Transient analysis of generalised semi-Markov processes using transient stochastic state classes

András Horváth
Dipartimento di Informatica - Università di Torino
horvath@di.unito.it

Lorenzo Ridi, Enrico Vicario
Dip. Sistemi e Informatica - Università di Firenze
ridi@dsi.unifi.it, enrico.vicario@unifi.it

***Abstract*—The method of stochastic state classes approaches the analysis of Generalised Semi Markov Processes (GSMP) through symbolic derivation of probability density functions over Difference Bounds Matrix (DBM) zones. This makes viable steady state analysis in both discrete and continuous time, provided that each cyclic behavior that changes the enabling status of generally distributed transitions visits at least one regeneration point. However, transient analysis is supported only in discrete time.**

We extend the approach providing a way to derive continuous time transient probabilities. To this end, stochastic state classes are extended with a supplementary age clock that enables symbolic derivation of the distribution of times at which the states of a zone can be reached. The approach is amenable to efficient implementation when model timings are given by expolynomial distributions, and it can in principle be applied to transient analysis with any given time bound for any GSMP. In the special case of models underlying a Markov Regenerative Process (MRP), the method can also be applied to symbolic derivation of local and global kernels, which in turn provide transient probabilities through numerical integration of generalized renewal equations. Since much of the complexity of this analysis is due to the local kernel, we propose a selective derivation of its entries depending on the specific transient measure targeted by the analysis.

***Keywords*-Generalised Semi Markov Process; Markov Regenerative Process; transient analysis; non-Markovian Stochastic Petri Net; Stochastic Time Petri Net; stochastic state class.**

I. Introduction

Discrete event systems encountered in real-time applications usually involve multiple concurrent generally distributed timers, often supported over finite time domains. Quantitative modeling of this kind of systems is conveniently supported by non-Markovian Stochastic Petri Nets. But, in the general case, this underlies a Generalised Semi-Markov Process (GSMP) [1], [2] for which simulation is the only approach to quantitative evaluation [3].

Analytical treatment becomes viable under restrictions on the degree of concurrency of non-exponential timers. In particular, most existing techniques develop on the so-called *enabling restriction*, which assumes that at most one generally-distributed transition is enabled in any reachable tangible marking, so that activity cycles of generally distributed transitions never overlap. In this case, the model underlies a Markov Regenerative Process (MRP) which regenerates at every change in the enabling status of non-exponential timed transitions and behaves as a (subordinated) continuous time Markov chain (CTMC) between any two regeneration points [4], [2], [5], [6]. In this case, conditional probabilities evolve according to a set of generalized renewal equations, formulated as a system of integral Volterra equations of the second type [7] defined by a local and a global kernel which capture the transient probability of the states before the first regeneration point and the cumulative distribution of the time to reach the states at the first regeneration point, respectively. Under enabling restriction, both kernels can be expressed in closed-form in terms of the exponential of the matrix describing the subordinated CTMC [8] and evaluated numerically through uniformisation. Generalised regeneration equations can then be solved by numerical approaches in the time domain or through Laplace-Stieltjes transform.

In principle, the analysis of a model with multiple concurrent generally distributed timers can be formulated through the theory of *supplementary variables* [9], [6], [10], extending the logical state (the marking) with the vector of ages of generally distributed enabled transitions. However, this results in a set of integro-differential equations whose practical solution is limited to one or two concurrently enabled non-exponential distributions, thus falling again within the limits of the enabling restriction [11]. The limit of the enabling restriction is overcome in [12], [13] but only for the case that all timed transitions are either exponential or deterministic. In this case, sampling the process at equidistant time points yields a General State Space Markov Chain whose kernel of transition probabilities is derived through numerical integration. Limiting and transient probabilities of tangible markings are derived by numerical integration of steady-state and transient equations taking the form of a set of Volterra differential equations [12] and a set of Fredholm integral equations [13], respectively.

More recently, [14], [15], [16] proposed a new analytical approach that manages the case of multiple concurrent generally distributed timers with possibly overlapping activity cycles, provided that every cyclic behavior that changes the enabling status of generally distributed transitions visits at least one regeneration point [15]. The approach relies on the stochastic expansion of non-deterministic state-classes described by Difference Bounds Matrix (DBM) zones [17], [18], [19] which are commonly employed in qualitative verification tools. This yields so-called stochastic state classes characterising sojourn time distributions which are organised then in a graph that explicitly describes transitions among classes. The stochastic state class graph abstracts the behaviour of a GSMP into an embedded Discrete Time Markov Chain (DTMC) that samples the process after each transition and thus allows evaluation of discrete time transient and steady state measures referred to the number of fired transitions. In addition, the distribution associated to stochastic classes allows derivation of average sojourn times between subsequent transitions and thus enables evaluation of steady probabilities of the overall continuous time process [16]. However, for what pertains to continuous time transient behaviour, the approach only supports derivation of the cumulative probability distribution of the time elapsed along a selected firing sequence that starts from a regenerative class where the process looses memory of its past history [15], [20]. This comprises a major limit for application in design and verification of real time systems where non-functional requirements are natively concerned with behavior within the short-term of deadlines or periods.

In this paper, we extend the approach of stochastic state classes so as to support derivation of continuous time transient probabilities. To this end, stochastic state classes are extended with a supplementary age clock that keeps track of the time elapsed from the initial state class. This makes class density functions be dependent on the time at which they are entered, and thus enables symbolic derivation of the distribution of times at which the states of a zone can be reached. The approach is amenable to efficient implementation when temporal parameters in the model are given by expolynomial distributions, and it can in principle be applied to transient analysis with any given time bound for any GSMP, regardless of the existence of regeneration points. In the special case of models underlying a Markov Regenerative Process (MRP), the method can be applied to the symbolic derivation of local and global kernels, also when regeneration periods break the limit of a subordinated CTMC and rather evolve according to complex stochastic structures with multiple concurrent non-exponential transitions. Transient probabilities can thus be derived through numerical integration of classical generalized renewal equations. In so doing, the local kernel turns out to be the major source of complexity, both for the number of non-null entries and for the fact that its entries are derived from multivariate distributions whenever multiple generally distributed transitions are concurrently enabled. We show that this complexity can be reduced through a structured approach to the derivation of transient probabilities of specific states that minimizes the size of the local kernel depending on the specific transient measure targeted by the analysis.

The rest of the paper is organised as follows. In Sect. II, we recall syntax and semantics of a variant of non-Markovian Stochastic Petri Nets, which we call stochastic Time Petri Nets (sTPN) to emphasise the interest on concurrent generally distributed transitions with possibly bounded supports. For the sake of simplicity, the treatment does not consider immediate and deterministic transitions, which could be encompassed in the theory by referring to the partitioned form of DBM domains [15]. In Sect. III, we introduce transient stochastic state classes, develop the calculus for their derivation and show how they support the derivation of transient probabilities. In Sect. IV, we discuss the application of the approach to MRPs showing how the symbolic form of the local and the global kernel can be derived. We illustrate how the local kernel can be reduced considering the desired transient measure. In Sect. V, we report computational experience based on a preliminary tool-chain of the Oris tool [15] and Wolfram Mathematica. Conclusions are drawn in Sect. VI.

II. Stochastic Time Petri Nets

A stochastic Time Petri Net (sTPN) [15], [16] is a tuple:

$$sTPN = \langle P, T, A^-, A^+, m_0, EFT, LFT, \mathcal{F}, \mathcal{C} \rangle$$

As in time Petri nets [21], [18], [17]: P is a set of places; T a set of transitions; $A^- \subseteq P \times T$ and $A^+ \subseteq T \times P$ are the usual pre- and post-conditions; m_0 is the initial marking. Firing times are characterised by $EFT : T \to \mathbb{R}_0^+$ and $LFT : T \to \mathbb{R}_0^+ \cup \{+\infty\}$ which associate each transition with an earliest and a latest firing time and by $\mathcal{F}$ providing for each transition $t \in T$ a probability distribution $F_t()$ supported in the firing interval $[EFT(t), LFT(t)]$. In addition, $\mathcal{C}$ associates each transition with a real-valued weight used to resolve conflicts.

The state of an sTPN is a pair $s = \langle m, \tau \rangle$, where $m : P \to \mathbb{N}$ is a marking and $\tau : T \to \mathbb{R}_0^+$ associates each transition with a time to fire. A transition t_0 is *enabled* if each of its input places contains at least one token, and it is *firable* if it is enabled and its time to fire $\tau(t_0)$ is not higher than that of any other enabled transition. When multiple transitions are firable, the choice is resolved through a random switch determined by the weight $\mathcal{C}$ as follows: $P\{t_0 \text{ is selected}\} = \frac{\mathcal{C}(t_0)}{\sum_{t_i \in T^f(s)} \mathcal{C}(t_i)}$ where $T^f(s)$ is the set of transitions that are firable in state s.

When a transition t_0 fires, the state $s = \langle m, \tau \rangle$ is replaced by $s' = \langle m', \tau' \rangle$, which we write as $s \xrightarrow{t_0} s'$. Marking m' is derived from m by removing a token from each input place of t_0, and by adding a token to each output place of t_0:

$$\begin{aligned} m_{tmp}(p) &= \begin{cases} m(p) - 1 & \text{if } \langle p, t_0 \rangle \in A^- \\ m(p) & else \end{cases} \\ m'(p) &= \begin{cases} m_{tmp}(p) + 1 & \text{if } \langle t_0, p \rangle \in A^+ \\ m_{tmp}(p) & else \end{cases} \end{aligned} \quad (1)$$

Transitions that are enabled both by the intermediate marking m_{tmp} and by m' are said *persistent*, while those that are enabled by m' but not by m_{tmp} are said *newly enabled*. If t_0 is still enabled after its own firing, it is always regarded as newly enabled [18], [17]. For any transition t_i that is persistent after the firing of t_0, the time to fire is reduced by the time elapsed in the previous state:

$$\tau'(t_i) = \tau(t_i) - \tau(t_0) \tag{2}$$

Whereas, for any transition t_a that is newly enabled after the firing of t_0, the time to fire takes a random value sampled in the firing interval according to $F_{t_a}()$:

$$\begin{array}{c} EFT(t_a) \leq \tau'(t_a) \leq LFT(t_a) \\ P\{\tau'(t_a) \leq x\} = F_{t_a}(x) \end{array} \tag{3}$$

Without loss of generality, we assume that $EFT(t) < LFT(t)$, which rules out immediate and deterministic transitions. These can be encompassed in the theory by resorting to a formulation based on the partitioned form of DBM zones described in [15]. As a side-effect of the assumption, $\mathcal{C}$ becomes irrelevant to the purposes of the analysis [15]. Still without loss of generality, we assume that $F^t()$ is absolutely continuous so that it can be expressed as the integral function of a probability density function $f_t()$ as

$$F_t(x) = \int_0^x f_t(y)dy \tag{4}$$

III. Transient stochastic state classes

State classes are largely employed in qualitative verification of models with non-deterministic continuously timed parameters as an abstraction that aggregates states reached through different timings of the same sequence of transitions. For models such as Time Petri Nets and Timed Automata, this reduces state space analysis to the enumeration of a succession relation among domains encoded as the space of solutions of sets of simple linear inequalities usually known as Difference Bounds Matrixes (DBM) [19], [22], [17].

Stochastic state classes extend the approach by decorating each state class with a probability density function supported over the DBM domain, that characterizes the distribution of times to fire of enabled transitions at the time when the class is entered [15].

Transient stochastic classes further extend the concept through the introduction of an additional age clock capturing the absolute time at which the class is entered. The distribution of the age clock provides a means to recover the distribution of probability of the absolute time at which the class can be reached, thus opening the way to the evaluation of transient probabilities. In so doing, the reachability graph of state classes is expanded into a transient stochastic graph, where a single state class can be expanded in multiple stochastic classes with the same marking and DBM domain but with different joint probability distributions of age clock and times to fire.

For technical convenience, the age clock encodes the opposite of the absolute time at which a class can be reached. In so doing the age is decreased as time advances. In so doing, the age follows the same slope of times to fire and domains remain encoded in the shape of a DBMs.

A. Transient Stochastic State Classes

Definition 3.1: A *state class* is a pair $S = \langle m, D\rangle$ where m is a marking and D is a (possibly infinite and continuous) set of values for the vector $\underline{\tau} =< \tau_0, \tau_1, ..., \tau_{N-1} >$ of times to fire of transitions enabled by m, with τ_i denoting the time to fire of transition t_i.

Definition 3.2: $S_2 = \langle m_2, D_2\rangle$ is the *successor* of $S_1 = \langle m_1, D_1\rangle$ through the firing of t_0, which we write as $S_1 \overset{t_0}{\to} S_2$, iff t_0 is firable in some state contained in S_1 and S_2 contains all and only the states that can be reached from some states in S_1.

Definition 3.3: Given an initial state class S_0, the relation $\overset{t_0}{\to}$ identifies a *state class graph* $\langle V, E\rangle$ where the set of vertices V is the transitive closure of the set of classes reachable through $\overset{t_0}{\to}$ starting from S_0 and the set of edges is E with $E \subseteq V \times T \times V$ containing a given triple $\langle S_1, t_0, S_2\rangle$ iff $S_1 \overset{t_0}{\to} S_2$.

Definition 3.4: A *transient stochastic state-class* (transient class for short) is a triple $\langle m, D, f_{<\tau_{age},\underline{\tau}>}()\rangle$ where m is a marking, $< \tau_{age}, \underline{\tau} >$ is a random variable called *clock vector* made up by the scalar variable τ_{age} representing the opposite of the time elapsed together with the vector $\underline{\tau} =< \tau_0, \tau_1, ..., \tau_{N-1} >$ representing the times to fire of transitions enabled by m, and $f_{<\tau_{age},\underline{\tau}>}()$ is the probability density function of $< \tau_{age}, \underline{\tau} >$ over the support D.

Definition 3.5: Given two transient stochastic classes $\Sigma = \langle m, D, f_{<\tau_{age},\underline{\tau}>}()\rangle$ and $\Sigma' = \langle m', D', f_{<\tau'_{age},\underline{\tau}'>}()\rangle$, we say that Σ' is a successor of Σ through t_0, and we write $\Sigma \overset{t_0}{\Longrightarrow} \Sigma'$, iff the following property holds: if the marking of the net is m and the clock-vector is a random variable $< \tau_{age}, \underline{\tau} >$ with support D and probability density function $f_{<\tau_{age},\underline{\tau}>}()$, then t_0 has a non-null probability to fire, and if it fires then the model reaches a new marking m' and a new clock-vector $< \tau'_{age}, \underline{\tau}' >$ with support D' and probability density function $f_{<\tau'_{age},\underline{\tau}'>}()$.

The transient stochastic class Σ' reached from Σ through t_0 can be derived following the steps of [15], with changes needed to account for the fact that τ_{age} does not behave as the time to fire of an enabled transition and rather encodes the opposite of the transient age. In the derivation, we use the two following notational conventions: *i)* if $D \subseteq \mathbb{R}^N$ and $n \in [0, N-1]$, then $D \downarrow_n$ denotes the projection of D that eliminates τ_n (i.e. $D \downarrow_n = \{\langle x_0, ...x_{n-1}, x_{n+1}, ..., x_{N-1}\rangle | \exists x_n \in \mathbb{R}$ such that $\langle x_0, ...x_{n-1}, x_n, x_{n+1}, ..., x_{N-1}\rangle \in D\}$); *ii)* given a vector $\underline{x} = \langle x_0, ..., x_{N-1}\rangle$ and a scalar δ, $\underline{x} + \delta$ denotes the vector $\langle x_0 + \delta, ..., x_{N-1} + \delta\rangle$. Let

$< \tau_{age}, \underline{\tau} >=< \tau_{age}, \tau_0, \tau_1, ...\tau_{N-1} >$ be the clock vector in the class Σ distributed over $D \subseteq \mathbb{R}_0^- \times {\mathbb{R}_0^+}^N$ according to $f_{<\tau_{age},\underline{\tau}>}(x_{age}, x_0, x_1, ..., x_{N-1})$. If firing of t_0 is possible (i.e. if the support of $< \tau_{age}, \tau_0, \tau_1, ...\tau_{N-1} >$ has a non null intersection with the set $\tau_0 \leq \tau_n \ \forall n \in [1, N-1]$), the distribution of the clock vector after the firing is derived through the following four steps:

i) Precedence conditioning: the assumption that t_0 fires before any other transition conditions the clock vector and thus yields a new random variable $< \tau^a_{age}, \underline{\tau}^a >=< \tau_{age}, \underline{\tau} | \tau_0 \leq \tau_n, n = 1, N-1 >$ distributed over $D^0 = D \bigcap \{\tau_0 \leq \tau_n, n = 1, N-1\}$ according to:

$$f_{<\tau_{age},\underline{\tau}^a>}(x_{age}, \underline{x}) = \frac{f_{<\tau_{age},\underline{\tau}>}(x_{age}, \underline{x})}{\int_{D^0} f_{<\tau_{age},\underline{\tau}>}(x_{age}, \underline{x}) dx_{age} d\underline{x}} \quad (5)$$

ii) Time advancement and projection: when t_0 fires, the age and the times to fire of all the other transitions are reduced by the time to fire value of t_0 given by τ^a_0, and τ^a_0 itself is eliminated through a projection. This yields a variable $< \tau^b_{age}, \underline{\tau}^b >=< \tau^a_{age} - \tau^a_0, \tau^a_1 - \tau^a_0, ..., \tau^a_{N-1} - \tau^a_0 >$, distributed over $D^b = D^0 \downarrow_0$ according to:

$$f_{<\tau^b_{age},\underline{\tau}^b>}(x_{age}, \underline{x}) = \int_{Min^0(x_{age},\underline{x})}^{Max^0(x_{age},\underline{x})} f_{<\tau^a_{age},\underline{\tau}^a>}(x_{age} + x_0, x_0, \underline{x} + x_0) dx_0 \quad (6)$$

where $\underline{x} = \langle x_1, ..., x_{N-1} \rangle$, and the interval $[Min^0(x_{age}, \underline{x}), Max^0(x_{age}, \underline{x})]$ provides the support representing all possible values of τ_0.

iii) Disabling: if a single transition, say t_1, is disabled at the firing of t_0, its elimination yields a new variable $< \tau^c_{age}, \underline{\tau}^c >=< \tau^b_{age}, \tau^b_2, ..., \tau^b_{N-1} >$, distributed over $D^c = D^b \downarrow_1$ according to:

$$f_{<\tau^c_{age},\underline{\tau}^c>}(x_{age}, x_2, ...x_{N-1}) = \int_{Min^1(x_{age},\underline{x})}^{Max^1(x_{age},\underline{x})} f_{<\tau^b_{age},\underline{\tau}^b>}(x_{age}, x_1, \underline{x}) dx_1 \quad (7)$$

where $\underline{x} = \langle x_2, ..., x_{N-1} \rangle$. When multiple transitions are disabled at the firing of t_0, the same step is repeated for each of them.

iv) Newly enabling: if a single transition t_N with firing density function $f_N(x_N)$ over the support $[EFT_N, LFT_N]$ is newly enabled at the firing of t_0, then the clock vector $< \tau'_{age}, \underline{\tau}' >=< \tau^c_{age}, \tau^c_2, ..., \tau^c_{N-1}, \tau_N >$ of the destination class Σ' is distributed over $D' = D^c \times [EFT_N, LFT_N]$ according to:

$$f_{<\tau^d_{age},\underline{\tau}^d>}(x_{age}, \underline{x}, x_N) = f_{<\tau^c_{age},\underline{\tau}^c>}(x_{age}, \underline{x}) \cdot f_N(x_N) \quad (8)$$

where $\underline{x} = \langle x_2, ..., x_{N-1} \rangle$. When multiple transitions are newly enabled, the product form is extended with a factor for each of them.

B. The graph of transient stochastic state classes

Definition 3.6: Given an initial transient class Σ_0, the relation $\overset{t_0}{\Rightarrow}$ identifies a *transient stochastic class graph* $\langle V, E \rangle$ where: the set of vertices V is the transitive closure of the set of classes reachable through $\overset{t_0}{\Rightarrow}$ with $\Sigma_0 \in V$; the set of edges E is $E \subseteq V \times T \times V$ and $\langle \Sigma_1, t_0, \Sigma_2 \rangle$ is part of E iff $\Sigma_1 \overset{t_0}{\Rightarrow} \Sigma_2$.

If the vector of times to fire $\underline{\tau}$ in the initial class Σ^0 is supported over a DBM domain D, then according to the properties of DBM zones (e.g. [19], [22], [17]) the domains of all reached classes are still in DBM form. Moreover, if static density functions of transitions in the STPN model are (monovariate) continuous functions, then density functions of reached transient stochastic classes are (multivariate) continuous piecewise functions with analytic representation over a finite partition of the support made by DBM subdomains [23]. In the special case that the static density function $f_t()$ associated with each model transition t is an expolynomial function (i.e. $f_t(y) = \sum_{k=1}^{K} c_k y^k e^{-\lambda_k y}$), then density functions of reachable transient classes accept a closed-form, which is efficiently computed in the Oris tool [15] through repeated application of the calculus of Sect. III-A within a conventional forward enumeration (semi-)algorithm.

Conditions that guarantee termination can be conveniently characterised with reference to the state class graph, due to the following univocal relation from transient stochastic classes to state classes:

Lemma 3.1: The transient stochastic graph includes a vertex $\Sigma = \langle m, D, f() \rangle$ iff the class graph includes a vertex $S = \langle m, D \downarrow_{\tau_{age}} \rangle$. Moreover, $\Sigma \overset{t}{\Rightarrow} \Sigma'$ with $\Sigma' = \langle m, D, f() \rangle$ iff $D \downarrow_{\tau_{age}} \overset{t}{\rightarrow} D' \downarrow_{\tau_{age}}$.

Proof: The proof runs by induction. *i)* By construction the support of times to fire in the initial vertex of the transient stochastic graph is the DBM domain in the initial vertex in the state class graph. *ii)* Besides, if Σ is a transient stochastic class with support D and S is a state class with domain D, then t has a non-null probability to occur from Σ iff it is an outgoing event from S (both conditions occur iff D has a non-empty subset where $\tau_0 \leq \tau_n \ \forall n \in [1, N]$). Moreover, the marking and the support of times to fire in the transient stochastic class Σ' reached from Σ through the firing of t_0 are by construction equal the marking and the domain in the class reached from S through the firing of t_0. ■

Lemma 3.1 permits to regard the state class graph as a kind of non-deterministic projection of the transient stochastic graph, or vice-versa, regard the transient stochastic graph as a stochastic and aged expansion of the class graph. As a particular consequence, any trace ρ in the transient class graph identifies a unique trace in the class graph, which we call projection. This permits to characterise a condition that rules out Zeno behaviours and guarantees finiteness in the number of transient classes that may be reached within any bounded age:

Lemma 3.2: If the state class graph is finite, and every cycle that it contains traverses at least one edge labeled by a transition t with non-null static earliest firing time (i.e. $EFT(t) > 0$), then, for any bound T, the number of transient stochastic classes including states with $(-\tau_{age}) \leq T$ is finite.

Proof: Ab absurdo, let ρ be an infinite path in the transient graph and let each class visited by ρ include at least one state satisfying $(-\tau_{age}) \leq T$. By transitive closure of Definition 3.5, a transient class includes a state s satisfying $(-\tau_{age}) \leq T$ iff the model accepts a behaviour that can reach s within a time not higher than T. According to this, ρ accepts a timing that never exceeds the bound T.

However, since the state class graph is finite, it must include a state class S_* that is visited infinitely often by the projection of ρ. The trace ρ can thus be decomposed into an infinite sequence of sub-traces $\{\rho_i\}_0^\infty$, and the projection of each sub-trace ρ_i on the class graph is a cycle that starts and terminates on class S_*. By hypothesis, each such cycle includes the firing of at least one transition with non-null static earliest firing time. Since transitions are finite, there is at least one transition with a non-null earliest firing time which is fired infinitely often along ρ, which contradicts the assumption that ρ accepts a timing that never diverges beyond T. ■

For the specific purposes of the subsequent treatment of this paper we are also interested in characterising the conditions that guarantee a finite number of transient classes before reaching a regeneration point where all times to fire values are independent from previous history and the age clock τ_{age} is the only memory-carrying variable.

Definition 3.7: We call *regenerative* a transient stochastic class where all times to fire are either newly enabled, or exponentially distributed, or deterministic, or bounded to take a deterministic delay with respect to a time to fire satisfying any of the previous conditions.
The conditions for a transient class $\Sigma = \langle D, m, f() \rangle$ to be regenerative are completely determined by the underlying state class $S = \langle D, m \rangle$ and by the condition of newly-enabling/persistence (somewhere referred to as restart/continue) of non-exponential transitions. In particular, it is relevant to note here that the same marking can be reached both in regenerative and non-regenerative classes.

A condition sufficient for guaranteeing termination in the enumeration of transient stochastic classes can be conveniently characterized in terms of cyclic behaviors occurring in the graph of underlying state classes:

Lemma 3.3: If the class graph is finite and every cycle that it contains visits at least one regenerative state class, then the number of transient classes visited before reaching the first regenerative transient stochastic class is finite.

Proof: Ab absurdo, let ρ be an infinite path in the transient stochastic graph that never reaches a regeneration class. Following the same steps of the proof of Lemma 3.2, ρ includes a finite prefix $\rho_{pre} \rightarrow \rho_{cycle}$ such that the projection of ρ_{cycle} on the class graph is a cycle. By hypothesis, this cycle visits a regeneration state class S_r, which implies that within the (finite) termination of $\rho_{pre} \rightarrow \rho_{cycle}$ the enumeration has visited a transient stochastic regeneration class. ■

The hypothesis of Lemmas 3.2 and 3.3 can apparently be relaxed in various directions, which cannot expand here. In particular, the special case of cyclic behaviours due to the firing of immediate or exponential transitions is attackable through the aggregation of transient classes visited along firing sequences that do not change the enabling status of generally distributed transitions. This issue was developed for steady state analysis in [16].

C. Derivation of transient probabilities

The transient stochastic graph abstracts the behaviour of an sTPN into a discrete state continuous time process $\mathbb{X} = \{X(t) = \Sigma_i,\ t \in \mathbb{R}_0^+\}$ (with $X(t) = \Sigma_i$ meaning that at time t, Σ_i is the most recently entered transient stochastic class). Transient analysis of continuous time behaviour of the model can thus be reduced to the evaluation of transient probabilities $\pi_i(t)$ associated with transient stochastic state classes:

$$\pi_i^\Sigma(t) = P\{X(t) = \Sigma_i\} \tag{9}$$

The embedded chain that samples the process at the firing of each transition $\mathbb{X}^e = \{X^e(n) = \Sigma_i,\ n \in \mathbb{N}\}$ (with $X^e(n) = \Sigma_i$ meaning that after n firings, Σ_i is the most recently entered transient stochastic class) is a DTMC as a direct consequence of Def. 3.5. Analysis of $\mathbb{X}^e$ provides the discrete distribution of the probability that each transient stochastic class is eventually reached:

$$\eta_i^\Sigma = P\{\exists n \in \mathbb{N}.\ X(n) = \Sigma_i\} \tag{10}$$

To obtain transient probabilities of the continuous time process, $\pi_i^\Sigma(t)$ can be expressed as the probability that Σ_i is entered at some time u not higher than t and the sojourn time after that entrance is not lower than $t - u$. Now, in the clock vector $\langle \tau_{age}, \underline{\tau} \rangle$ of Σ_i, the variable τ_{age} encodes the opposite of the time at which Σ_i was entered, and $\underline{\tau}$ encodes the vector of times to fire of enabled transitions. The sojourn time in Σ_i can be characterised by $min\{\underline{\tau}\} \stackrel{def}{=} \min_{h \in [0,N-1]}\{\tau_h\}$. Probabilities $\pi_i^\Sigma(t)$ can thus be derived from the state density function $f^i_{\langle age, \underline{\tau} \rangle}(x_{age}, \underline{x})$ of Σ_i as

$$\pi_i^\Sigma(t) = \eta_i^\Sigma \int_{\substack{u \in [0,t] \\ u + min\{\underline{x}\} > t}} f^i_{\langle age, \underline{\tau} \rangle}(-u, \underline{x}) du\ d\underline{x} \tag{11}$$

Aggregation of $\pi_i^\Sigma(t)$ over the set of transient stochastic classes that share a common marking provides transient probabilities of reachable markings:

$$P\{M(t)=m\}=\pi_m(t)=\sum_{M(\Sigma_i)=m}\pi_i^\Sigma(t) \qquad (12)$$

where $M(\Sigma_i)$ denotes the marking of the transient class Σ_i.

IV. Exploiting regeneration points

The analysis approach proposed in Sect. III derives transient probabilities directly from probability density functions of enumerated transient classes. This is blind to the presence of regeneration points and can thus be applied to any GSMP. However, in so doing, the analysis does not take any advantage from possible regeneration points after which the process repeats itself in a stochastic sense.

An analysis technique that is based on regeneration points is provided by the theory of MRPs and it is based on the stochastic description of a regeneration period (the interval between two consecutive regeneration points). In Sect. IV-A we show that transient stochastic state classes are applicable also to the analysis of a regeneration period. The full analysis of a regeneration period can require a lot of calculation. For this reason, in Sect. IV-B we propose two techniques that, by considering the transient measure to be computed, apply a selective and, consequently, less heavy analysis of regeneration periods.

A. Analysis of a regeneration period by transient stochastic state classes

We denote by $Z(t)$ the marking of the process at time t, by $S_0=0,S_1,S_2,\ldots$ the successive regeneration time instants and by $Y_0=Z(0),Y_1=Z(S_1+),Y_2=Z(S_2+),\ldots$ the markings reached right after regenerations.

The theory of MRPs applies two quantities for the description of a regeneration period [24]. What happens inside the period is described by the local kernel whose entries are given by the probability

$$L_{ij}(t)=P\left(Z(t)=j,S_1>t|Y_0=i\right) \qquad (13)$$

i.e., by the probability that, having started in marking i, no regeneration has occured up to time t and the marking at time t is j. In order to characterise where the process ends up right after regeneration, the global kernel is used whose entries are defined as

$$G_{ij}(t)=P\left(Y_1=j,S_1\leq t|Y_0=i\right) \qquad (14)$$

which corresponds to the probability that, having started in marking i, the first regeneration period is not longer than t time units and the marking reached by the first regeneration is marking j. It is easy to see that we have

$$\sum_j L_{ij}(t)+G_{ij}(t)=1,\ \forall i,\forall t. \qquad (15)$$

Based on the two kernels, conditional transient probabilities can be expressed as the solution of generalized renewal equations [7] that take the form of a set of Volterra integral equations of the second type and can be solved in various approaches developed for the evaluation of MRPs:

$$\begin{array}{l}P(Z(t)=j|Z(0)=i)=\\ L_{ij}(t)+\sum_k\int_{x=0}^t G_{ik}(x)P(Z(t-x)=j|Z(0)=k)dx\end{array} \qquad (16)$$

Transient stochastic state classes can be applied to derive the analytical form of the entries of the kernels $L_{ij}(t)$ and $G_{ij}(t)$ given in (13) and (14). Naturally, in order to determine those entries of the kernels that correspond to a regeneration period starting in marking i, we have to perform the enumeration of the transient stochastic state classes starting from marking i. The difference with respect to the enumeration described in Sect. III is that we are interested only in a single regeneration period. Accordingly, we stop the generation of successors in case we arrive to a regeneration point. A regeneration point is reached upon entrance into a class where all active transitions are newly enabled or exponential and such class will be called regenerative. In order to simplify the analysis, having reached a regenerative class we do not even perform the enabling of the newly enabled transitions. Consequently, a regenerative class will be characterised by a density function of a single variable describing the distribution of the elapsed time. Those transient stochastic state classes that are regenerative contribute to the global kernel while the others to the local kernel. Having performed the enumeration starting from marking i the contribution of the resulting classes in the kernels can be calculated based on (11) as

$$L_{ij}(t)=\sum_{k:\Sigma_k \text{ is not regenerative and } M(\Sigma_k)=j}\pi_k^\Sigma(t) \qquad (17)$$

$$G_{ij}(t)=\sum_{k:\Sigma_k \text{ is regenerative and } M(\Sigma_k)=j}\pi_k^\Sigma(t) \qquad (18)$$

In order to fully characterise the kernels the enumeration has to be performed starting from the initial marking and from all the other markings that can be reached by regeneration.

Once the kernels are characterised, solution techniques developed for MRPs can be applied to calculate the desired transient or steady state measures. For a survey of solution techniques see [25].

B. Selective analysis of a regeneration period

The integral given in (11), which is used then in (17) and (18) to determine the entries of the kernels, presents different levels of difficulty for regenerative classes and for non regenerative classes. For a regenerative class the density function depends on a single variable corresponding to the elapsed time and hence the integral is performed according to this single variable. For what concerns non regenerative classes the integration with respect to the time to fire variables requires to consider a potentially complicated domain, described by a DBM, over which the state density function is different from zero. For this reason it is convenient to perform the analysis with as few entries of the local kernel as possible. In the following we discuss two simple cases in which the number of necessary entries of the local kernel can be kept low.

In the first case, assume that we want to compute the distribution of the time to reach a given marking for the first time, i.e., our goal is to determine a first passage time distribution. This problem can be faced without computing any entry of the local kernel as follows. We modify the model in such a way that the considered marking becomes absorbing, i.e., the model gets blocked as soon as the marking is reached. This way the time instance of entering this marking is a regeneration point. Consequently, it is not necessary to describe what happens inside a regeneration period and the global kernel of the modified model is enough to characterise the desired first passage time distribution. Having computed the entries of the global kernel, $G_{ij}(t)$, the fact that the process is not described inside a regeneration period can be formally introduced by setting the entries of the local kernel as

$$L_{ij}(t) = \begin{cases} 1 - \sum_k G_{ik}(t) & i = j \\ 0 & i \neq j \end{cases}$$

which implies that the process remains in the same marking for the whole regeneration period. Note that this way we simplified the original MRP into a semi-Markov process (SMP) and hence the computation itself can be carried out by techniques for SMPs.

In the second case, assume that we are interested in the transient probabilities of a single marking, in particular, marking 1. In order to compute the desired probability, we need to compute all entries of the global kernel, $G_{ij}(t)$, and those entries of the local kernel that leads to marking 1, $L_{i1}(t)$. The remaining entries of the local kernel can be simply set to

$$L_{ij}(t) = \begin{cases} 1 - \sum_k G_{ik}(t) - L_{i1}(t) & i = j \\ 0 & i \neq j \end{cases} \qquad j \neq 1$$

because inside a regeneration period all we have to know is if the process is in marking 1 or somewhere else.

By performing the analysing in the above described manner we decrease the number of local kernel entries that have to be computed. The size of the kernels can also get reduced if there exist markings that are not involved in the analysis objectives and do not play a role in the global kernel.

V. Computational experience

In Sects. V-A and V-B, we discuss the application of the approach on two examples, the former relying on straight enumeration of transient classes and the latter developing on regeneration points. Both examples include a low number of markings, but they still comprise non-toy cases that underly a complex stochastic process whose transient analysis has not been performed before.

Numerical results were obtained through a preliminary implementation in the Oris tool [15]. Wolfram Mathematica was also employed to solve generalized renewal equations in the analysis based on regeneration points, and to validate all obtained results through stochastic simulation.

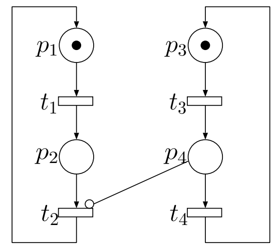

Figure 1. Petri net representation of the preemptive G/G/1/2/2 queue.

A. Example without regeneration points

Fig. 1 depicts the Petri net model of the G/G/1/2/2 preemptive queue. There are two clients in the system and their arrival is represented by transitions t_1 and t_3. Service, represented by transitions t_2 and t_4, is preemptive in such a way that the arrival of client 2 interrupts the service of client 1. The interruption is according to the preemptive repeat different policy which means that the service of client 1 restarts from crash when transition t_2 gets enabled again. All transitions are with uniformly distributed firing times on the interval [1,2].

A simpler version of the model, namely, the M/G/1/2/2 preemptive queue has been analysed in several papers dealing with non-Markovian stochastic processes with non-overlapping activity cycles. If the arrival times are not exponential the activity cycles overlap. Moreover, with the seemingly simple timings chosen for the transitions, the length of the regeneration periods have a distribution with infinite support. This means that there are markings starting from which infinite firing sequences without regeneration points can occur. Even this situation can be handled by the technique proposed in the paper. In order to determine the transient behaviour in the interval $[0, T]$, we have to generate all the classes in which the support of the density is such that the elapsed time is between 0 and T with positive probability. We experimented with $T = 8$ and there are about 100 classes where the elapsed time can be less than 8. The generation of these classes requires 10 seconds with a 2GHz processor and 2GB RAM and they directly provide the transient behaviour of the model for the interval [0,8]. The transient probabilities are depicted in Fig. 2.

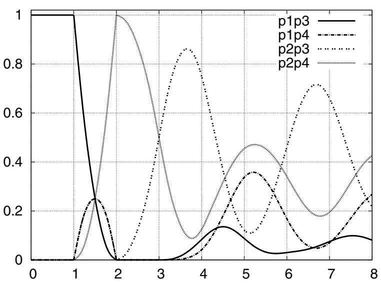

Figure 2. Transient probabilities of the preemptive G/G/1/2/2 queue.

B. Example with regeneration points

A model in which the distribution of the length of the regeneration periods has finite support is depicted in Fig. 3. This model, introduced in [15], represents a system composed of two production cells which operate in alternate manner passing control to each other (transitions start1 and start2). The two cells are identical with the single exception that transitions start1 and start2 have different earliest firing times. Both cells are composed of two parallel activities, named JobA and JobB, which are associated with uniform duration. JobA in both cells requires a resource called res which may fail during usage according to an exponential distribution with rate parameter 0.3. If failure occurs then JobA is interrupted and a recovery activity (recA) together with a repair action is started. Both the recovery and the repair require a duration with uniform distribution.

We concentrate on the situation when both resources are in the failed state (i.e., when both place failed1 and failed2 are marked) and compute two related performance indices.

First, we consider the distribution of the time to reach double failure for the first time. As it is described in Sect. IV-B, this kind of measure can be computed without computing any entry of the local kernel. Originally there are 6 markings in the model that can be reached through regeneration. We stop the model as soon as double failure is reached and this adds a 7th regenerative marking (an absorbing one) to the model. Accordingly, the global kernel is 7 by 7. The enumeration of the classes has to be performed starting from the 6 different non-absorbing markings. In total, the 6 enumerations result in 100 classes and takes approximately a minute. About 30 classes are regenerative and the global kernel is built based on these classes. Once the global kernel is defined, MRP or SMP transient analysis techniques can be used to compute the desired probability. As in this paper our focus is not the transient solution of MRPs, we have applied a simple forward discretization scheme for the solution and then validated the results by simulation. The time needed for the transient analysis is negligible compared to the enumeration phase. The results of the computation are depicted in Fig. 4.

Second, we consider the probability that at time t the system is in such a marking that both resources are failed. As we are interested in a particular situation only, once again we can perform the computations based on a selective analysis of the kernels as described in Sect. IV-B. The enumeration has to be performed for the 6 different regenerative markings which results in a total of about 150 classes. The number is higher than before because the model is not stopped anymore when double failure is reached. About 30 of the classes are regenerative and these are needed to construct the global kernel. For what concerns the local kernel, we need only those entries that corresponds to double failure. There are only 3 such non regenerative classes in which both failed places are marked. The resulting kernels are 7 by 7 as before because we need the 6 markings that can be reached through regeneration and a 7th state where we keep track of being in double failed state. This 7th state is reached only by the local kernel. The enumeration takes about one and a half minute while the rest of the computation time is negligible. The results of the computation, which were validated by simulation, are depicted in Fig. 5. In Fig. 6 and 7 we give an example for a global and a local kernel entry, respectively.

VI. Conclusions

In this paper a technique for the transient analysis of generalised semi-Markov processes has been proposed. This technique, which is an extension of the method of stochastic state classes, is based on keeping track of the elapsed time in the state density function of the classes. The approach is applicable, in principle, to any generalised semi-Markov process and it results in the possibility of symbolic derivation of transient measures in the case when all timings are expolynomial. It has been shown that, if the model visits regeneration points, the approach can be used to characterise the regeneration period even if the involved stochastic process is with multiple concurrent non-exponential activities. Based on a preliminary implementation in Oris, the method has been tested and resulted to be applicable to models whose transient analysis has not been carried out before. Future work includes the introduction of approximations in order to limit the complexity of the calculations and to make the graph of transient classes finite even if the process does not visit regeneration points.

References

[1] P. W. Glynn, "A GSMP formalism for discrete-event systems," *Proceedings of the IEEE*, vol. 77, pp. 14–23, 1989.

[2] G. Ciardo, R. German, and C. Lindemann, "A characterization of the stochastic process underlying a Stochastic Petri Net," *IEEE Trans. Softw. Eng.*, vol. 20, no. 7, pp. 506–515, 1994.

[3] P.J.Haas and G.S.Shedler, "Stochastic petri net representation of discrete event simulations," *IEEE Transactions on SW Engineering*, vol. 15, no. 4, pp. 381–393, 1989.

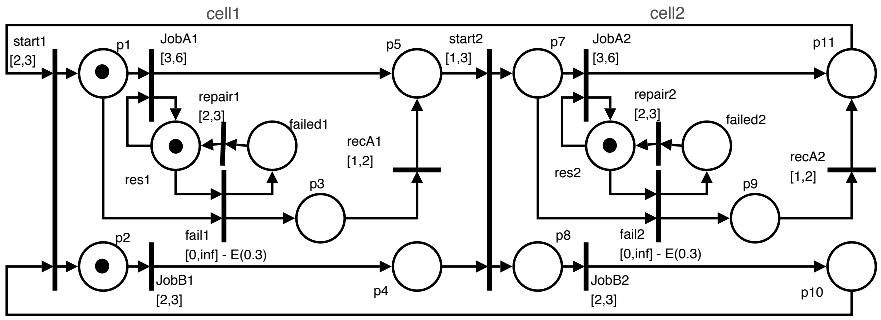

Figure 3. Petri Net modeling a system with two production cells. E(0.3) denotes an exponential transition with rate 0.3 and all other transitions are uniform on the indicated interval.

Figure 4. Probability of having reached double failure at least once as function of time for two different time scales

Figure 5. Probability of being in double failure as function of time for two different time scales

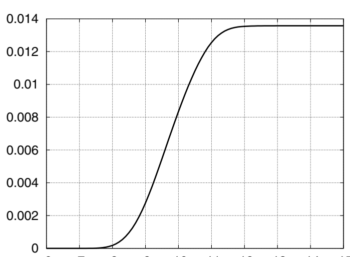

Figure 6. Probability as function of time that a regeneration period started in marking p1,p2,res1,res2 finishes in marking p1,p2,res1,res2 before time t

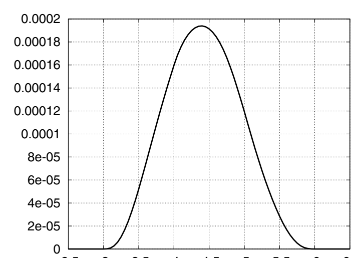

Figure 7. Probability as function of time that during a regeneration period started in marking p1,p2,res1,res2 the system is doubly failed at time t

[4] H. Choi, V. G. Kulkarni, and K. Trivedi, "Markov Regenerative Stochastic Petri Nets," *Perf. Eval.*, vol. 20, pp. 337–357, 1994.

[5] A. Bobbio and M. Telek, "Markov Regenerative SPN with non-overlapping activity cycles," *Int. Computer Performance and Dependability Symp. - IPDS95*, pp. 124–133, 1995.

[6] R. German and C. Lindemann, "Analysis of Stochastic Petri Nets by the method of supplementary variables," in *Performance Evaluation*, vol. 20, 1994, pp. 317–335.

[7] E.Cinlar, *Introduction to Stochastic Processes*. Prentice Hall, Englewood Cliffs, 1975.

[8] R. German, D. Logothetis, and K. S. Trivedi, "Transient analysis of markov regenerative stochastic petri nets: a comparison of approaches," in *PNPM '95: Proceedings of the Sixth International Workshop on Petri Nets and Performance Models*. Washington, DC, USA: IEEE Computer Society, 1995, p. 103.

[9] D. Cox, "The analysis of non-markovian stochastic processes by the inclusion of supplementary variables," *Proceedings of the Cambridge Phylosophical Society*, vol. 51, pp. 433–440, 1955.

[10] M. Telek and A. Horváth, "Supplementary variable approach applied to the transient analysis of age-MRSPNs," in *Proc. Int. Performance and Dependability Symp.*, 1998, pp. 44–51.

[11] R. German and M. Telek, "Formal relation of Markov renewal theory and supplementary variables in the analysis of Stochastic Petri Nets," in *Proc.* 8^{th} *International Workshop on Petri Nets and Performance Models*, 1999, pp. 64–73.

[12] C. Lindemann and G.S.Schedler, "Numerical analysis of Deterministic and Stochastic Petri Nets with concurrent deterministic transitions," *Performance Evaluation*, vol. 27-28, pp. 565–582, 1996.

[13] C. Lindemann and A. Thuemmler, "Transient analysis of Deterministic and Stochastic Petri Nets with concurrent deterministic transitions," *Performance Evaluation*, vol. 36-37, pp. 35–54, 1999.

[14] G. Bucci, R.Piovosi, L. Sassoli, and E. Vicario, "Introducing probability within state class analysis of dense time dependent systems," *Proc. of the* 2^{nd} *Int. Conf. on the Quant. Evaluation of Sys.(QEST)*, September 2005.

[15] L. Carnevali, L. Sassoli, and E. Vicario, "Using stochastic state classes in quantitative evaluation of dense-time reactive systems," *IEEE Trans. on Soft. Eng.*, Nov. 2009.

[16] A. Horváth and E. Vicario, "Aggregated stochastic state classes in quantitative evaluation of non-markovian stochastic Petri nets," in *Proc. of 6th International Conference on the Quantitative Evaluation of Systems (QEST)*, Budapest, Hungary, Sept 2009.

[17] E. Vicario, "Static analysis and dynamic steering of time dependent systems using Time Petri Nets," *IEEE Trans. on Soft. Eng.*, August 2001.

[18] B. Berthomieu and M. Diaz, "Modeling and verification of time dependent systems using Time Petri Nets," *IEEE Trans. on Soft. Eng.*, vol. 17, no. 3, March 1991.

[19] D. Dill, "Timing assumptions and verification of finite-state concurrent systems," *Proc.Workshop on Computer Aided Verification Methods for Finite State Systems*, 1989.

[20] L. Carnevali, L. Ridi, and E. Vicario, "Partial stochastic characterization of timed runs over dbm domains," in *Proceedings of the 9th International Workshop on Performability Modeling of Computer and Communication Systems (PMCCS 2009)*, Eger, Hungary, September 2009.

[21] P.Merlin and D.J.Farber, "Recoverability of communication protocols," *IEEE Trans.on Communications*, vol. 24, no. 9, 1976.

[22] J.Bengtsson and W.Yi, "Timed automata: Semantics, algorithms and tools," in *Concurrency and Petri Nets*. LNCS 3098, 2004.

[23] L. Carnevali, L. Grassi, and E. Vicario, "State-density functions over DBM domains in the analysis of non-Markovian models," *IEEE Trans. on Soft. Eng.*, vol. 35, pp. 178 – 194, March-April 2009.

[24] V. G. Kulkarni, *Modeling and analysis of stochastic systems*. Chapman and Hall, 1995.

[25] D. Logothetis, K. Trivedi, and A. Puliafito, "Markov regenerative models," in *Proceedings of the International Computer Performance and Dependability Symposium on Computer Performance and Dependability Symposium (IPDS'95)*. Washington, DC, USA: IEEE Computer Society, 1995.