A characterization of the stochastic process underlying a stochastic Petri net

Gianfranco Ciardo[*] Reinhard German[†] Christoph Lindemann[‡]

Department of Computer Science
College of William and Mary
Williamsburg, VA 23187, USA
ciardo@cs.wm.edu

Institut für Technische Informatik
Technische Universität Berlin
10587 Berlin, Germany
rge@cs.tu-berlin.de lind@cs.tu-berlin.de

Abstract

Stochastic Petri nets (SPNs) with generally distributed firing times are isomorphic to generalized semi-Markov processes (GSMPs), but simulation is the only feasible approach for their solution. We explore a hierarchy of SPN classes where modeling power is reduced in exchange for an increasingly efficient solution. Generalized stochastic Petri nets (GSPNs), deterministic and stochastic Petri nets (DSPNs), semi-Markovian stochastic Petri nets (SM-SPNs), timed Petri nets (TPNs), and generalized timed Petri nets (GTPNs) are particular entries in our hierarchy. Additional classes of SPNs for which we show how to compute an analytical solution are obtained by the method of the embedded Markov chain (DSPNs are just one example in this class) and state discretization, which we apply not only to the continuous-time case (PH-type distributions), but also to the discrete case.

1 Introduction

About one decade ago, Molloy [16], Natkin [18], and Symons [19] independently proposed associating exponentially distributed firing delays to transitions of a Petri net. Generalized stochastic Petri nets (GSPNs), introduced by Ajmone Marsan, Balbo, and Conte in [2], relax this condition by also allowing transitions firings in constant zero time. Stochastic activity networks (SANs) [15] and stochastic reward nets (SRNs) [9] are two other classes of Petri nets in which transition firing is either exponentially distributed or constant zero. GSPNs, SANs, and SRNs are isomorphic to continuous-time Markov chains (CTMCs).

The need for non-exponentially distributed transition firing times in SPNs has been observed by several authors. Bechta, Geist, Nicola, and Trivedi defined extended stochastic Petri nets (ESPNs) [4] in which the firing delay of timed transitions may have arbitrary distribution. The numerical solution for ESPNs is applicable when the underlying stochastic behavior is a semi-Markov process. Ciardo proposed several extensions to ESPNs and called this modeling formalism semi-Markov SPNs (SM-SPNs) [8]. Deterministic and stochastic Petri nets (DSPNs) introduced by Ajmone Marsan and Chiola [3] as an extension to GSPNs include exponentially distributed and constant timing. If at most one deterministic transition is enabled in a marking, the steady state solution can be computed using an embedded Markov chain. Timed Petri nets (see [20] as a recent survey paper) and generalized timed Petri nets [13] employ a discrete time scale for their underlying Markov process. Timed transition in TPNs and GTPNs fire in three phases and the next transition to fire is preselected according to a probability distribution.

Recently, the class of extended DSPNs has been introduced [11]. In extended DSPNs, transitions with arbitrary distributed firing times are allowed under the restriction that at most one transition with non-exponentially distributed firing time is enabled in each marking. General formulas for the steady-state solution of extended DSPNs were derived by the method of supplementary variables. In case the non-exponential distributions are piecewise specified by polynomials,

[*] G. Ciardo was a Visiting Professor at the Technische Universität Berlin with support from the German Academic Exchange Office (DAAD).

[†] R. German was supported by the Siemens Corporate Research & Development.

[‡] C. Lindemann was supported by the Federal Ministry for Research and Development of Germany (BMFT) under grant ITR 9003.

an efficient numerical solution is possible. Furthermore, Choi, Kulkarni, and Trivedi introduced the class of Markov regenerative SPNs (MR-SPN) in [6, 7] which is equivalent to the class of extended DSPNs. The authors observed that the stochastic process underlying a MR-SPN is a Markov regenerative process and derived general formulas for the transient and steady-state solution. The transient solution method employs inversion of matrices containing expressions of Laplace-Stieltjes transforms and the inversion of Laplace transforms.

In this paper, we explore various subclasses of SPNs, obtained by imposing restrictions on the combinations of firing distributions types allowed or on the effect of a transition firing on the other enabled transitions. This leads to a hierarchy of SPN classes where modeling power is reduced in exchange for an increasingly efficient solution. GSPNs, DSPNs, SM-SPNs, TPNs, and GTPNs are particular entries in our hierarchy. Furthermore, we show that state discretization can be applied to both the continuous-time and the discrete-time case. The class of Discrete-time SPNs introduced by Molloy [17] is then extended by allowing arbitrary discrete firing time distributions rather than only the geometric distribution.

We introduce the semi-regenerative SPNs (SR-SPNs) for which we show how to compute the steady state solution by embedding a Markov chain at appropriately defined regeneration points. The evolution of a SR-SPN between regeneration points is not restricted to be a CTMC, as for MR-SPNs and extended DSPNs. Thus, we relax the restriction that in any marking at most one timed transition with non-exponentially distributed firing delay is enabled. In Section 4, we present a SR-SPNs of a transmission line, where two deterministic transitions are concurrently enabled. We consider in particular SR-SPNs where all transition firing distributions can be piecewise defined by polynomials multiplied with exponential expressions. This class of probability distribution is referred to as *expolynomial distributions* and includes the exponential as well as the constant distribution as special cases.

The paper is organized as follows. Section 2 defines SPNs and describes their behavior. A hierarchical classification of SPNs according to the underlying stochastic process is presented in Section 3 and the feasibility of their numerical solution is discussed. To illustrate the numerical solution method of SR-SPNs, a SR-SPN of a simple transmission line is analyzed in Section 4. Finally, concluding remarks are given.

2 Stochastic Petri nets

Throughout this paper, we adopt the common formalism introduced for Petri nets in which transition firings is augmented with time [9]. We consider *stochastic Petri nets* (SPN) in which firings of timed transitions is an atomic operation and two types of transitions exist: *immediate transitions*, which fire without delay, and *timed transitions*, which fire after a random firing delay. The firing of immediate transitions has priority over the firing of timed transitions. Each immediate transition has associated a weight which determines its firing probability in case this transition is conflicting with some other immediate transition. The firing delay of each timed transition is specified by a probability distribution function. As a consequence, the reachability set of a SPN can be divided into *vanishing* and *tangible* markings depending on whether an immediate transition is enabled. The tangible markings of a SPN correspond to the states of an underlying stochastic process, the *marking process*. Firing weights of immediate transitions, average firing delays of timed transitions, and arc multiplicities may be marking-dependent.

A timed transition is denoted by t, the set of all timed transitions by T. A tangible marking is denoted by μ, the tangible reachability set by $\mathcal{S}$. $\mathbf{E}(\mu)$ is the set of transitions enabled in marking μ and $\mathcal{S}_t = \{\mu \in \mathcal{S} \mid t \in \mathbf{E}(\mu)\}$ is the set of markings where $t \in T$ is enabled. $F_t(\mu, \cdot)$ is the probability distribution function for the firing time of t in μ. If this distribution is not marking dependent, we write $F_t(\cdot)$.

We consider only the case where the firing time distribution may depend on the marking through a "scaling factor" [1]. Define the average firing time of transition t in marking μ as:

$$\forall \mu \in \mathcal{S}_t : \ f_t(\mu) = \int_0^\infty (1 - F_t(\mu, \theta))\, d\theta$$

Then, we require that

$$\forall \mu \in \mathcal{S}_t : \ F_t(\mu, \theta) = \hat{F}_t\left(\frac{\theta}{f_t(\mu)}\right)$$

where $\hat{F}_t(\cdot)$ is the "normalized firing time distribution" of t, which has average one and is not a marking-dependent quantity.

To specify the influence of the firing of a transition on the firing process of other transitions enabled in the current marking, *execution policies* have been introduced in [1]. We allow different execution policies for timed transitions in a SPN which may also depend on the marking [10]. Define $e_{t,s}(\mu) \in \{R, C\}$

to be the execution policy to be used for transition s when transition t fires in marking μ. If $e_{t,s}(\mu) = R$, transition s "Restarts" (samples a new random delay from the associated distribution); if $e_{t,s}(\mu) = C$, it "Continues".

2.1 Stochastic behavior

The tangible marking μ of the SPN as a function of the time θ is described by a continuous-time stochastic process, the *marking process*: $\{\mu\{\theta\}, \theta \geq 0\}$ or by a discrete-time bivariate stochastic process [9]: $\{(\theta^{[n]}, \mu^{[n]}) : n \in \mathbb{N}\}$, where $\theta^{[n]}$ is the instant of time when a timed transition fires and $\mu^{[n]}$ is marking reached after this firing. $\theta^{[0]}$ is zero and $\mu^{[0]}$ is the initial marking.

Consider now the *remaining firing time* (RFT) of each timed transition after a change of the marking. The RFT of transition t enabled in marking $\mu^{[n]}$, $\tau_t^{[n]}$, specifies the time to be spent in markings enabling t before transition t can fire. The transition t with the minimum RFT enabled in $\mu^{[n]}$ fires at time $\theta^{[n+1]} = \theta^{[n]} + \tau_t^{[n]}$. If the firing time distributions of two or more transitions enabled in a marking have jumps at the same instants of time, the probability of them having the same RFT is positive. We do not consider this case, although weights can be used to define a probability mass function over these transitions.

At time $\theta^{[0]}$, the RFT of each timed transition enabled in the initial marking is given by a random sample, $\text{rand}(F_t(\mu^{[0]}, \cdot))$, from the firing time distribution associated with this transition (all other RFTs are undefined). If transition $t \in \mathbf{E}(\mu^{[n]})$ has the minimum RFT, at time $\theta^{[n]}$, the RFT $\tau_s^{[n+1]}$ of any other transition $s \in \mathbf{E}(\mu^{[n+1]})$ at time $\theta^{[n+1]} = \theta^{[n]} + \tau_t^{[n]}$ is:

$$\begin{array}{ll} \text{rand}(F_s(\mu^{[n+1]}, \cdot)) & \text{if } e_{t,s}(\mu^{[n]}) = R \vee s \notin \mathbf{E}(\mu^{[n]}) \\ (\tau_s^{[n]} - \tau_t^{[n]}) \cdot \frac{f_s(\mu^{[n+1]})}{f_s(\mu^{[n]})} & \text{if } e_{t,s}(\mu^{[n]}) = C \wedge s \in \mathbf{E}(\mu^{[n]}) \end{array}$$

Using the terminology of [1], this behavior corresponds to a "race policy", since the minimum RFT determines the next transition to fire. After the firing of a timed transition, the next tangible marking is reached either directly or after the firing of immediate transitions. The probability of branching to marking ν after the firing of timed transition t in marking μ is $\delta_{\mu,\nu}^t$.

The following firing time distributions are important in practical applications:

- *constant*: $X \sim \text{Const}(c), c \geq 0 \Leftrightarrow \Pr\{X \leq \theta\} = 0$ if $\theta < c$, 1 if $\theta \geq c$.
- *geometric*: $X \sim \text{Geom}(p, \omega), 0 < p \leq 1, \omega \geq 0 \Leftrightarrow \Pr\{X \leq \theta\} = 1 - (1-p)^{\lfloor \frac{\theta}{\omega} \rfloor}$, where ω is the length of the unit step. The constant distribution is a special case: $\text{Const}(c)$ is equivalent to $\text{Geom}(1, c)$.
- *discrete*: $X \sim \text{Discr} \Leftrightarrow$ the distribution function of X is obtained as a weighted sum of a (finite or countably infinite) number of constant distributions. The geometric distribution is a special case. It is possible to approximate any distribution arbitrarily well by using a sufficiently large number of elements in the weighted sum.
- *exponential*: $X \sim \text{Expo}(\lambda), \lambda > 0 \Leftrightarrow \Pr\{X \leq \theta\} = 1 - e^{-\lambda\theta}$. This distribution approaches $\text{Const}(0)$ as λ increases.
- *uniform*: $X \sim \text{Unif}(a, b), b > a \geq 0 \Leftrightarrow \Pr\{X \leq \theta\} = 0$ if $\theta < a$, $(\theta - a)/(b - a)$ if $a \geq \theta \geq b$, and 1 if $\theta \geq b$. This distribution approaches $\text{Const}(b)$ as a approaches b.
- *polynomial*: $X \sim \text{Poly} \Leftrightarrow$ the distribution function of X is piecewise defined by polynomials in θ (expressions of the form $\sum_{i=0}^{n} a_i \theta^i, a_i \in \mathbb{R}$) and has finite support $[\theta_{min}, \theta_{max}]$. The finite discrete and uniform distributions are special cases. It is possible to approximate any distribution arbitrarily well by using either a sufficiently large number of polynomials of small degree (e.g., constants, as for the discrete distributions) or by using a polynomial of sufficiently large degree.
- *expolynomial*: $X \sim \text{Expoly} \Leftrightarrow$ the distribution function of X is piecewise defined by expolynomials in θ (expressions of the form $\sum_{i=0}^{n} \sum_{j=0}^{m} a_{ij} \theta^i e^{-\lambda_{ij}\theta}, a_{ij} \in \mathbb{R}, \lambda_{ij} \in [0, +\infty)$). The polynomial and exponential distributions are special cases.

3 SPNs with efficient solution

In this section, we describe several types of behavior which might render the solution analytically tractable. This leads to a hierarchy of SPN classes where modeling power is reduced in exchange for an increasingly efficient solution. The classes are defined by the underlying stochastic process.

3.1 Markov SPNs

The main obstacle to an analytical solution is the presence of the RFT in the state description. If the firing time distribution of t is memoryless, the RFT of t in μ has the same distribution as the entire firing time, $F_t(\mu, \cdot)$, hence, there is no need to include it in the state description. Accordingly, two classes of SPNs were defined: we call them "CTMC-SPNs", where all distributions are exponential [16, 18], and "DTMC-SPNs", where all distributions are geometric [17], since the marking process $\{\mu(\theta) : \theta \geq 0\}$ is a continuous-time Markov chain (CTMC) or a discrete-time Markov chain (DTMC), respectively.

As the geometric distribution is memoryless only at discrete time instants multiple of the "time step" ω, exponential and geometric distributions cannot be freely mixed. A special case of memoryless distribution is the constant zero, the distribution of the immediate transitions.

GSPNs [2] are a special case of CTMC-SPNs where the mass at zero is either zero or one. By using state-expansion, any phase-type distribution with any mass at zero still results in an underlying CTMC [1].

The "discrete-time SPNs" defined in [17] allow only geometric distributions with the same step ω, possibly with parameter one, that is, the constant ω, since $\mathrm{Const}(\omega)$ is equivalent to $\mathrm{Geom}(1, \omega)$. A discretization analogous to the one used to expand a phase-type distribution can be applied to the discrete-time case [8]. First, a geometric distribution with unit step $i\omega$ can be discretized as shown in Figure 1, where t_1 has step 3ω and t_2 has step ω. Weights are needed to decide whether transition t_1 or t_2 will fire, given that both attempt to fire, an event which has probability pq when the underlying DTMC is in state $(100c)$. This allows more generality than in [17], since the timing of a transition (i.e., $F_{t_1} \sim \mathrm{Geom}(p, 3\omega)$) and its ability to fire when competing with other transitions (i.e., $w_{t_1} = w_1$) are described by different quantities. Then, $\mathrm{Const}(i\omega)$ is equivalent to $\mathrm{Geom}(1, i\omega)$, hence TPNs are also reducible to a DTMC with unit step ω, if all constant firing times are a multiple of ω. The process described by a TPN is probabilistic even if its firing times are not random variables since, whenever two transitions have the minimum RFT, the conflict must be resolved probabilistically using the weight information. Finally, since any discrete distribution can be obtained as a weighted combination of constants, any SPN whose firing distributions have as support a subset of $\{i\omega : i \in \mathbb{N}\}$ can be reduced to a DTMC with unit step ω.

For both CTMC-SPNs and DTMC-SPNs, steady-state and transient analysis can be performed using standard numerical techniques [9]. Assume that the state space of the process is $\mathcal{S}$ and the initial proba-

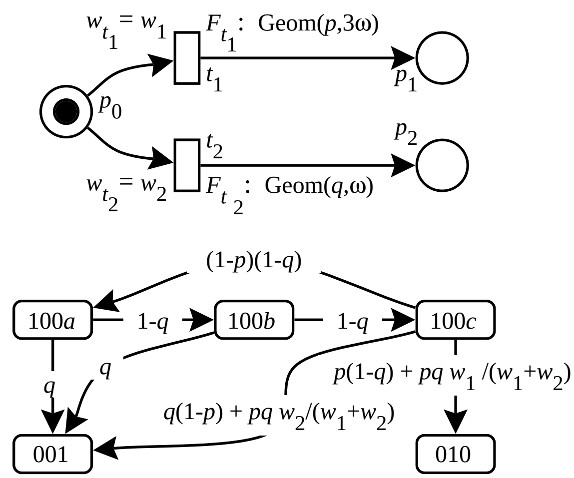

Figure 1: Discretizing two Geom distributions.

bility distribution is $\pi(0)$. The steady-state probability vector π of a CTMC described by the infinitesimal generator $\mathbf{Q}$, or of a DTMC described by the one-step transition probability matrix $\mathbf{P}$, is the solution of

$$\pi\mathbf{Q} = 0 \quad \text{or} \quad \pi(\mathbf{P} - \mathbf{I}) = 0 \quad \text{subject to} \quad \sum_{i \in \mathcal{S}} \pi_i = 1$$

assuming that $\mathcal{S}$ contains only one recurrent class of states (if this is not the case, $\mathcal{S}$ can be partitioned into a transient class and two or more recurrent classes, which can be solved independently). Sparsity-preserving iterative methods such as Gauss-Seidel or Successive Over-Relaxation can be effectively used for the solution. For transient analysis of the continuous case, the transient probability vector at time θ is the solution of

$$\frac{d\pi(\theta)}{d\theta} = \pi(\theta)\mathbf{Q} \quad \text{with initial condition} \quad \pi(0)$$

and can be computed using Jensen's method, also called Uniformization or Randomization [12]. For the discrete case, the Power method can be used:

$$\pi((i+1)\omega) = \pi(i\omega)\mathbf{P} \quad \text{starting from} \quad \pi(0)$$

(the iterations halt when $i\omega \geq \theta$). Ergodicity is not required for transient analysis.

3.2 Semi-Markov SPNs

If the firing of a transition t in marking $\mu^{[n]}$ causes transition s to restart its firing in $\mu^{[n+1]}$, $e_{t,s}(\mu^{[n]}) =$

R, the RFT of s has to be resampled in $\mu^{[n+1]}$. If all transition pairs behave this way, the marking process is a *semi-Markov process* (SMP), that is, it enjoys absence of memory immediately after every state change [1, 4]. If s has an exponential distribution, the choice between Restart and Continue is irrelevant and we assume $e_{t,s} = R$ in this case. The time instants $\theta^{[n]}, n \in \mathbb{N}$ are called *regeneration points* [5].

For transient and steady-state analysis of a SMP, the evolution of the process during the regeneration points has to be studied. Equation (1) describes the kernel $\mathbf{K}(\theta) = [k_{ij}(\theta)]$ of a SMP. Since the future of the marking process after a regeneration point becomes a probabilistic replica of the future of the process after time zero, if started in the same state, the kernel is also given by Equation (2).

$$k_{ij}(\theta) = \Pr\{\mu^{[n+1]} = j, \theta^{[n+1]} - \theta^{[n]} \le \theta \mid \mu^{[n]} = i\} \quad (1)$$
$$= \Pr\{\mu^{[1]} = j, \theta^{[1]} \le \theta \mid \mu^{[0]} = i\} \quad (2)$$

Equation (3) describes the vector $\mathbf{h}(\theta) = [h_i(\theta)]$ of holding time distributions in the states of the SMP, which can be reduced to Equation (4).

$$h_i(\theta) = \Pr\{\theta^{[n+1]} - \theta^{[n]} \le \theta \mid \mu^{[n]} = i\} \quad (3)$$
$$= \Pr\{\theta^{[1]} \le \theta \mid \mu^{[0]} = i\} \quad (4)$$

The matrix $\mathbf{\Pi}(\theta)$ of transient solutions of a SMP is given by the following system of integral equations:

$$\mathbf{\Pi}(\theta) = (\mathbf{I} - \mathrm{diag}(\mathbf{h}(\theta))) + \int_0^\theta \mathbf{\Pi}(\theta - y) d\mathbf{K}(y) \quad (5)$$

where $\mathrm{diag}(\mathbf{h}(\theta))$ represents a square matrix having the elements of $\mathbf{h}(\theta)$ on the main diagonal and zeros elsewhere.

For steady-state analysis, an *embedded Markov chain* (EMC) can be defined. The one-step transition probability matrix $\mathbf{P}$ of the EMC is computed by studying the evolution of the SMP between the regeneration points. Having obtained the matrix $\mathbf{P}$, the steady-state solution of the EMC can be derived by solving the linear system of global balance equations. Subsequently, the vector $\mathbf{c}$ of conversion factors has to be computed. The entries of $\mathbf{c}$ represent the expected holding times in the states of the SMP between two regeneration points. The solution vector of the EMC is multiplied with the vector $\mathbf{c}$ and normalized to obtain the steady-state probability vector of the SMP.

The one-step transition probability matrix $\mathbf{P}$ of the EMC is derived from the kernel: $\mathbf{P} = \lim_{\theta\to\infty} \mathbf{K}(\theta)$ and the vector $\mathbf{c}$ of conversion factors is given by $\mathbf{c} = \int_0^\infty (1 - \mathbf{h}(\theta))\, d\theta$. The steady-state solution γ of the EMC can be obtained by solving:

$$\gamma \cdot (\mathbf{P} - \mathbf{I}) = \mathbf{0} \quad \text{subject to} \quad \sum_i \gamma_i = 1 \quad (6)$$

and the steady-state solution of the SMP is given by:

$$\pi = \frac{\gamma \cdot \mathrm{diag}(\mathbf{c})}{\gamma \cdot \mathbf{c}} \quad (7)$$

For a SM-SPN, the entries of kernel and the vector of holding times are given by (for simplicity, we assume that simultaneous firings have zero probability):

$$k_{ij}(\theta) = \sum_{t\in\mathbf{E}(i)} \delta_{ij}^t \cdot \int_0^\theta \prod_{s\in\mathbf{E}(i), s\neq t} (1 - F_s(i,y))\, dF_t(i,y)$$

$$h_i(\theta) = 1 - \prod_{t\in\mathbf{E}(i)} (1 - F_t(i,\theta))$$

Equation (5) can be solved directly or by employing Laplace-Stieltjes transforms as recently proposed in [6, 7]. This solution method may cause numerical difficulties and its computational cost is significant for large models. However, when all firing time distributions are expolynomial distributions, the entries of $\mathbf{P}$ and $\mathbf{c}$ can be obtained by symbolic integration, and the steady-state solution can then be obtained by solving (6) and (7).

3.3 Semi regenerative SPNs

If there is a marking μ and two transitions $t, s \in \mathbf{E}(\mu)$ such that $e_{t,s}(\mu) = C$, t can fire before s, and s does not have an exponential distribution, the marking process is not semi-Markov. However, under certain conditions, it might be possible to find regeneration points, at which the process enjoys absence of memory. This process is called a *semi regenerative process* (SRP) in [5], so we call *semi regenerative SPN* (SR-SPN) a SPN whose marking process is a SRP. For the transient and steady-state analysis, the evolution of the process between the regeneration points must be studied. Since this can be an arbitrary stochastic process, the marking process of a SR-SPN is more general than in a MR-SPNs [6] or an extended DSPN [11]. The set of regeneration points of a SR-SPN can be expressed as

$$\{\theta^{[n_k]} : k \in \mathbb{N}, n_k \in \mathbb{N}, n_0 = 0, n_{k+1} > n_k\}$$

where each regeneration point $\theta^{[n_k]}$ must satisfy the condition that, when the SPN enters marking $\mu^{[n_k]}$ at time $\theta^{[n_k]}$ by firing transition t, any transition enabled in $\mu^{[n_k]}$ restarts its firing process.

$$\forall\theta^{[n_k]}, \forall s \in \mathbf{E}(\mu^{[n_k]}),\ e_{t,s}(\mu^{[n_k-1]}) = R$$

In the following we concentrate on steady-state analysis. As for SMP, an EMC is defined at regeneration points of a SRP. For the one-step transition probability matrix $\mathbf{P}$ of the EMC, the evolution of the SRP between the regeneration points must be studied. The steady-state solution of the EMC can be computed as in the case of SMPs, but, the conversion factors constitute a matrix rather than a vector. The steady-state probability vector of the SRP is derived by multiplying the steady-state probability vector of the EMC with the matrix of conversion factors and normalizing according to Equations (6) and (8).

The one-step transition probability matrix $\mathbf{P} = [p_{ij}]$ of the EMC and the matrix $\mathbf{C} = [c_{ij}]$ of conversion factors are defined by:

$$\begin{aligned} p_{ij} &= \Pr\{\mu^{[n_{k+1}]} = j \mid \mu^{[n_k]} = i\} \\ &= \Pr\{\mu^{[n_1]} = j \mid \mu^{[0]} = i\} \end{aligned}$$

$$\begin{aligned} c_{ij} &= \mathrm{E}\{\text{time in } j \text{ during } [\theta^{[n_k]}, \theta^{[n_{k+1}]}) \mid \mu^{[n_k]} = i\} \\ &= \mathrm{E}\{\text{time in } j \text{ during } [0, \theta^{[n_1]}) \mid \mu^{[0]} = i\} \end{aligned}$$

The steady-state solution of the EMC can still be obtained by solving the linear system of equations (6). The solution of the EMC is converted to that of the SRP multiplying by the conversion factors and normalizing:

$$\gamma' = \gamma \cdot \mathbf{C}, \qquad \pi = \frac{\gamma'}{\sum_i \gamma'_i} \tag{8}$$

For a structural definition of SR-SPNs, the sets $\mathcal{S}_E$, T_G, and T_R, are introduced. $\mathcal{S}_E$ is the set of all markings in which only exponential transitions are enabled. T_G is the set of all general (non-exponential) transitions of the SPN. $T_R \subseteq T_G$ is a set which contains regenerative transitions. A transition $t \in T_G$ is called regenerative, if all other transitions of the SPN have to restart when t becomes enabled, fires, or becomes disabled. Note that T_R does not have to contain all general transitions.

Definition 1 (SR-SPN) *A SPN is a SR-SPN, if a set T_R of regenerative transitions can be found, such that $\mathcal{S}_E$ and $\mathcal{S}_t, t \in T_R$ constitute a partition of $\mathcal{S}$.*

The definition of the regeneration points of a SR-SPN depends on whether a regenerative transition is enabled or not. For states $\mu^{[n_k]} \in \mathcal{S}_E$, the next regeneration point is chosen to be the instant of time after the transition with the minimum firing delay has fired: $\theta^{[n_{k+1}]} = \theta^{[n_k+1]}$. For states $\mu^{[n_k]} \in \mathcal{S}_t, t \in T_R$, the next regeneration point is chosen to be the instant of time after t has fired or has become disabled.

The possible evolution of the SR-SPN during the enabling period of a regenerative transition t is described by the *subordinated (stochastic) process* of $t \in T_R$. The matrix of transient state probabilities for this process is $\mathbf{\Pi}_t(\theta) = [\pi^t_{ij}(\theta)]$:

$$\pi^t_{ij}(\theta) = Pr\{\text{state } j \text{ at time } \theta \mid \text{state } i \text{ at time } 0\}$$

Based on $\mathbf{\Pi}(\theta)$, $\mathbf{P}$ and $\mathbf{C}$ can be defined row-wise. For all regenerative transitions $t \in T_R$ the rows corresponding to states $i \in \mathcal{S}_t$ are defined by:

$$\mathbf{P}_i = \mathbf{u}_i \cdot\ \mathbf{\Omega}_t \cdot \mathbf{\Delta}_t \tag{9}$$

$$c_{ij} = \begin{cases} \psi^t_{ij} & \text{if } j \in \mathcal{S}_t \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $\mathbf{u}_i$ is the i-th row unity vector, $\mathbf{\Delta}_t = \left[\delta^t_{ij}\right]$ is the matrix of branching probabilities after a firing of t, and $\mathbf{\Omega}_t = \left[\omega^t_{ij}\right]$ and $\mathbf{\Psi}_t = \left[\psi^t_{ij}\right]$ are the transient probabilities and the expected holding times of the states of the subordinated process:

$$\mathbf{\Omega}_t = \int_0^\infty \mathbf{\Pi}_t(\theta) dF_t(\theta) \tag{11}$$

$$\mathbf{\Psi}_t = \int_0^\infty \mathbf{\Pi}_t(\theta)\left(1 - F_t(\theta)\right) d\theta \tag{12}$$

The entries of $\mathbf{P}$ and $\mathbf{C}$ for rows corresponding to states $i \in \mathcal{S}_E$ are given by

$$p_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{\lambda_{ij}}{\lambda_i} & \text{if } i \neq j \end{cases} \qquad c_{ij} = \begin{cases} \frac{1}{\lambda_i} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{13}$$

where λ_{ij} is the rate leading from state i to state j and λ_i is the sum of all outgoing rates for state i.

If a regenerative transition t is never enabled together with other transitions, the steady-state solution of a SR-SPN is insensitive to the distribution of t, since equations (11) and (12) reduce to:

$$\mathbf{\Omega}_t = \mathbf{I}, \quad \mathbf{\Psi}_t = \mathrm{diag}(f_t(i))$$

The efficient numerical computation of the Equations (11) and (12) is the critical step for the practical application of SR-SPNs. In the next section, an example is presented, in which the subordinated process is a SMP.

The numerical solution of a SR-SPNs with large state space can be performed efficiently if the subordinated processes are CTMCs. In this case, at most one regenerative transition may be enabled in each marking. SR-SPNs with this restriction are equivalent to the class of extended DSPN defined in [11] and Markov regenerative SPNs defined in [6]. In case of a subordinated CTMC, the matrix of transient state probabilities for the subordinated process is given by the matrix exponential of the generator matrix $\mathbf{Q}_t$ for the subordinated CTMC of transition t:

$$\forall t \in T_R: \ \mathbf{\Pi}_t(\theta) = e^{\mathbf{Q}_t \theta} \tag{14}$$

In Appendix A we show how to generalize Jensen's method for an efficient calculation of the rows of the equations (11) and (12) in case of expolynomial regenerative transitions.

The analysis can be generalized to the case where the firing time distribution of a regenerative transition depends on the marking through a scaling factor. This can be done by scaling the generator matrix by the corresponding scaling factors:

$$\hat{\mathbf{Q}}_t = \left[\hat{q}_{ij}^t\right] = \left[f_t(i) \cdot q_{ij}^t\right] \tag{15}$$

In this case the matrices $\mathbf{\Omega}_t$ and $\mathbf{\Psi}_t$ are given by:

$$\mathbf{\Omega}_t = \int_0^\infty e^{\hat{\mathbf{Q}}_t \theta} d\hat{F}_t(\theta) \tag{16}$$

$$\mathbf{\Psi}_t = \int_0^\infty e^{\hat{\mathbf{Q}}_t \theta}(1 - \hat{F}_t(\theta)) d\theta \cdot \mathrm{diag}(f_t(i)) \tag{17}$$

Equations (16) and (17) generalize the results presented in [14] (see Appendix B).

3.4 Generalized semi-Markov SPNs

If there is a marking i and two transitions $t, s \in \mathbf{E}(i)$ such that $e_{t,s} = e_{s,t} = C$, and t and s do not have an exponentially distributed firing delay, the underlying process is too difficult to study as a SRP, or it might even not be a SRP. The underlying stochastic process is a generalized semi-Markov process. It is usual to resort to simulation to obtain quantitative results for such processes. However, it is also possible to derive the state equations for the transient or steady-state case by means of supplementary variables [11]. The resulting equations constitute a system of partial differential equations which can be analyzed numerically by replacing the differential quotients by finite difference quotients.

Figure 2 summarizes the classes of SPNs and relates them to the underlying stochastic processes.

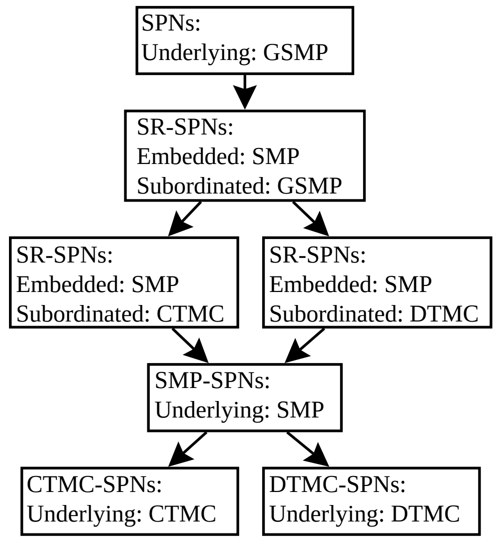

Figure 2: SPN hierarchy.

4 An example

In this section, a SPN model of a simple transmission line is considered, to illustrate the steady-state analysis of a SR-SPN. It is assumed that, after the generation of a message, its transmission begins and a timeout clock is started. Conflict for the medium can delay the start of transmission. If the timeout elapses before the transmission of the message is completed, the transmission will be repeated. A measure of interest is the number of successful transmissions per unit time. Figure 3 shows a SPN model of the system. The generation of a message is modeled by transition t_0, which has an arbitrary firing time distribution with average firing time $f_{t_0} = \tau_0$. The timeout and the transmission of the message are represented by transitions t_1 and t_3 with constant firing times of τ_1 and τ_3, respectively. The delay to acquire the medium is modeled by transition t_2, which has an exponentially distributed firing time with rate λ. The multiplicity of some input arcs is marking-dependent to ensure that places p_1, p_2, and p_3 are empty after the firing of t_1 or t_3.

In this SPN, t_1 and t_3 have constant firing times, are concurrently enabled, and start their firing process at different instants of time. Nevertheless, the analysis is possible using firings of the transitions t_0 and t_1 as regeneration points: $T_R = \{t_0, t_1\}$. The reachability set of the SR-SPN consists of three markings and is

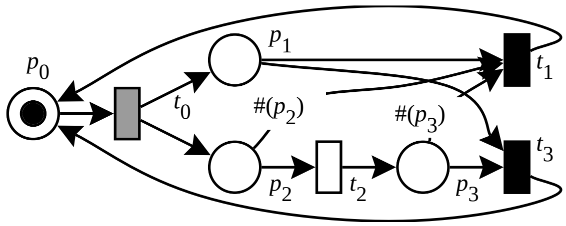

Figure 3: SR-SPN of a transmission line.

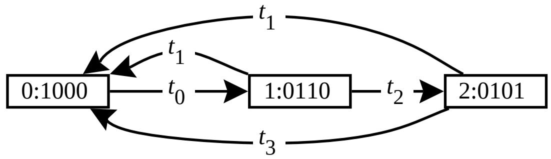

Figure 4: The reachability graph for the model.

shown in Figure 4:

$$\mathcal{S} = \{0, 1, 2\},\ \ 0 : [1, 0, 0, 0],\ \ 1 : [0, 1, 1, 0],\ \ 2 : [0, 1, 0, 1]$$

The sets $\mathcal{S}_{t_0} = \{0\}$ and $\mathcal{S}_{t_1} = \{1, 2\}$ constitute a partition of the reachability set $\mathcal{S}$. Since transition t_0 is exclusively enabled, the solution of the SPN is insensitive to the distribution of t_0. It only depends on the average firing time τ_0. The process subordinated to transition t_1 is a SMP and shown in Figure 5.

The matrix $\mathbf{\Pi}_{t_1}(\theta)$ of the transient state probabilities for the SMP subordinated to t_1 can be obtained symbolically:

$$\left[\begin{array}{ccc} 1 & 0 & 0 \\ 1 - e^{-\lambda(\theta - \min(\theta, \tau_3))} & e^{-\lambda\theta} & e^{-\lambda(\theta - \min(\theta, \tau_3))} - e^{-\lambda\theta} \\ u(\theta - \tau_3) & 0 & 1 - u(\theta - \tau_3) \end{array}\right]$$

where $u(\cdot)$ is the unit step function. The matrices $\mathbf{\Omega}_{t_1}$ of state probabilities of the SMP in the instant of firing of t_1 and $\mathbf{\Psi}_{t_1}$ of expected holding times in states up

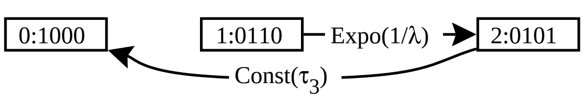

Figure 5: The SMP subordinated to t_1.

to firing of t_1 are:

$$\mathbf{\Omega}_{t_1} = \int_0^\infty \mathbf{\Pi}_t(\theta) du(\theta - \tau_1) = \mathbf{\Pi}_{t_1}(\tau_1)$$

$$\mathbf{\Psi}_{t_1} = \int_0^\infty \mathbf{\Pi}_t(\theta)(1 - u(\theta - \tau_1))d\theta = \int_0^{\tau_1} \mathbf{\Pi}_t(\theta) d\theta$$

Employing equations (9) and (10) yields the one-step transition probability matrix $\mathbf{P}$ of the EMC and the matrix of conversion factors $\mathbf{C}$:

$$\mathbf{P} = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right], \quad \mathbf{C} = \left[\begin{array}{ccc} \tau_0 & 0 & 0 \\ 0 & \psi_{11}^{t_1} & \psi_{12}^{t_1} \\ 0 & 0 & \psi_{22}^{t_1} \end{array}\right]$$

All changes of marking caused by firings of transition t_1 lead to marking 0. Marking 2 cannot be reached in a regeneration point, hence does not appear in the EMC (this implies that the last row and column of $\mathbf{P}$ and the last row of $\mathbf{C}$ do not need to be computed in practice). The steady-state probability vector of the EMC is computed by solving the linear system of equations (6). The steady-state marking probability vector of the SR-SPN is obtained by multiplying the steady-state probability vector of the EMC with the matrix $\mathbf{C}$ of conversion factors and a subsequent normalization according to Equation (8).

Conclusion

In this paper, we have classified the SPNs into various classes, according to the nature of their underlying stochastic process. The most complex class we identified, SR-SPNs, corresponds to semi-regenerative processes and effectively extends the class of SPNs for which an analytical solution is known.

We illustrate the type of behavior that can be modeled by the SR-SPNs using a simple system transmitting messages as an example. In it, two deterministic transitions can become enabled concurrently, but not necessarily at the same time. A similar model was given as an example of a DSPN which cannot be solved with previously known methods [3].

References

[1] M. Ajmone Marsan, G. Balbo, A. Bobbio, G. Chiola, G. Conte, and A. Cumani. The effect of execution policies on the semantics and analyis of Stochastic Petri Nets. *IEEE Trans. Softw. Eng.*, 15(7):832–846, July 1989.

[2] M. Ajmone Marsan, G. Balbo, and G. Conte. A class of Generalized Stochastic Petri Nets for the performance evaluation of multiprocessor systems. *ACM Trans. Comp. Syst.*, 2(2):93–122, May 1984.

[3] M. Ajmone Marsan and G. Chiola. On Petri Nets with deterministic and exponentially distributed firing times. In G. Rozenberg, editor, *Adv. in Petri Nets 1987, Lecture Notes in Computer Science 266*, pages 132–145. Springer-Verlag, 1987.

[4] J. Bechta Dugan, K. S. Trivedi, R. M. Geist, and V. F. Nicola. Extended Stochastic Petri Nets: applications and analysis. In E. Gelenbe, editor, *Performance '84*, Paris, France, Dec. 1984. North-Holland.

[5] E. Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

[6] H. Choi, V. G. Kulkarni, and K. S. Trivedi. Markov regenerative stochastic Petri nets. In G. Iazeolla and S. S. Lavenberg, editors, *Performance '93*. North-Holland, Sept. 1993.

[7] H. Choi, V. G. Kulkarni, and K. S. Trivedi. Transient analysis of deterministic and stochastic Petri nets. In M. Ajmone Marsan, editor, *Application and Theory of Petri Nets 1993, Lecture Notes in Computer Science*, volume 691, pages 166–185. Springer-Verlag, 1993.

[8] G. Ciardo. *Analysis of large stochastic Petri net models*. PhD thesis, Duke University, Durham, North Carolina, 1989.

[9] G. Ciardo, A. Blakemore, P. F. J. Chimento, J. K. Muppala, and K. S. Trivedi. Automated generation and analysis of Markov reward models using Stochastic Reward Nets. In C. Meyer and R. J. Plemmons, editors, *Linear Algebra, Markov Chains, and Queueing Models*, volume 48 of *IMA Volumes in Mathematics and its Applications*, pages 145–191. Springer-Verlag, 1993.

[10] G. Ciardo and C. Lindemann. Analysis of deterministic and stochastic Petri nets. In *Proc. of the Fifth Int. Workshop on Petri Nets and Performance Models (PNPM93)*, Toulouse, France, Oct. 1993.

[11] R. German and C. Lindemann. Analysis of Stochastic Petri Nets by the Method of Supplementary Variables. In G. Iazeolla and S. S. Lavenberg, editors, *Performance '93*. North-Holland, Sept. 1993.

[12] D. Gross and D. Miller. The randomization technique as a modeling tool and solution procedure for transient Markov processes. *Oper. Res.*, 32(2):926–944, Mar.-Apr. 1984.

[13] M. A. Holliday and M. K. Vernon. A Generalized Timed Petri Net model for performance analysis. *IEEE Trans. Softw. Eng.*, 12:1297–1310, Dec. 1987.

[14] C. Lindemann and R. German. Modeling discrete event systems with state-dependent deterministic service times. *Discrete Event Dynamic Systems: Theory and Applications*, 3:249–270, July 1993.

[15] J. F. Meyer, A. Movaghar, and W. H. Sanders. Stochastic activity networks: structure, behavior, and application. In *Proc. of the Int. Workshop on Timed Petri Nets*, Torino, Italy, July 1985.

[16] M. K. Molloy. Performance analysis using stochastic Petri nets. *IEEE Trans. Comp.*, 31(9):913–917, Sept. 1982.

[17] M. K. Molloy. Discrete Time Stochastic Petri Nets. *IEEE Trans. Softw. Eng.*, 11(4):417–423, Apr. 1985.

[18] S. Natkin. Reseaux de Petri stochastiques. These de docteur ingeneur, CNAM-Paris, Paris, France, June 1980.

[19] F. J. W. Symons. *Modeling and Analysis of Communication Protocols Using Numerical Petri Nets*. PhD thesis, University of Essex, Great Britain, 1978.

[20] W. L. Zuberek. Timed Petri nets definitions, properties, and applications. *Microelectronics and Reliability*, 31:627–644, 1991.

A Extending Jensen's method to expolynomial distributions

In this appendix we derive an efficient numerical computation of equations (11) and (12) for a transition t with an expolynomial distributed firing time and a subordinated CTMC. The presented formulas generalize Jensen's method and the formulas presented in [11] for polynomial distributions.

For $\theta \in (0, \infty), m \in \mathbb{N}, \lambda \in [0, \infty)$, define

$$\mathbf{E}^{(\theta)} = e^{\mathbf{Q}\theta}, \quad \mathbf{L}^{(\theta,m,\lambda)} = \int_0^\theta y^m e^{-\lambda y} e^{\mathbf{Q}y} dy$$

The solution of (11) and (12) is a weighted sum of matrices $\mathbf{E}^{(\theta)}$ and $\mathbf{L}^{(\theta,m,\lambda)}$. $\mathbf{E}^{(\theta)}$ can be calculated with Jensen's method: define

$$\mathbf{A} = \frac{1}{q}\mathbf{Q} + \mathbf{I}, \quad q = 1.02 \cdot \max_i |q_{ii}|$$

The rows of $\mathbf{E}^(\theta)$ are obtained as

$$\mathbf{E}^{(\theta)}(i) \approx \sum_{k=L}^{R} \mathbf{\Phi}(k) \cdot \beta(k, q\theta)$$

where $\mathbf{\Phi}$ is calculated by iterative vector-matrix multiplications:

$$\mathbf{\Phi}(0) = \mathbf{u}_i, \quad \mathbf{\Phi}(k+1) = \mathbf{\Phi}(k) \cdot \mathbf{A}$$

$\mathbf{u}_i$ is the i-th row unity-vector and $\beta(k, q\theta)$ is the k-th Poisson probability in $q\theta$ which can be calculated iteratively by:

$$\beta(0, q\theta) = e^{-q\theta}, \quad \beta(k+1, q\theta) = \beta(k, q\theta) \cdot \frac{q\theta}{k+1}$$

The truncation points L and R of the summation can be estimated for a given error tolerance [9].

For the calculation of $\mathbf{L}^{(\theta,m,\lambda)}$, the power series of the matrix exponential can be substituted:

$$\mathbf{L}^{(\theta,m,\lambda)} = \int_0^\theta y^m \cdot e^{-\lambda y} \cdot \left(\sum_{k=0}^{\infty} \mathbf{A}^k \cdot \beta(k, qy) \right) dy$$

$$= \sum_{k=0}^{\infty} \mathbf{A}^k \frac{(k+m)!}{k!(q+\lambda)^{m+1}} \left(1 - \sum_{h=0}^{k+m} \beta(h, (q+\lambda)\theta) \right)$$

Applying right truncation leads to:

$$\mathbf{L}^{(\theta,m,\lambda)}(i) \approx \sum_{k=0}^{R} \mathbf{\Phi}(k) \cdot \gamma(k) \cdot \eta(k)$$

where $\gamma(k)$ and $\eta(k)$ are calculated iteratively:

$$\gamma(0) = \frac{m!}{(q+\lambda)^{m+1}}, \quad \gamma(k+1) = \gamma(k) \cdot \frac{k+m+1}{k+1}$$

$$\eta(0) = 1 - \sum_{h=0}^{m} \beta(h, (q+\lambda)\theta)$$

$$\eta(k+1) = \eta(k) - \beta(k+m+1, (q+\lambda)\theta)$$

A criterion to bound the truncation error can be derived. The entries of each row of the matrix exponential sum to one. The truncation error of each entry of $\mathbf{L}^{(\theta,m,\lambda)}$ is bounded by ε, if R satisfies

$$R = \min_{r \in \mathbb{N}} \left\{ s - \sum_{h=0}^{r} \gamma(h)\eta(h) \le \varepsilon \right\}$$

where s is the (exact) sum of each row of $\mathbf{L}^{(\theta,m,\lambda)}$. If $\lambda = 0$, s is given by

$$s = \int_0^\theta y^m dy = \frac{\theta^{m+1}}{m+1}$$

If $\lambda > 0$, s is given by

$$s = \int_0^\theta y^m e^{-\lambda y} dy = \frac{m!}{\lambda^{m+1}} \cdot \left(1 - \sum_{h=0}^{m} \beta(h, \lambda\theta) \right)$$

The computational effort for the calculation of one row of $\mathbf{L}^{(\theta,m,\lambda)}$ depends mainly on the number of vector-matrix multiplications. If a weighted sum of matrices $\mathbf{E}^{(\theta)}$ and $\mathbf{L}^{(\theta,m,\lambda)}$ has to be calculated, it is possible to factor out the matrix powers, thus avoiding repeated vector-matrix multiplications. The asymptotical complexity for the numerical solution in case of an expolynomial distribution is therefore of the same order as in the case of a deterministic firing time.

B Formulas for marking-dependence through a scaling factor

In this appendix we derive the equations (16) and (17) presented for regenerative SPNs with a transition t with a general distributed firing time depending on the marking through a scaling factor and with a subordinated CTMC. Define the normalized state probabilities of the subordinated CTMC as (θ_j is the absolute and $\hat{\theta}$ is the normalized elapsed firing time of t):

$$\hat{\pi}^t_{ij}(\hat{\theta}) = \pi^t_{ij}(f_t(j) \cdot \hat{\theta}) = \pi^t_{ij}(\theta_j)$$

The matrix of normalized state probabilities is given by the matrix exponential of the scaled generator matrix $\hat{\mathbf{Q}}$ (this can be shown by the underlying system of differential equations [14]):

$$\hat{\mathbf{\Pi}}_t(\hat{\theta}) = \left[\hat{\pi}^t_{ij}(\hat{\theta}) \right] = e^{\hat{\mathbf{Q}}\hat{\theta}}$$

Define ω^t_{ij} as the probability, that the subordinated CTMC is in state j after t fires, given it was in state i initially, and ψ^t_{ij} as the expected holding time in state j up to the firing of t. Integrating by substitution:

$$\omega^t_{ij} = \int_0^\infty \pi^t_{ij}(\theta_j) dF_t(j, \theta_j) = \int_0^\infty \hat{\pi}^t_{ij}(\hat{\theta}) d\hat{F}_t(\hat{\theta})$$

$$\psi^t_{ij} = \int_0^\infty \pi^t_{ij}(\theta_j)(1 - F_t(j, \theta_j)) d\theta_j$$

$$= \int_0^\infty \hat{\pi}^t_{ij}(\hat{\theta})(1 - \hat{F}_t(\hat{\theta}) f_t(j) d\hat{\theta}$$

Substitution of the entries of the matrix exponential leads to equations (16) and (17).