of 135 regenerative epochs: 99 through the analysis under the bounded regeneration restriction, 18 through the analysis under the enabling restriction, and 18 through the heuristic-based approximate analysis. On a machine equipped with an Intel i5-5200U 2.20 GHz and 8 GB RAM, the evaluation takes nearly 40 min, spending less than 0.1 s to perform non-deterministic analysis and classification of regenerative epochs; nearly 40 s, 0.3 s, and 0.4 s to analyze the state space of regenerative epochs under the bounded regeneration restriction, under the enabling restriction, and beyond both restrictions, respectively; approximately 100 s, 180 s, and 2.5 s to evaluate the kernel entries of regenerative epochs under the bounded regeneration restriction, under the enabling restriction, and beyond both restrictions, respectively; nearly 23 s to evaluate the heuristic criterion; and, approximately 34 min to solve the Markov renewal equations. Numbers show that non-deterministic analysis has relatively negligible computational complexity, and thus it can be efficiently used to select the solution technique applied to each regenerative epoch. Notably, the heuristic criterion has a significantly lower cost with respect to the evaluation of the kernel entries of restricted epochs, which much depends on the number of encountered regenerations. Overall, results suggest that approximate analysis could be applied also to epochs under enabling or bounded regeneration restrictions to limit state space exploration and reduce evaluation complexity.

To illustrate possible rewards of interest, Fig. 4a plots the average number of messages waiting to be served at time t in each station and in the overall system, i.e., $w_n(t) = \sum_{j \in \Psi} \pi_{ij}(t) \cdot j(\texttt{Waiting}n)\ \forall n \in \{1,2,3\}$ and $w(t) = \sum_{j \in \Psi} \pi_{ij}(t) \cdot \sum_{n=1}^{3} j(\texttt{Waiting}n)$, respectively, where i is the initial regeneration (i.e., a stochastic state class with the marking of Fig. 3, where all enabled transitions are newly-enabled) and Ψ is the set of markings reached within t_n.

To evaluate the impact of different heuristics in approximate analysis, we evaluate the total defect in the evaluation of transient probabilities of markings, i.e., $\epsilon(t) := \sum_{j \in \Psi} \epsilon_{ij}(t)$ where i is the initial regeneration and Ψ the set of markings, which can be easily computed a posteriori as $\epsilon(t) = 1 - \sum_{j \in \Psi} \pi_{ij}(t)$. Fig. 4b plots the total error at the time limit $t_n = 30$ as a function of the

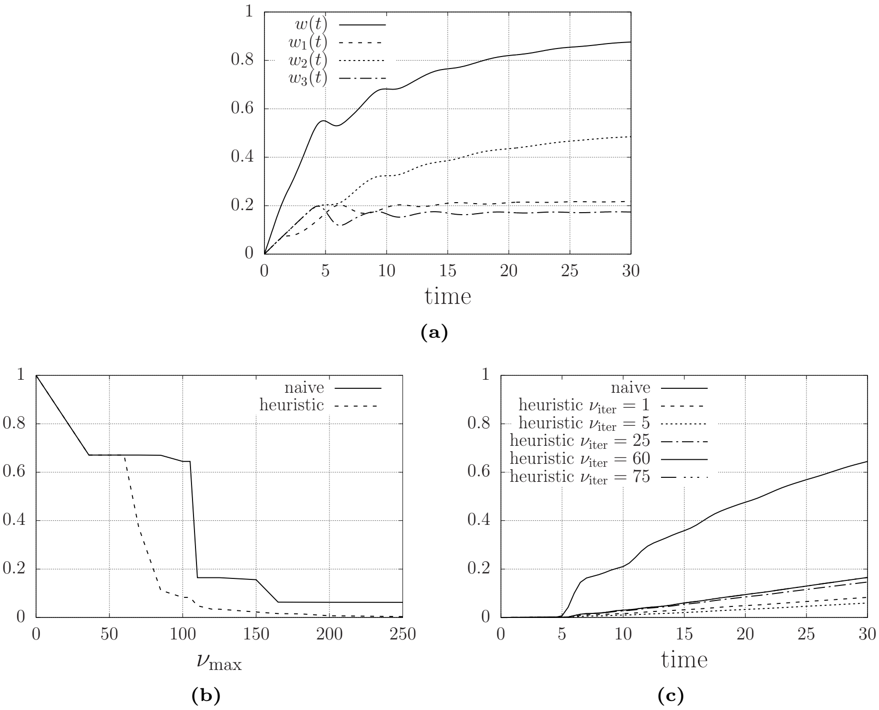


Fig. 4. (a) Average number of messages waiting to be served at time t at station s, i.e., $w_s(t)\ \forall s \in \{1,2,3\}$, and in the overall system, i.e., $w(t)$; (b) total error $\epsilon(t_n)$ (committed in the evaluation of transient probabilities of markings at the time limit t_n) as a function of the number of classes enumerated in unrestricted epochs; (c) total error $\epsilon(t)$ obtained with 70 stochastic state classes enumerated in unrestricted epochs.

threshold $\nu_{\max}$, comparing results with those obtained with a *naive approximate analysis* that explores all stochastic transient trees of unrestricted epochs, enumerating $\nu_{\max}/U$ stochastic state classes in each tree, where U is the number of unrestricted epochs. As expected, $\epsilon(t_n)$ decreases as $\nu_{\max}$ increases, and the two approaches achieve approximately the same values of $\epsilon(t_n)$ for very small values of $\nu_{\max}$. Conversely, when $\nu_{\max}$ becomes larger than 60, the heuristic-based analysis achieves significantly lower values of $\epsilon(t_n)$, in the order of $8 \cdot 10^{-2}$ for $\nu_{\max} = 100$ and $7 \cdot 10^{-3}$ from $\nu_{\max} = 200$ on, with respect to values in the order of 0.65 and $6 \cdot 10^{-2}$ attained by naive analysis, respectively. Overall, these results could be used to select a convenient value of $\nu_{\max}$ in a trade-off between the result accuracy and the computational complexity.

Fig. 4c plots the total error attained by the two approaches as a function of time, with $\nu_{\max} = 100$, $\nu_{\text{start}} = 2$, and increasing values of ν_{iter}. All curves are around zero until time 5, due to the very low probability that the server has reached station 3 by that time. From time 5 on, the error attained by naive analysis rapidly increases, being nearly 0.21, 0.48, and 0.64 at $t = 10$, $t = 20$, and $t = 30$, respectively. Conversely, $\epsilon(t)$ increases with a much smaller slope for heuristic-based analysis. As expected, the cases with lower values of ν_{iter} achieves better results; for instance, for $\nu_{\text{iter}} = 1$, $\epsilon(t)$ is approximately equal to 0.016, 0.049, and 0.083 at $t = 10$, $t = 20$, and $t = 30$, respectively. Values of $\epsilon(t)$ slightly increase with ν_{iter}, though remaining nearly in the same order of magnitude, showing that heuristic-based analysis yields sufficiently accurate results while permitting to limit the computational cost.

5 Conclusions

We leverage the low computational cost of non-deterministic analysis to drive the integration of different solution techniques in the evaluation of the kernels of an MRP, distinguishing regenerative epochs that can be analyzed through exact approaches from those that need approximate evaluation, due to infinite sequences of discrete events that never visit a regeneration. For the latter epochs, we present a novel approach based on the partial enumeration of stochastic state classes, which are iteratively explored according to a heuristic criterion based on the probability that a regeneration is reached. In so doing, the approximation is limited to the kernel entries of a subset of regenerative epochs, and transient probabilities of markings can be safely and accurately approximated.

Notably, the approximate analysis algorithm is designed to permit the integration of other solution techniques, which can be equivalently analytical or simulative. Other heuristic criteria could be used as well to select the next node to visit in partial enumeration of stochastic state classes, possibly taking into account an estimate of the mean time until when a regeneration is reached. Experimental results show that the heuristic-based approximate analysis provides accurate results while maintaining a moderate computational cost, suggesting that approximation could be used also for regenerative epochs characterized by finite stochastic transient trees, in order to reduce the number of stochastic state classes needed to compute the kernels.

References

1. Amparore, E. G., Buchholz, P., and Donatelli, S. A structured solution approach for markov regenerative processes. In *QEST* (2014), pp. 9–24.
2. Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and Van der Vorst, H. *Templates for the Solutions of Linear Systems: building blocks for Iterative Methods.* SIAM, 1994.
3. Berthomieu, B., and Diaz, M. Modeling and Verification of Time Dependent Systems Using Time Petri Nets. *IEEE Trans. SW Eng. 17*, 3 (1991), 259–273.
4. Bucci, G., Carnevali, L., Ridi, L., and Vicario, E. Oris: a tool for modeling, verification and evaluation of real-time systems. *STTT 12*, 5 (2010), 391–403.
5. Çinlar, E. Markov renewal theory: A survey. *Management Science 21*, 7 (1975), 727–752.
6. Choi, H., Kulkarni, V. G., and Trivedi, K. S. Markov regenerative stochastic Petri nets. *Perf. Eval. 20*, 1-3 (1994), 337–357.
7. Ciardo, G., German, R., and Lindemann, C. A characterization of the stochastic process underlying a stochastic Petri net. *IEEE Trans. SW Eng. 20*, 7 (1994), 506–515.
8. German, R., and Lindemann, C. Analysis of stochastic petri nets by the method of supplementary variables. *Perf. Eval. 20*, 1 (1994), 317–335.
9. German, R., Logothetis, D., and Trivedi, K. S. Transient analysis of markov regenerative stochastic petri nets: A comparison of approaches. In *Int. Workshop on Petri Nets and Performance Models* (1995), IEEE, pp. 103–112.
10. Horváth, A., Paolieri, M., Ridi, L., and Vicario, E. Transient analysis of non-Markovian models using stochastic state classes. *Perf. Eval. 69*, 7-8 (2012), 315–335.
11. Ibe, O. C., and Trivedi, K. S. Stochastic petri net models of polling systems. *IEEE Journal on Selected Areas in Communications 8*, 9 (1990), 1649–1657.
12. Kulkarni, V. *Modeling and analysis of stochastic systems.* Chap. & Hall, 1995.
13. Kwiatkowska, M., Norman, G., and Parker, D. Prism: Probabilistic symbolic model checker. In *Int. Conf. on Computer Performance Evaluation: Modelling Techniques and Tools (TOOLS)* (2002), Springer Berlin Heidelberg, pp. 200–204.
14. Lime, D., and Roux, O. H. Expressiveness and analysis of scheduling extended time Petri nets. In *IFAC Conf. on fieldbus and their appl.* (2003), Elsevier Science.
15. Lindemann, C., and Thümmler, A. Transient analysis of deterministic and stochastic Petri nets with concurrent deterministic transitions. *Perf. Eval. 36-37*, 1-4 (1999), 35–54.
16. Paolieri, M., Horváth, A., and Vicario, E. Probabilistic Model Checking of Regenerative Concurrent Systems. *IEEE Trans. SW Eng. 42*, 2 (2016), 153–169.
17. Telek, M., and Horváth, A. Transient analysis of age-mrspns by the method of supplementary variables. *Perf. Eval. 45*, 4 (2001), 205–221.
18. Vicario, E., Sassoli, L., and Carnevali, L. Using Stochastic State Classes in Quantitative Evaluation of Dense-Time Reactive Systems. *IEEE Trans. SW Eng. 35*, 5 (2009), 703–719.
19. Zimmermann, A. Modeling and evaluation of stochastic Petri nets with TimeNET 4.1. In *Int. ICST Conf. on Perf. Eval. Meth. and Tools)* (2012), pp. 54–63.