

Approximate transient analysis for subclasses of deterministic and stochastic Petri nets

Gianfranco Ciardo Guangzhi Li
Department of Computer Science
College of William and Mary
Williamsburg, VA 23187-8795, USA
{ciardo, gli}@wm.cs.edu

Abstract

Transient analysis of non-Markovian Stochastic Petri nets is a theoretically interesting and practically important problem. In this paper, we first present a method to compute bounds and an approximation on the average state sojourn times for a subclass of deterministic and stochastic Petri nets (DSPNs) where there is a single persistent deterministic transition that can become enabled only in a special state. Then, we extend this class by allowing the transition to become enabled in any state, as long as the time between successive enablings of the deterministic transition is independent of this state, and develop a new approximate transient analysis approach. In addition to renewal theory, we only make use of discrete and continuous Markov chain concepts. As an application, we use the model of a finite-capacity queue with a server subject to breakdowns, and assess the quality of our approximations.

Keywords: renewal processes, Markov-regenerative processes, approximate transient solution of stochastic models.

1 Introduction

Stochastic Petri nets (SPNs) are well-suited for the model-based performance and dependability evaluation of complex systems. In the past few years, many papers have been published dealing with the analysis of non-Markovian SPNs where, under certain structural restrictions, the firing times of some transitions may be generally distributed. Particular attention has been given to deterministic transition firing times, an important tool for modeling discrete-event dynamic systems. Examples of activities that might have a constant duration are transfer times of fixed-size data packets in a distributed computing system with no interference or loss, timeouts in real-time systems, and repair times of components in fault-tolerant systems.

Deterministic and stochastic Petri nets (DSPNs) have been introduced in [1] as a continuous-time modeling tool that includes both exponentially distributed and constant time transitions. In

DSPNs, transition firing is atomic and the transition with the smallest firing delay is selected to fire next. Under the structural restriction that at most one deterministic transition is enabled in any marking, an analytical method to solve DSPNs in steady state exists, based on the idea of the embedded Markov chain [6, 7, 12]. In this case, the markings of a DSPN are the same as those of the untimed Petri net, thus standard structural analysis techniques can be employed. In particular, minimal-support place and transition invariants can be calculated [14].

However, the exact transient study of DSPNs is much more difficult. In [3, 4], a method based on Laplace-Stieltjes transforms is proposed, but the numerical solution is very complex. A solution approach based on supplementary variables has also been proposed [8], and further improved through the use of automatic stepsize control [10]. In [9], the two approaches are compared in terms of memory requirements, efficiency, and accuracy.

In this paper, we first present a new method to obtain bounds and an approximation for the average total sojourn time in each state up to a given time t , for a special class of DSPNs satisfying the restriction that they have only one deterministic transition d that becomes enabled only upon entering a unique state i^* , and d is “persistent”, that is, once it becomes enabled, it can become disabled only because of its own firing. The idea is to treat this stochastic process as a renewal process. By computing the expected number of renewal periods up to time t and the average sojourn time in each state during a renewal period, we compute bounds on the average total sojourn time in each state, and an approximation for it.

Then, we extend this class of DSPNs by allowing d to become enabled in any state i , as long as the time between successive enablings of d is independent of i . Our idea is to treat the stochastic process as a subordinated Markov chain and an absorbing Markov chain alternatively, and compute the sojourn time in each state recursively.

The paper is organized as follows. Section 2 reviews renewal theory and DSPN terminology. Section 3 provides the theoretical results for our first class of DSPNs, and the corresponding computational method. Section 4 present a second, more general class of DSPNs, and a different computational approach applicable to this class. A complete example is illustrated in Section 5. Finally, concluding remarks are in Section 6.

2 Background

We briefly review the essential concepts of renewal theory and DSPNs. For further information, the reader can consult [2, 17] for the former, and [1, 3, 7] for the latter. Table 1 summarizes the symbols used in this paper.

2.1 Renewal theory

Definition 2.1 If the sequence of nonnegative random variables $\{T^{[1]}, T^{[2]}, \dots\}$ is independent and identically distributed, the counting process $\{N(t) = \max\{n \in \mathbf{N} : T^{[1]} + \dots + T^{[n]} \leq t\} : t \geq 0\}$

Symbol	Meaning
d, τ	the deterministic transition in the DSPN and its firing time
i^*	the only state where d can become enabled, for the first class of DSPNs
$\mathcal{S}, \widehat{\mathcal{S}}, \widetilde{\mathcal{S}}$	state space for the DSPN, SMC, and AMC
i, ρ_i	a generic state and its reward rate
$Y^{[n]}$	length of the n^{th} stay in the AMC
$T^{[n]}$	$\tau + Y^{[n]}$, length of the n^{th} renewal period
$S^{[n]}$	n^{th} renewal time
$F, F^{[n]}$	distribution of the generic renewal period, and its n -fold convolution
$N(t)$	number of renewals up to time t
$A(t)$	$t - S^{[N(t)]}$, age at time t
$m(t)$	$E[N(t)]$, the renewal function
\mathbf{P}	matrix to transform state probabilities from the SMC to the AMC
$T_i^{[n]}$	amount of time spent in state i during the n^{th} renewal period
$T_i^{[n]}(t)$	amount of time spent in state i during the n^{th} renewal period intersected with $[0, t]$
$X(t), \pi(t), \sigma(t)$	DSPN state, probabilities, and sojourn times at (global) time t
$\widehat{X}(t), \widehat{\pi}(t), \widehat{\sigma}(t)$	SMC state, probabilities, and sojourn times at (local) time t
$\widetilde{X}(t), \widetilde{\pi}(t), \widetilde{\sigma}(t)$	AMC state, probabilities, and sojourn times at (local) time t
$\widehat{X}^{[n]}$	initial state of the SMC for the n^{th} renewal period

Table 1: Symbols used in the paper

is said to be a renewal process. □

Thus a renewal process is a counting process such that the time $T^{[1]}$ until the first event occurs has some distribution F , the time $T^{[2]}$ between the first and second event has, independent of the time of the first event, the same distribution F , and so on. When an event occurs, we say that a renewal has taken place. Then,

$$S^{[0]} = 0, \quad \text{and} \quad \forall n \geq 1, \quad S^{[n]} = \sum_{k=1}^n T^{[k]}$$

are the renewal times.

The distribution of $N(t)$ is determined by an important relationship: the number of renewals up to time t is greater than or equal to n iff the n^{th} renewal occurs by time t :

$$N(t) \geq n \iff S^{[n]} \leq t.$$

From this relation, we obtain

$$P\{N(t) = n\} = P\{N(t) \geq n\} - P\{N(t) \geq n + 1\} = P\{S^{[n]} \leq t\} - P\{S^{[n+1]} \leq t\}.$$

Since the random variables $\{T^{[1]}, T^{[2]}, \dots\}$ are independent and have a common distribution F , $S^{[n]}$ has distribution $F^{[n]}$, the n -fold convolution of F with itself. Hence,

$$P\{N(t) = n\} = F^{[n]}(t) - F^{[n+1]}(t).$$

Definition 2.2 The mean value of $N(t)$ is called the renewal function $m(t)$:

$$m(t) = E[N(t)] = \sum_{n=1}^{\infty} P\{N(t) \geq n\} = \sum_{n=1}^{\infty} P\{S^{[n]} \leq t\} = \sum_{n=1}^{\infty} F^{[n]}(t). \quad \square$$

2.2 DSPNs

We assume that the reader is familiar with the definition of Petri nets with inhibitor arcs. Then, a DSPN is obtained by associating a firing time with exponential, zero, or positive constant distribution to each transition (which is then called an exponential, immediate, or deterministic transition, respectively), and its underlying stochastic process is $\{X(t) : t \geq 0\}$, where $X(t) \in \mathcal{S}$ is the marking at time t , and \mathcal{S} is the set of possible markings, or states, which we assume finite. For $i \in \mathcal{S}$, let $\pi_i(t)$ and $\sigma_i(t)$ be the probability that the DSPN is in state i at time t , and the expected time spent in state i up to time t , respectively:

$$\pi_i(t) = \Pr\{X(t) = i\} \quad \text{and} \quad \sigma_i(t) = \int_0^t \pi_i(u) du.$$

If at most one deterministic transition is enabled in any marking, $\{X(t) : t \geq 0\}$ is a Markov regenerative process (MRGP) and its stationary analysis can be carried on by embedding.

Let $\hat{\mathcal{S}}$ be the set of markings where a deterministic transition is enabled, $\mathcal{S}^D \subseteq \hat{\mathcal{S}}$ be the set of markings where a deterministic transition becomes enabled, and \mathcal{S}^E be the set of markings where no deterministic transition is enabled. Hence, the state space of the DSPN is $\hat{\mathcal{S}} \cup \mathcal{S}^E$. For each marking $i \in \mathcal{S}^D$ enabling a deterministic transition, a continuous-time Markov chain (CTMC) is defined, to model the evolution of the DSPN until that transition is disabled or fires. This “subordinated Markov chain” (SMC) $\{\hat{X}_i(t) : t \geq 0\}$, with state space $\hat{\mathcal{S}}_i \subseteq \hat{\mathcal{S}}$, is defined so that any state in which the deterministic transition becomes disabled is considered absorbing. The SMC is studied in the transient at time τ , the firing time of the deterministic transition (the transition itself is not explicitly modeled in the SMC). For every state $j \in \hat{\mathcal{S}}_i$, we compute $\hat{\pi}_{j|i}(\tau)$, the probability of being in j at time τ , and for every transient state $j \in \hat{\mathcal{S}}_i$ we compute $\hat{\sigma}_{j|i}(\tau)$, the amount of time spent in j up to time τ , given that $\hat{X}_i(0) = i$.

Then, a discrete time Markov chain (DTMC), the so-called “embedded Markov chain” (EMC) with state space $\tilde{\mathcal{S}} = \mathcal{S}^D \cup \mathcal{S}^E$ is defined, together with a vector of expected holding times h . For $i \in \mathcal{S}^E$, h_i is the inverse of the sum of the rates leaving i , and the one-step probabilities are defined as for the classical embedding of a CTMC into a DTMC: if i goes to j with rate $\lambda_{i,j}$ in the DSPN, i goes to j with probability $\lambda_{i,j} \cdot h_i$ in the EMC. For $i \in \mathcal{S}^D$, instead, h_i is the sum of $\hat{\sigma}_{k|i}(\tau)$

over all transient states $k \in \hat{\mathcal{S}}_i$ and the EMC contains a transition from i to j with probability $\sum_{k \in \hat{\mathcal{S}}_i} \hat{\pi}_{k|i}(\tau) \cdot \mathbf{P}_{k,j}$, where matrix $\mathbf{P} \in \mathbb{R}^{\hat{\mathcal{S}} \times \hat{\mathcal{S}}}$, whose rows are probability vectors, translates SMC states into the EMC states: if k is reached by disabling d , then j is the same as k , that is, $\mathbf{P}_{k,j} = 1$ iff $k = j$; otherwise, $\mathbf{P}_{k,j}$ is the probability that the DSPN reaches j by firing d in k .

If the EMC is ergodic, we can compute the stationary probability p_i of each state $i \in \tilde{\mathcal{S}}$. Otherwise, we can compute the expected number of visits n_i to each transient state i until absorption. The analogous quantities for the original DSPN are then obtained, respectively, as:

$$\lim_{t \rightarrow \infty} \pi_k(t) = \begin{cases} \frac{p_k \cdot h_k}{\sum_{j \in \tilde{\mathcal{S}}} p_j \cdot h_j} & \text{if } k \in \mathcal{S}^E \\ \sum_{i \in \mathcal{S}^D \wedge k \in \hat{\mathcal{S}}_i} \frac{p_i \cdot \hat{\sigma}_{k|i}(\tau)}{\sum_{j \in \tilde{\mathcal{S}}} p_j \cdot h_j} & \text{if } k \in \hat{\mathcal{S}} \end{cases} \quad (\text{ergodic case})$$

$$\lim_{t \rightarrow \infty} \sigma_k(t) = \begin{cases} n_k \cdot h_k & \text{if } k \in \mathcal{S}^E \\ \sum_{i \in \mathcal{S}^D \wedge k \in \hat{\mathcal{S}}_i} n_i \cdot \hat{\sigma}_{k|i}(\tau) & \text{if } k \in \hat{\mathcal{S}} \end{cases} \quad (\text{non-ergodic case, } k \text{ transient}).$$

However, the transient analysis of DSPNs has proved much harder. As mentioned in the introduction, the algorithms known so far require either inversion of Laplace-Stieltjes transforms, or numerical integration and solution of differential equations, resulting in substantial computation.

3 A first class of DSPNs

We first consider a restricted class of DSPNs where (1) there is exactly one deterministic transition d , with constant firing time τ , (2) d is persistent, that is, if it becomes enabled, it can become disabled only by its own firing, and (3) d can become enabled only in marking i^* . These conditions can be easily checked during the generation of the state space and, in certain cases, even *a priori*, through structural arguments at the net level.

Fig. 1 shows the behavior of the MRGP we consider. Each regeneration period can be described as follows:

1. Transition d becomes enabled in state i^* at time 0.
2. The DSPN evolves according to a SMC with infinitesimal generator $\hat{\mathbf{Q}}$ and state space $\hat{\mathcal{S}}$
3. At time τ , transition d fires, let $k \in \hat{\mathcal{S}}$ be the state immediately before the firing.
4. The firing of d in k causes the DSPN to reach state $l \in \tilde{\mathcal{S}}$ with probability $\mathbf{P}_{k,l}$.
5. The DSPN then evolves according to an ‘‘absorbing Markov chain’’ (AMC) $\{\tilde{X}(t) : t \geq 0\}$ with infinitesimal generator $\tilde{\mathbf{Q}}$ and state space $\tilde{\mathcal{S}}$ (the same state space the EMC would have if we performed the embedding required for stationary analysis) until reaching state i^* , which is

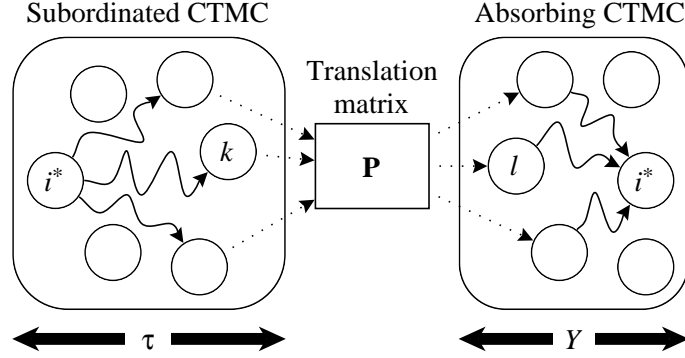


Figure 1: A depiction of our class of MRGPs.

considered absorbing for the AMC. Let Y be the time between d firing and becoming enabled again.

$T = \tau + Y$ is the length of a generic regeneration period; the state space of the DSPN is $\mathcal{S} = \hat{\mathcal{S}} \cup \tilde{\mathcal{S}}$ and $\hat{\mathcal{S}} \cap \tilde{\mathcal{S}} = \{i^*\}$; $\tilde{\pi}_i(t)$ is the probability of state $i \in \tilde{\mathcal{S}}$ and $\tilde{\sigma}_i(t)$ is the expected sojourn time in state $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$ for the AMC at time t , given that it starts with an initial probability distribution $\tilde{\pi}_j(0) = \sum_{i \in \hat{\mathcal{S}}} \hat{\pi}_i(\tau) \cdot \mathbf{P}_{i,j}$, for $j \in \tilde{\mathcal{S}}$, at time 0.

3.1 Bounding analysis

Our assumptions imply that there is only one type of regeneration point, hence we can identify a renewal process whose renewal periods can be divided into two portions: the first one has a constant duration τ (the interval from when d becomes enabled to when it fires), the second one has a continuous phase-type (PH) distribution (the interval from the time d fires until it becomes enabled again).

Before starting our discussion, we observe that we can assume that we are interested in studying the behavior of the DSPN up to time t , for t greater than τ . This is because, for $t \leq \tau$, we know that the DSPN is still in the SMC at time t , and an exact transient analysis of the SMC will provide all the desired information with no approximation.

Theorem 3.1 Consider a DSPN with (only) one deterministic transition d that can become enabled only upon entering a unique state i^* and is persistent. Then, the number of visits to state i^* up to time t constitutes a renewal process.

Proof: We are assuming that the DSPN just entered state i^* at time 0, that is, $X(0) = i^*$ (if this were not the case, we would obtain a delayed renewal process). Then, if we let $S^{[0]} = 0$ and $S^{[n]}$ be the time at which state i^* is entered again for the n^{th} time, it is easy to see that the sequence of times

$\{T^{[n]} = S^{[n]} - S^{[n-1]} : n \geq 1\}$ are independent and identical distributed with the same cumulative distribution function F . Hence, process $\{N(t) = \max\{n \in \mathbf{N} : S^{[n]} = T^{[1]} + \dots + T^{[n]} \leq t\} : t \geq 0\}$ is a renewal process. \square

Definition 3.1 [18] A random variable $K \in \mathbf{N}$ is a stopping time for the sequence of independent random variables $\{A_1, A_2, \dots\}$ iff the event $\{K = n\}$, for any $n \in \mathbf{N}$, is independent of $\{A_{n+1}, A_{n+2}, \dots\}$. \square

Lemma 3.1 $N(t) + 1$ is a stopping time for the sequence $\{T^{[1]}, T^{[2]}, \dots\}$.

Proof: We can simply observe that $N(t) + 1 = n$ iff $\sum_{k=1}^{n-1} T^{[k]} \leq t < \sum_{k=1}^n T^{[k]}$, and the lemma follows from the independence of the random variables $\{T^{[1]}, T^{[2]}, \dots, T^{[n]}\}$ and $\{T^{[n+1]}, T^{[n+2]}, \dots\}$. \square

Lemma 3.2 For $i \in \mathcal{S}$, let $T_i^{[n]}$ be the sojourn time in state i during the n^{th} renewal period. Then $N(t) + 1$ is a stopping time for the sequence $\{T_i^{[1]}, T_i^{[2]}, \dots\}$.

Proof: For any $n > 0$, the random variables $\{T_i^{[n+1]}, T_i^{[n+2]}, \dots\}$ are independent of $\{T^{[1]}, \dots, T^{[n]}\}$, so $N(t) + 1 = n$ is independent of $\{T_i^{[n+1]}, T_i^{[n+2]}, \dots\}$. \square

Lemma 3.3 For i in \mathcal{S} , $E \left[\sum_{n=1}^{N(t)+1} T_i^{[n]} \right] = E[T_i] \cdot (m(t) + 1)$, where T_i is the sojourn time in state i during a generic renewal period.

Proof: Since $N(t) + 1$ is a stopping time for the sequence $\{T_i^{[1]}, T_i^{[2]}, \dots\}$, we can apply Wald's equation [18] and obtain

$$E \left[\sum_{n=1}^{N(t)+1} T_i^{[n]} \right] = E[T_i] \cdot E[N(t) + 1] = E[T_i] \cdot (m(t) + 1). \quad \square$$

Using the previous result, we can then obtain the following bounds for the average cumulative sojourn time in each state.

Theorem 3.2 The average time spent in state i up to time t is bounded by:

$$\begin{aligned} E[T_i] \cdot (m(t) + 1) - \tau &\leq \sigma_i(t) \leq E[T_i] \cdot (m(t) + 1), & \text{if } i \in \hat{\mathcal{S}}; \\ E[T_i] \cdot (m(t) + 1) - (E[T^{[N(t)+1]}] - \tau) &\leq \sigma_i(t) \leq E[T_i] \cdot (m(t) + 1), & \text{if } i \in \tilde{\mathcal{S}} \setminus \{i^*\}. \end{aligned}$$

Proof: We have

$$E \left[\sum_{n=1}^{N(t)} T_i^{[n]} \right] \leq \sigma_i(t) < E \left[\sum_{n=1}^{N(t)+1} T_i^{[n]} \right].$$

Lemma 3.3 proves the upper bound. For the lower bound, observe that

$$E \left[\sum_{n=1}^{N(t)} T_i^{[n]} \right] = E[T_i] \cdot (m(t) + 1) - E[T_i^{[N(t)+1]}]$$

and that $T_i^{[N(t)+1]}$ is at most τ for $i \in \hat{\mathcal{S}}$, and at most $T^{[N(t)+1]} - \tau$ for $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$. \square

The lower bound of Theorem 3.2 will not be tight for most states, since, in general, the time spent in a particular state i during the last renewal period will only be a small fraction of the entire renewal period itself. However, in practical applications, the vector of expected sojourn times is used to obtain a high-level measure by assigning a reward ρ_i to each state i , and computing the weighted sum $r(t) = \sum_{i \in \mathcal{S}} \sigma_i(t) \cdot \rho_i$. Then, the following bounds on $r(t)$ can be obtained.

Theorem 3.3 The value of $r(t)$, the expected accumulated reward up to time t corresponding to the reward rate assignment ρ_i , for $i \in \mathcal{S}$, is bounded by:

$$\begin{aligned} & (\sum_{i \in \mathcal{S}} E[T_i] \cdot \rho_i) \cdot (m(t) + 1) - \tau \cdot \max_{i \in \hat{\mathcal{S}}} \{\rho_i\} - (E[T^{[N(t)+1]}] - \tau) \cdot \max_{i \in \tilde{\mathcal{S}} \setminus \{i^*\}} \{\rho_i\} \leq r(t) \\ & \leq (\sum_{i \in \mathcal{S}} E[T_i] \cdot \rho_i) \cdot (m(t) + 1). \end{aligned}$$

Proof: For the upper bound, consider that

$$\begin{aligned} r(t) &= \sum_{i \in \mathcal{S}} \sigma_i(t) \cdot \rho_i \leq \sum_{i \in \mathcal{S}} E[T_i] \cdot (m(t) + 1) \cdot \rho_i \\ &= \left(\sum_{i \in \mathcal{S}} E[T_i] \cdot \rho_i \right) \cdot (m(t) + 1). \end{aligned}$$

For the lower bound, we have

$$\begin{aligned} r(t) &= \sum_{i \in \mathcal{S}} \sigma_i(t) \cdot \rho_i \\ &\geq \sum_{i \in \mathcal{S}} \left(E[T_i] \cdot (m(t) + 1) - E[T_i^{[N(t)+1]}] \right) \cdot \rho_i \\ &= \left(\sum_{i \in \mathcal{S}} E[T_i] \cdot \rho_i \right) \cdot (m(t) + 1) - \sum_{i \in \mathcal{S}} E[T_i^{[N(t)+1]}] \cdot \rho_i \\ &= \left(\sum_{i \in \mathcal{S}} E[T_i] \cdot \rho_i \right) \cdot (m(t) + 1) - \sum_{i \in \hat{\mathcal{S}}} E[T_i^{[N(t)+1]}] \cdot \rho_i - \sum_{i \in \tilde{\mathcal{S}} \setminus \{i^*\}} E[T_i^{[N(t)+1]}] \cdot \rho_i. \end{aligned}$$

And, by the same argument as in Theorem 3.2,

$$\begin{aligned} \sum_{i \in \hat{\mathcal{S}}} E[T_i^{[N(t)+1]}] \cdot \rho_i &\leq \tau \cdot \max_{i \in \hat{\mathcal{S}}} \{\rho_i\} \quad \text{and} \\ \sum_{i \in \tilde{\mathcal{S}} \setminus \{i^*\}} E[T_i^{[N(t)+1]}] \cdot \rho_i &\leq (E[T^{[N(t)+1]}] - \tau) \cdot \max_{i \in \tilde{\mathcal{S}} \setminus \{i^*\}} \{\rho_i\}. \quad \square \end{aligned}$$

Theorems 3.2 and 3.3 tell us that, if we can compute the average sojourn time during a generic renewal period for each state $i \in \mathcal{S}$, $E[T_i]$, the expected length of the last renewal interval, $E[T^{[N(t)+1]}]$, and the renewal function, $m(t)$, then we can derive bounds on the average sojourn time for each state $i \in \mathcal{S}$ and on the cumulative reward up to time t .

The first one of these quantities is straightforward. If $i \in \hat{\mathcal{S}}$, $E[T_i]$ is simply $\hat{\sigma}_i(\tau)$, the same quantity that needs to be computed for the stationary analysis of the DSPN. If $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$, $E[T_i]$ is simply $\tilde{\sigma}_i(\infty) = \lim_{t \rightarrow \infty} \tilde{\sigma}_i(t)$, a quantity that can be computed by solving the nonhomogeneous linear system

$$\tilde{\sigma}(\infty) \cdot \tilde{\mathbf{Q}}^* = -\tilde{\pi}(0)^*,$$

where the superscript “*” indicates the restriction of $\tilde{\mathbf{Q}}$ and $\tilde{\pi}(0)$ to the transient states (i.e., with the row and column corresponding to state i^* removed). See [5] for a detailed discussion of efficient methods to compute $\tilde{\sigma}(\infty)$.

We then focus on the computation of the renewal function $m(t)$ and of the expected length of the last renewal interval $E[T^{[N(t)+1]}]$. The difficulty with using the formula

$$m(t) = \sum_{n=1}^{\infty} F^{[n]}(t)$$

to compute the renewal function is that the determination of $F^{[n]}(t) = P\{T^{[1]} + \dots + T^{[n]} \leq t\}$ requires the computation of an n -dimensional integral. Ross [17] proposed an efficient algorithm that requires only one-dimensional integrals in input.

Theorem 3.4 (Ross [17]) For a renewal process where the renewal interval has pdf $f(x)$, define

$$m_r = \int_0^{\infty} \sum_{k=0}^{r-1} (1 + m_{r-k}) \frac{e^{-\lambda x} (\lambda x)^k}{k!} f(x) dx, \quad \text{for } r = 1, 2, \dots, n \text{ and } \lambda = n/t.$$

Then, if $m(t)$ is continuous at t , m_n converge to $m(t)$ as n goes to infinity. □

This theorem tells us that if we know the distribution F of the renewal interval, we can approximate the renewal function arbitrarily well.

The reason for having to bound $E[T^{[N(t)+1]}]$ and $E[T_i^{[N(t)+1]}]$ instead of computing their actual value is due to the well-known inspection paradox: the length of $T^{[N(t)+1]}$ is in general greater than that of a typical renewal interval. For example, consider the case of renewal intervals having an exponential distribution with parameter λ , and assume that the renewal process has been “going on forever” (i.e., in the limit for $t \rightarrow \infty$). Then, the expected length of the renewal interval containing a given time instant t is $2/\lambda$, *twice* as long as that of the generic interval. An intuitive explanation for this is that both the *age*, $t - S^{[N(t)]}$, and the *remaining lifetime*, $S^{[N(t)+1]} - t$, are exponentially distributed with parameter λ . For finite values of t , however, the age is bounded by t , so, in our example, its expected value is actually that of an exponential random variable

truncated at t , $(1 - e^{-\lambda t})/\lambda$. The expected remaining lifetime is instead, of course, $1/\lambda$, hence $E[T^{[N(t)+1]}] = (2 - e^{-\lambda t})/\lambda$, less than $2/\lambda$, but still larger than the average length of a typical renewal interval, $1/\lambda$.

Unless the distribution F is analytically known and tractable, it is computationally difficult to take into account the fact that the age is no greater than t . However, we can always ignore this restriction and compute an upper bound on $E[T^{[N(t)+1]}]$ by using the relation [15, p. 269]

$$\lim_{t \rightarrow \infty} E[T^{[N(t)+1]}] = E[T^2]/E[T]. \quad (1)$$

Since our renewal interval is described as the constant τ plus the time to absorption in a CTMC, and each visit to a state of the SMC or CTMC is exponentially distributed, we can tighten the bounds in the acyclic case.

Theorem 3.5 If the SMC is acyclic, for any state i in $\hat{\mathcal{S}}$, $\sigma_i(t)$ satisfies

$$E[T_i] \cdot (m(t) + 1) - \min\{\tau, 2E[T_i]\} \leq \sigma_i(t).$$

If the AMC is acyclic, for any state i in $\tilde{\mathcal{S}} \setminus \{i^*\}$, $\sigma_i(t)$ satisfies

$$E[T_i] \cdot (m(t) + 1) - \min\{E[T^{[N(t)+1]}] - \tau, 2E[T_i]\} \leq \sigma_i(t).$$

Proof: For a state $i \in \hat{\mathcal{S}}$ (the case $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$ is analogous), exactly one of these three events must occur:

- e_1 : In the last renewal period (the one containing time t), the DSPN does not visit state i ;
- e_2 : The DSPN is in state i at time t ;
- e_3 : In the last renewal period, the DSPN visits state i , but, at time t , it is not in state i .

Conditioning on these mutually exclusive events, we have

$$\begin{aligned} E[T_i^{[N(t)+1]}] &= E[T_i^{[N(t)+1]}|e_1] \cdot \Pr\{e_1\} + E[T_i^{[N(t)+1]}|e_2] \cdot \Pr\{e_2\} + E[T_i^{[N(t)+1]}|e_3] \cdot \Pr\{e_3\} \\ &= 0 \cdot \Pr\{e_1\} + 2E[T_i] \cdot \Pr\{e_2\} + E[T_i] \cdot \Pr\{e_3\} \\ &\leq 2E[T_i]. \end{aligned}$$

By the same argument as in Theorem 3.2, we obtain the result. □

3.2 Approximate analysis

We now consider a slightly different approach, where we seek to heuristically approximate the exact value of $\sigma_i(t)$. Let $A(t)$ be the age of the current renewal interval at time t , $A(t) = t - S^{[n]}$, and

$A_i(t)$ be the time spent in state $i \in \mathcal{S}$ up to time t during the current renewal interval, so that $A(t) = \sum_{i \in \mathcal{S}} A_i(t)$. Then,

$$\begin{aligned} \sigma_i(t) &= E[T_i^{[1]} + T_i^{[2]} + \dots + T_i^{[N(t)]} + A_i(t)] \\ &= \sum_{n=0}^{N(t)} E[T_i^{[1]} + \dots + T_i^{[N(t)]} | N(t) = n] \cdot \Pr\{N(t) = n\} + \sum_{n=0}^{N(t)} E[A_i(t) | N(t) = n] \cdot \Pr\{N(t) = n\} \end{aligned}$$

We can approximate the first summation by:

$$m(t) \cdot E[T_i] = \begin{cases} m(t) \cdot \hat{\sigma}_i(\tau) & \text{if } i \in \hat{\mathcal{S}} \\ m(t) \cdot \tilde{\sigma}_i(\infty) & \text{if } i \in \tilde{\mathcal{S}} \setminus \{i^*\} \end{cases}$$

and the second summation, after defining $a = t - m(t) \cdot E[T] \approx E[A(t)]$, by:

$$\begin{cases} \hat{\sigma}_i(\tau) & \text{if } i \in \hat{\mathcal{S}} \text{ and } a > \tau \\ \hat{\sigma}_i(a) & \text{if } i \in \hat{\mathcal{S}} \text{ and } a \leq \tau \\ \tilde{\sigma}_i(a - \tau) & \text{if } i \in \tilde{\mathcal{S}} \setminus \{i^*\} \text{ and } a > \tau \\ 0 & \text{if } i \in \tilde{\mathcal{S}} \setminus \{i^*\} \text{ and } a \leq \tau \end{cases}.$$

This heuristic is appealing because it attempts to capture the fact that the DSPN alternates between SMC and AMC periods, *starting in the SMC at time 0*, by allocating our best guess of the age at time t , a , to the SMC, and only the remaining part, if any, to the AMC. Moreover, it also allows to capture the fact that, at time 0, the DSPN (hence the SMC) is in state i^* , an essential characteristic of transient analysis.

We observe that, while this approximation cannot be shown to be a lower or upper bound for $\sigma_i(t)$, it is nevertheless guaranteed to fall within the bounds we defined in the previous section, since, for $i \in \hat{\mathcal{S}}$,

$$\hat{\sigma}_i(\tau) \cdot (m(t) + 1) - \tau \leq m(t) \cdot \hat{\sigma}_i(\tau) + \hat{\sigma}_i(a) \leq m(t) \cdot \hat{\sigma}_i(\tau) + \hat{\sigma}_i(\tau) = \hat{\sigma}_i(\tau) \cdot (m(t) + 1),$$

while, for $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$,

$$\tilde{\sigma}_i(\infty) \cdot (m(t) + 1) - (E[T^{[N(t)+1]}] - \tau) \leq m(t) \cdot \tilde{\sigma}_i(\infty) \leq m(t) \cdot \tilde{\sigma}_i(\infty) + \tilde{\sigma}_i(a - \tau) \leq \tilde{\sigma}_i(\infty) \cdot (m(t) + 1).$$

3.3 Computational algorithm

Our approach hinges on the ability to obtain the distribution F of the renewal interval T . Since the length of this interval is $T = \tau + Y$, where τ is a constant, we know that $F(u) = 0$ for $u < \tau$. For $u \geq \tau$, instead, $F(u)$ is the probability that the AMC is absorbed in state i^* by time $u - \tau$, given that it is started with the initial probability distribution $\tilde{\pi}(0)$ previously defined. Thus, computing F numerically only requires us to perform a transient analysis on the AMC to obtain the probability

of being in state i^* at times $0, y_1, y_2, \dots, y_{MAX}$, appropriately chosen so that this discretization of F is a good approximation of its continuous behavior (in particular $F(\tau + y_{MAX})$ should be very close to one). The detailed steps required for our bounding approach are then the following (these should be performed only if the conditions of Theorem 3.1 hold):

1. Derive the SMC and the AMC defined by the model, and compute their state spaces $\hat{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ and infinitesimal generators $\hat{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$. Also, derive, for each state $i \in \hat{\mathcal{S}}$, the probability $\mathbf{P}_{i,j}$ that state $j \in \tilde{\mathcal{S}}$ is reached when d fires in i ,
2. Perform an instantaneous and cumulative transient analysis of the SMC: starting from the initial state i^* , and for each state $i \in \hat{\mathcal{S}}$ of the SMC, compute the probability of being in i at time τ , $\hat{\pi}_i(\tau)$, and the cumulative sojourn time in i during the interval $[0, \tau]$, $\hat{\sigma}_i(\tau)$.
3. For each state $i \in \hat{\mathcal{S}}$ of the SMC, obtain (from the model) the probability $\mathbf{P}_{i,j}$ that state $j \in \tilde{\mathcal{S}}$ of the AMC is reached when d fires in i , and use these quantities to compute the initial probability distribution of the AMC, $\tilde{\pi}_j(0) = \sum_{i \in \hat{\mathcal{S}}} \hat{\pi}_i(\tau) \cdot \mathbf{P}_{i,j}$.
4. Compute the probability $\Pr\{Y \leq y\}$ that the AMC is in state i^* at times $y = 0, y_1, y_2, \dots, y_{MAX}$, then define a discretization of F using the values

$$F(u) = \begin{cases} 0 & \text{if } u < \tau, \\ \Pr\{Y \leq y\} & \text{if } u = \tau + y, \quad \text{for } y \in \{0, y_1, \dots, y_{MAX}\}. \end{cases}$$

5. Compute the expected sojourn time $\tilde{\sigma}_i(\infty)$ in each transient state $i \in \tilde{\mathcal{S}} \setminus \{i^*\}$ of the AMC until absorption, starting from the initial distribution $\tilde{\pi}(0)$, by solving the linear system

$$\tilde{\sigma}(\infty) \cdot \tilde{\mathbf{Q}}^* = -\tilde{\pi}(0)^*.$$

6. For each time t of interest, compute an approximation of the average number of renewal periods $m(t)$ using Ross's method.
7. Estimate $E[T^{[N(t)+1]}]$. For finite values of t , $E[T^{[N(t)+1]}] \leq E[T^2]/E[T]$, hence $E[T^2]/E[T]$ is a safe upper bound. $E[T^2]$ can be obtained from the discretization of F , but $E[T]$ is even simpler: $E[T] = \tau + \sum_{i \in \tilde{\mathcal{S}} \setminus \{i^*\}} \tilde{\sigma}_i(\infty)$. In cases where the function F is analytically known, as in our example of Sect. 5, we might be able to estimate a tighter upper bound that takes into account the actual value of t .
8. For each time t of interest, compute the bounds for the average total sojourn time in each state or the total accumulated reward using Theorems 3.2 and 3.3.

For our heuristic approximation approach, the steps are similar, except that we also need to perform a transient analysis solution, for the SMC at time a , if $a < \tau$, or for the AMC at time $a - \tau$, if $a > \tau$, for each time t of interest, where $a = t - m(t) \cdot E[T]$.

We conclude this section with a few observation about the complexity of the approach we propose. The transient analysis of the SMC can be performed using the uniformization algorithm, which has a complexity $O(\eta(\hat{\mathbf{Q}}) \cdot \hat{q} \cdot \tau)$, where $\eta(\cdot)$ is an operator indicating the number of nonzero entries in the argument matrix and \hat{q} is the uniformization rate for the SMC, i.e., $\hat{q} = \max_i \{-\hat{\mathbf{Q}}_{i,i}\}$ [16].

The study of the AMC at times y_1, y_2, \dots, y_{MAX} can also be performed using uniformization. The same AMC is studied at those different times. It is then possible to compute all these MAX probabilities essentially at the cost of the last one, $\Pr\{Y \leq y_{MAX}\}$, since the number of iterations required by the uniformization algorithm is dictated by the largest transient time at which solution is required. Hence, for this portion of the analysis, the complexity is $O(\eta(\tilde{\mathbf{Q}}) \cdot \tilde{q} \cdot y_{MAX})$, where \tilde{q} is now the uniformization rate for the AMC. In this case, y_{MAX} can be of course quite large, since we need to discretize F along its entire range of values. However, steady-state detection [13] can be employed to reduce the computational cost when the value chosen for y_{MAX} happens to be needlessly large (when $\tilde{q} \cdot y_{MAX}$ is so large that the uniformized DTMC reaches numerical steady-state in far fewer than $\tilde{q} \cdot y_{MAX}$ iterations).

Finally, the third computational component is spent to obtain an approximation to $m(t)$ using Ross's method. This requires $O(n^2 + n \cdot MAX)$ operations, where n is a value chosen beforehand. Since, essentially, the idea uses an n -stage Erlang to approximate a constant, small values of n (e.g., 10 or 20) are usually adequate, and the cost of computing $m(t)$ is comparatively negligible.

4 A more general class of DSPNs

We now study a more general class of DSPN than the one we considered so far. We still assume that there is exactly one deterministic transition d , with constant firing time τ , and that d is persistent. However, instead of requiring that there exists only one state i^* where d can become enabled, we allow d to become enabled in any state $i \in \hat{\mathcal{S}}$, as long as the length of the renewal period is independent of i . While this condition can be hard to check in general, it can be easily verified by inspection for many models; this is the case in the example we present in Sect. 5.

Given the independence of the time to absorption in the AMC and the initial state of the SMC, we can still identify a renewal process $\{N(t) : t \geq 0\}$ as for the simpler case previously considered. $N(t)$ counts the number of times d becomes enabled in $(0, t]$. Now, however, we also need to consider the processes $\{\hat{X}^{[n]} : n \in \mathbf{N}\}$ and $\{\tilde{X}^{[n]} : n \in \mathbf{N}\}$, where $\hat{X}^{[n]}$ and $\tilde{X}^{[n]}$ are the initial states of the SMC and of the AMC, respectively, for the n^{th} renewal period. Given our assumptions, it is easy to see that these two processes are DTMCs, and that they are independent of $\{N(t) : t \geq 0\}$.

We now discuss an approximation for a type of Markov regenerative process having iid renewal periods $(T^{[1]}, T^{[2]}, \dots)$ that are independent of the Markovian states $(i^{[1]}, i^{[2]}, \dots)$ in which the periods are started. Then, we will specialize the computational procedure to the case where the renewal periods are given by a constant time τ spent in the SMC plus the time to absorption in

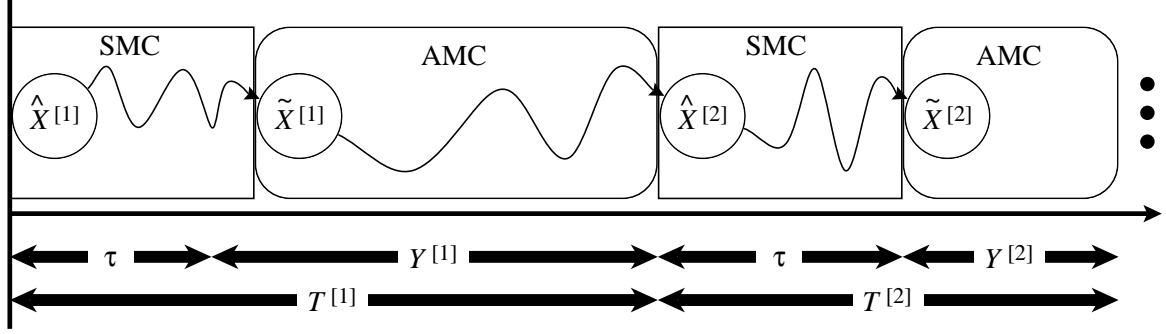


Figure 2: The evolution in time of our MRGP.

the AMC.

Let $T_i^{[n]}(t)$ be the amount of time spent in state i during $[S^{[n-1]}, S^{[n]}] \cap [0, t]$, that is, during the intersection of the n^{th} period with the interval of interest for our study. Clearly,

- $T_i^{[n]}(t) = T_i^{[n]}$, if $n \leq N(t)$,
- $T_i^{[n]}(t) = A_i(t)$, if $n = N(t) + 1$, and
- $T_i^{[n]}(t) = 0$, if $n > N(t) + 1$,

and the total amount of time spent in state i during $[0, t]$ is

$$T_i^{[1]} + \dots + T_i^{[N(t)]} + A_i(t) = \sum_{n=1}^{\infty} T_i^{[n]}(t). \quad (2)$$

The renewal periods can start in any state of $\hat{\mathcal{S}}$, hence we consider the computation of $\sigma_i(t)$ conditioned on the initial state:

$$\begin{aligned} \sigma_i(t|X^{[1]} = k) &= \sum_{n=1}^{\infty} E [T_i^{[n]}(t)|X^{[1]} = k] \\ &= E [T_i^{[1]}(t)|X^{[1]} = k] + \sum_{n=2}^{\infty} E [T_i^{[n]}(t)|X^{[1]} = k] \\ &= E [T_i^{[1]}(t)|X^{[1]} = k] + \sum_{l \in \hat{\mathcal{S}}} \sum_{n=2}^{\infty} E [T_i^{[n]}(t)|X^{[1]} = k \wedge X^{[2]} = l] \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} \\ &= E [T_i^{[1]}(t)|X^{[1]} = k] + \int_0^{\infty} \sum_{l \in \hat{\mathcal{S}}} \sum_{n=2}^{\infty} E [T_i^{[n]}(t)|X^{[1]} = k \wedge X^{[2]} = l \wedge T^{[1]}(t) = u] \\ &\quad \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} d\Pr\{T^{[1]}(t) \leq u|X^{[1]} = k \wedge X^{[2]} = l\} \end{aligned}$$

$$\begin{aligned}
&= E [T_i^{[1]}(t)|X^{[1]} = k] + \int_0^t \sum_{l \in \hat{\mathcal{S}}} \sum_{n=2}^{\infty} E [T_i^{[n]}(t)|X^{[2]} = l \wedge T^{[1]}(t) = u] \\
&\quad \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} d\Pr\{T^{[1]}(t) \leq u\} \\
&= E [T_i^{[1]}(t)|X^{[1]} = k] + \sum_{l \in \hat{\mathcal{S}}} \left(\int_0^t \sigma_i(t-u|X^{[1]} = l) d\Pr\{T^{[1]}(t) \leq u\} \right) \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} \\
&\approx E [T_i^{[1]}(t)|X^{[1]} = k] + \sum_{l \in \hat{\mathcal{S}}} \sigma_i(t - E[T^{[1]}(t)]|X^{[1]} = l) \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\}
\end{aligned}$$

Note that the only approximation is introduced in the last step of the derivation, where the complex integral in parentheses is substituted with the expected value $\sigma_i(t - E[T^{[1]}(t)]|X^{[1]} = l)$. Also, observe that, in principle, the computation of $\Pr\{X^{[2]} = l|X^{[1]} = k\}$ would require us to compute limiting probabilities (i.e., the solution of a linear system for the absorption probabilities in the states $\tilde{\mathcal{S}} \cap \hat{\mathcal{S}}$ of the AMC in our case). However, we can use instead the results from the transient cumulative analysis of the first renewal period, which must be performed anyway to obtain $E[T_i^{[1]}(t)|X^{[1]} = k]$. This is because, given the independence of the length of the renewal period from the state at the beginning of the period, we can write, for any $k, l \in \hat{\mathcal{S}}$ such that entrance in k and l are renewal points, and any time u ,

$$\begin{aligned}
\frac{\Pr\{X^{[1]}(u) = l|X^{[1]}(0) = k\}}{\Pr\{T^{[1]} \leq u\}} &= \frac{\Pr\{X^{[1]}(u) = l \wedge X^{[1]}(0) = k\}}{\Pr\{T^{[1]} \leq u\} \cdot \Pr\{X^{[1]}(0) = k\}} \\
(\text{since } X^{[1]}(u) = l \text{ implies } T^{[1]} \leq u) &= \frac{\Pr\{X^{[1]}(u) = l \wedge X^{[1]}(0) = k \wedge T^{[1]} \leq u\}}{\Pr\{T^{[1]} \leq u \wedge X^{[1]}(0) = k\}} \\
&= \Pr\{X^{[1]}(u) = l|X^{[1]}(0) = k \wedge T^{[1]} \leq u\} \\
&= \Pr\{X^{[2]} = l|X^{[1]} = k\}
\end{aligned}$$

In other words, the relative values of the absorption probabilities in the various states of the AMC are time invariant. While the absorption probabilities sum to one only in the limit, for $t \rightarrow \infty$, a simple normalization of the absorption probabilities at any finite time t yields the same values. Furthermore, again because of our independence assumption, the absorption probabilities are proportional to the transient cumulative sojourn times in the absorbing states, so only a transient cumulative analysis is needed.

4.1 Computational algorithm

We can now specialize the above idea to our particular application. The detailed steps to obtain an approximation σ^{app} to the vector $\sigma(t)$ are:

1. Derive the SMC and the AMC defined by the model, their state spaces $\hat{\mathcal{S}}$ and $\tilde{\mathcal{S}}$, their infinitesimal generators $\hat{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$, and for each state $i \in \hat{\mathcal{S}}$, the probability $\mathbf{P}_{i,j}$ that state

- $j \in \tilde{\mathcal{S}}$ is reached when d fires in i . Derive the initial probability distribution $\hat{\pi}(0)$ and choose the time horizon t . For each $i \in \hat{\mathcal{S}} \cup \tilde{\mathcal{S}}$, initialize σ_i^{app} to zero.
2. If $t \leq \tau$, perform a cumulative transient analysis of the SMC: starting from the initial probability $\hat{\pi}(0)$, compute the expected cumulative sojourn time in each state $i \in \hat{\mathcal{S}}$, during the interval $[0, t]$, $\hat{\sigma}_i(t)$, and add it to σ_i^{app} . Stop.
 3. Otherwise, perform an instantaneous and a cumulative transient analysis of the SMC: starting from the initial probability vector $\hat{\pi}(0)$, compute the probability of being in each state $i \in \hat{\mathcal{S}}$ at time τ , $\hat{\pi}_i(\tau)$, and the expected cumulative sojourn time in i during the interval $[0, \tau]$, $\hat{\sigma}_i(\tau)$. Add $\hat{\sigma}_i(\tau)$ to σ_i^{app} .
 4. Compute the initial probability of each state $j \in \tilde{\mathcal{S}}$ of the AMC as $\tilde{\pi}_j(0) = \sum_{i \in \hat{\mathcal{S}}} \hat{\pi}_i(\tau) \cdot \mathbf{P}_{i,j}$.
 5. Perform a cumulative transient analysis of the AMC: starting from the initial probability distribution $\tilde{\pi}(0)$, for each state $i \in \tilde{\mathcal{S}}$, compute the cumulative sojourn time in i during the interval $[0, t - \tau]$, $\tilde{\sigma}_i(t - \tau)$. For each transient state $i \in \tilde{\mathcal{S}} \setminus \hat{\mathcal{S}}$, add $\tilde{\sigma}_i(t - \tau)$ to σ_i^{app} . Compute the total amount of time spent in the absorbing states of the AMC up to time $t - \tau$ as $\alpha = \sum_{i \in \tilde{\mathcal{S}} \cap \hat{\mathcal{S}}} \tilde{\sigma}_i(t - \tau)$.
 6. Update the time horizon by setting t to α and the initial probability vector of the SMC: $\hat{\pi}_i(0) \leftarrow 0$ if $i \in \hat{\mathcal{S}} \setminus \tilde{\mathcal{S}}$; $\hat{\pi}_i(0) \leftarrow \tilde{\sigma}_i(t - \tau)/\alpha$ if $i \in \hat{\mathcal{S}} \cap \tilde{\mathcal{S}}$. Go to step 2

With any approximation, it is important to assess both its complexity and its quality. The following theorem bounds the number of iterations $I(t)$ required by our second approximation approach, thus allowing us to estimate its complexity in terms of the number of transient solutions of the SMC and AMC we must perform

Theorem 4.1 For any $t > 0$, the number of iterations $I(t)$ required by our computational procedure satisfies

$$\left\lceil \frac{t}{\tau + E[Y]} \right\rceil \leq I(t) \leq \left\lceil \frac{t}{\tau} \right\rceil$$

Proof: At each iteration, the time horizon is decreased by at least τ , hence when we stop, we have $(I(t) - 1) \cdot \tau < t$, which implies $I(t) < t/\tau + 1$. We know $I(t)$ is an integer, so $I(t) \leq \lceil t/\tau \rceil$. On the other hand, at each iteration we decrease the time horizon by $\tau + E[Y|Y < t - \tau]$, which is no greater than $\tau + E[Y]$. Hence, $(\tau + E[Y]) \cdot I(t) \geq t$, which implies the lower bound. \square

The second theorem addresses instead the quality of the approximation, in a limiting sense, by showing that, as the time horizon t increase, we obtain asymptotically exact results. This, paired with the fact that we also obtain exact results for $t \leq \tau$, and likely excellent approximations for $t < E[T]$, should increase our confidence in this approximation.

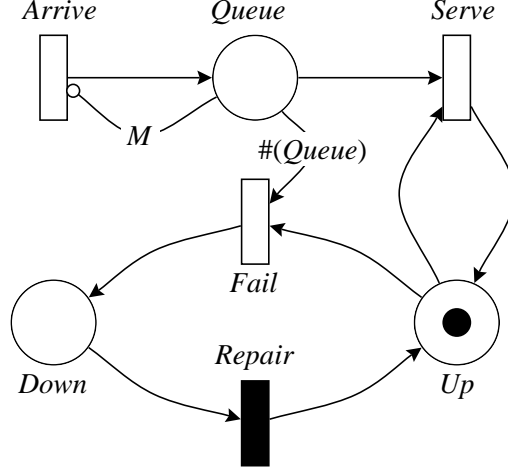


Figure 3: An example of our class of DSPNs.

Theorem 4.2 If the renewal periods have finite expectation and the DSPN is ergodic, in the sense that $\lim_{t \rightarrow \infty} \sigma_i(t|X^{[1]} = k)/t$ exists and equals $\pi_i > 0$, independent of k , then the value σ_i^{app} computed by our procedure satisfies

$$\lim_{t \rightarrow \infty} \frac{\sigma_i^{app}(t)}{\sigma_i(t)} = 1.$$

Proof: We have

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\sigma_i^{app}(t)}{t} &= \sum_{l \in \hat{\mathcal{S}}} \lim_{t \rightarrow \infty} \frac{\sigma_i(t - E[T^{[1]}(t)|X^{[1]} = l])}{t} \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} \\ &= \sum_{l \in \hat{\mathcal{S}}} \lim_{t \rightarrow \infty} \frac{\sigma_i(t - E[T^{[1]}(t)|X^{[1]} = l])}{t - E[T^{[1]}(t)|X^{[1]} = l]} \cdot \frac{t - E[T^{[1]}(t)|X^{[1]} = l]}{t} \cdot \Pr\{X^{[2]} = l|X^{[1]} = k\} \\ &= \pi_i \cdot \sum_{l \in \hat{\mathcal{S}}} \Pr\{X^{[2]} = l|X^{[1]} = k\} \\ &= \pi_i. \end{aligned}$$

Hence,

$$\lim_{t \rightarrow \infty} \frac{\sigma_i^{app}(t)}{\sigma_i(t)} = \lim_{t \rightarrow \infty} \frac{\sigma_i^{app}(t)}{t} \cdot \lim_{t \rightarrow \infty} \frac{t}{\sigma_i(t)} = \pi_i \cdot \frac{1}{\pi_i} = 1. \quad \square$$

5 Example

We now present an application of our techniques. Consider the DSPN shown in Fig. 3, modeling a finite-capacity queue where the server is subject to breakdowns. We use the notation $\#(place, i)$

to indicate the number of tokens in *place* when the marking is i and $\lambda(\textit{transition}, i)$ to indicate the firing rate for *transition* when the marking is i (the marking is omitted if the “current state” is intended).

Transitions *Arrive* and *Serve* model the arrival and service processes, respectively. Tokens in place *Queue* represent customers in the queue, including the one(s) in service. The maximum number of customers that can fit in the queue is M (the inhibitor arc from *Queue* to *Arrive* enforces this limit). The server is working when there is a token in place *Up*, and it is being repaired when there is instead a token in *Down*. The failure of the server is represented by transition *Fail*. We consider two cases. In the first model, the firing of *Fail* removes all the tokens from *Queue* (the “flushing arc” with cardinality $\#(\textit{Queue})$ from *Queue* to *Fail* achieves this effect); this model satisfies the stricter assumptions of our first approximation. In the second model, the flushing arc is not present, so customers in the queue are not lost because of a failure; this second model satisfies the assumptions for our second approximation, but not for the the first one. We stress that arrivals can restart again immediately after a failure, without having to wait for the completion of the repair. If this were not the case, we could of course add an inhibitor arc from place *Down* to transition *Arrive*, and the process would be even simpler (e.g., in the first model, the SMC would consist of the single absorbing marking $\{\textit{Down}\}$).

The repair of the server is the only activity with a deterministic duration. The conditions for applying our approximations are satisfied as long as the timing of the other transitions in our model is as follows:

- Transition *Arrive* has an exponential distribution. A PH distribution can also be used, but it must be reset by the firing of transition *Fail* if we want to use the first approximation. This is essential because, when a failure occurs and the deterministic transition *Repair* becomes enabled, no memory of the past, including the phase in which the arrival process was, can be maintained in this case.
- Transition *Serve* can have an arbitrary PH distribution, again reset by the firing of transition *Fail* if we want to use the first approximation. This is equivalent to assuming that the failure of the server destroys any work in progress. Note that the rate of the service can be marking-dependent, allowing us to model a multiple or infinite server behavior, as long as the failures and repairs affect the entire service station as a whole, and not individual servers.
- Transition *Fail* can have an arbitrary PH distribution. No restrictions need to be placed on the reset behavior of this transition, since, like *Repair*, it is persistent.

We consider the following transient measures, time averaged over the interval $[0, t]$:

- The expected customer throughput:

$$\frac{1}{t} \cdot \sum_{i \in \tilde{\mathcal{S}}} \sigma_i(t) \cdot \lambda(\textit{Serve}, i).$$

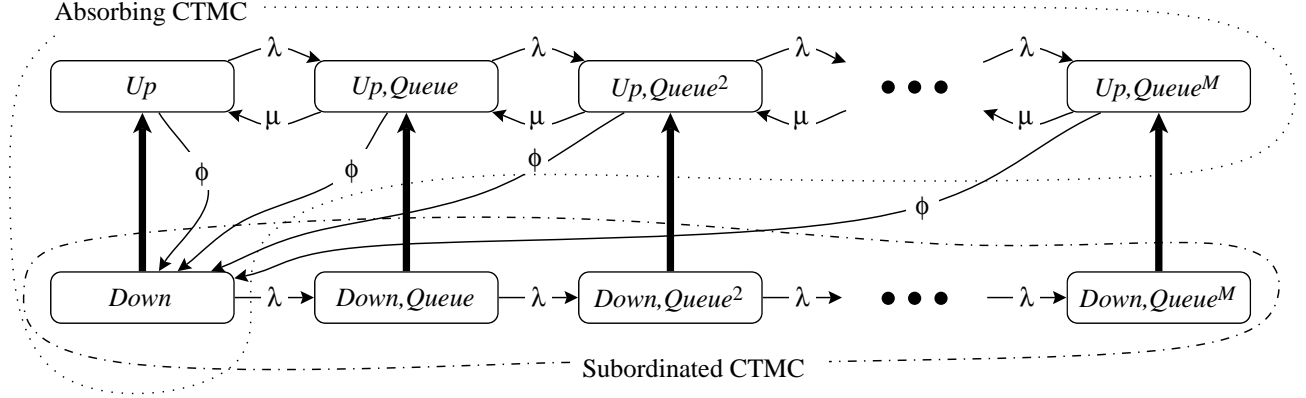


Figure 4: The process underlying our DSPN (exponential failure process).

- The expected number of customers waiting to be serviced or in service:

$$\frac{1}{t} \cdot \sum_{i \in \mathcal{S}} \sigma_i(t) \cdot \#(Queue, i)$$

(including customers lost due to a server failure; there does not seem to be a simple way to restrict this measure to customers who complete service).

Fig. 4 shows the stochastic process corresponding to the first model when the nondeterministic transitions have exponential distributions with rate λ (*Arrive*), μ (*Serve*), and ϕ (*Fail*). Markings are represented using bag notation: for example, marking $\{Up, Queue^2\}$ means that there is one token in place *Up*, two tokens in place *Queue*, and no tokens elsewhere. The heavy arcs indicate the firing of the deterministic transition. If the failure process has instead an Erlang distribution with K phases, each exponentially distributed with rate ϕ , the underlying process is shown in Fig. 5, where the component “ P ” of the state indicates the phase of the Erlang distribution. The analogous stochastic processes for the second model are exactly the same, except that the arc with failure rate ϕ leaving from $\{Up, Queue^n\}$ (in the exponential case) or from $\{Up, Queue^n, P^K\}$ (in the Erlang case) goes to $\{Down, Queue^n\}$, not to $\{Down\}$, and the AMC now includes all the down states, not just $\{Down\}$.

5.1 Bounds and first approximation

In our example, the distribution F of the renewal periods T is actually known in closed form, since $T = \tau + Y$ and Y is either $Expo(\phi)$ or $Erlang(K, \phi)$, depending on the type of failure process. This is because the arrival and service processes do not affect the duration of the renewal intervals.

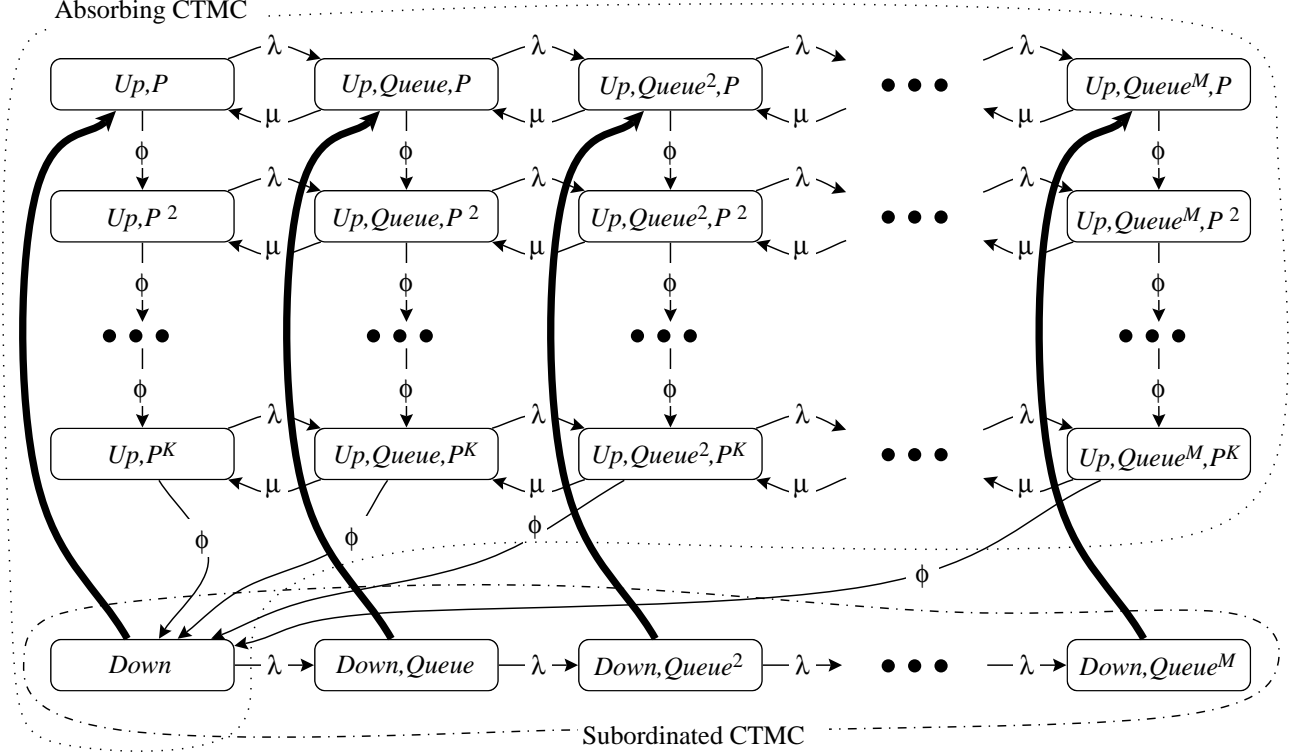


Figure 5: The process underlying our DSPN (Erlang failure process).

Hence, we can compute the value of $E[T^{[N(t)+1]}]$ analytically. If we ignore that the age of the renewal interval is bounded by t , we can use the relation (1) and obtain:

$$\lim_{t \rightarrow \infty} E[T^{[N(t)+1]}] = \frac{E[T^2]}{E[T]} = \frac{\tau^2 + 2\tau E[Y] + E[Y^2]}{\tau + E[Y]} = \left(\tau + \frac{1}{\phi} \right) + \frac{1}{\phi^2 \tau + \phi}$$

if the failure process is exponential (hence $E[Y^2] = 2/\phi^2$) and

$$\lim_{t \rightarrow \infty} E[T^{[N(t)+1]}] = \frac{E[T^2]}{E[T]} = \frac{\tau^2 + 2\tau E[Y] + E[Y^2]}{\tau + E[Y]} = \left(\tau + \frac{K}{\phi} \right) + \frac{1}{\phi^2 \tau / K + \phi}$$

if it is Erlang (hence $E[Y^2] = (K^2 + K)/\phi^2$).

We can also attempt to use the fact that the time spent in the AMC up to time t (which we can assume greater than τ) cannot exceed $t - \tau$. For a worst-case scenario with the exponential failure process, we assume that the DSPN is in the AMC at time t , so its expected ‘‘age’’ in the AMC is $1 - e^{-\phi(t-\tau)}/\phi$ and

$$E[T^{[N(t)+1]}] \leq \left(\tau + \frac{1}{\phi} \right) + \frac{1 - e^{-\phi(t-\tau)}}{\phi}.$$

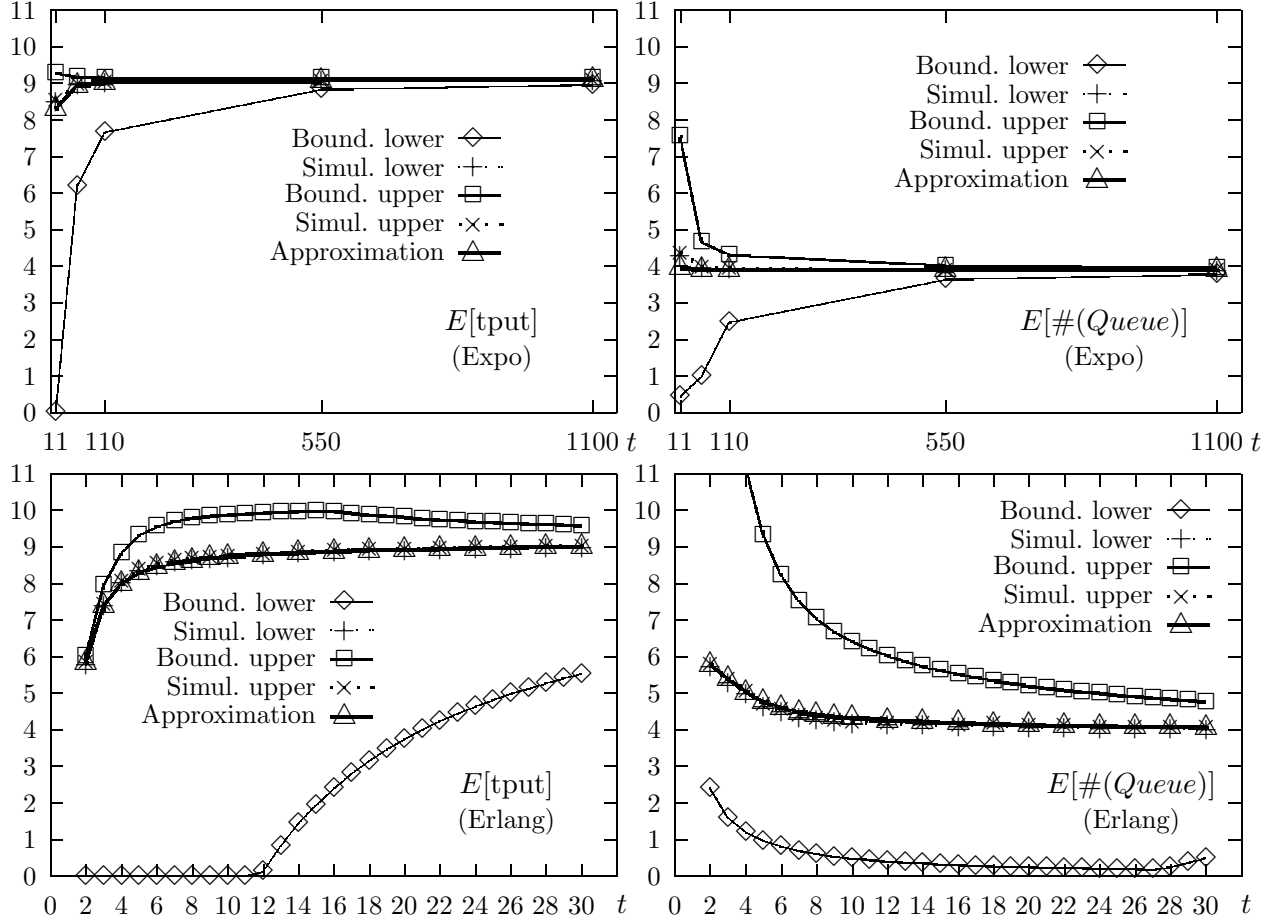


Figure 6: Numerical results for an exponential or Erlang failure process (first model).

In the Erlang case, we also assume that the the failure process is in the first phase of the Erlang distribution at time t , not just in any state of the AMC, resulting in

$$E[T^{[N(t)+1]}] \leq \left(\tau + \frac{K}{\phi} \right) + \frac{1 - e^{-\phi(t-\tau)}}{\phi}.$$

These alternative bounds explicitly state that the age is at most τ , spent in the SMC, plus an exponentially distributed amount bounded by $t - \tau$, spent in the AMC, and the remaining lifetime has the entire expectation $E[Y]$.

We can then combine the two approaches and obtain the upper bound on $E[T^{[N(t)+1]}]$,

$$\left(\tau + \frac{1}{\phi} \right) + \min \left\{ \frac{1}{\phi^2 \tau + \phi}, \frac{1 - e^{-\phi(t-\tau)}}{\phi} \right\} \quad \text{and} \quad \left(\tau + \frac{K}{\phi} \right) + \min \left\{ \frac{1}{\phi^2 \tau / K + \phi}, \frac{1 - e^{-\phi(t-\tau)}}{\phi} \right\}.$$

for the exponential and Erlang case, respectively. These are used to compute the results for our model, reported in Fig. 6. Each set of plots shows the lower and upper bounds using our bounding

technique, the lower and upper value of the 95% confidence intervals computed using simulation in the tool SPNP [11], and the value obtained using our first approximation (our second approximation can also be applied to this model, of course, but it is computationally more expensive and achieves analogous results in this case). The numerical values of the parameters are $M = 10$ (this limit is used also in the simulation for comparison purposes, even if the simulation does not require to limit the number of tokens in place *Queue*, as long as the system is stable), $\lambda = 10$, $\mu = 12$, $\tau = 1$, and $\phi = 0.1$ (exponential case) or $K = 5$ and $\phi = 0.5$ (Erlang case).

For an exponential failure process, we studied the system at time 11, 55, 110, 550, and 1100 (i.e., 1, 5, 10, 50, and 100 times the average renewal period). It is apparent that steady state is reached quite fast, and that our bounding technique is correct, but results in bounds that might be too wide in practice. Hence, for an Erlang failure process, we focus on the earlier portion of the system evolution, from time 2 to time 30, before steady state is reached. Especially for the throughput, the upper bound is meaningful.

Even more interesting, however, is the performance of our heuristic approximation, which follows very closely the results obtained from simulation. We believe this to be a very encouraging result.

One of the reasons for the loose lower bounds when t is small is that Theorem 3.3 maximizes the reward accumulated during the last renewal period, $T^{[N(t)+1]}$, by assuming that the reward is always $\max\{\rho_i\}$. In our case, this corresponds to assuming that the server is always busy during the period in AMC (for throughput bounds), or that the queue is always full (for queue length bounds).

5.2 Second approximation

We now consider the second model, with no customer loss due to failure. Our second approximation can be applied to this case, since the renewal periods are identically distributed and independent of the state (number of customers) at the time of a failure.

The results are shown in Fig. 7, for both the exponential and Erlang case, and for values of t up to 55 (five times the expected renewal period). It is apparent how our second approximation tracks very closely the simulation results.

6 Conclusion

The stationary analysis of DSPNs has been studied in the past and, if at most one deterministic transition is enabled in any marking, efficient solution algorithms are known. However, the transient analysis of the same class of DSPNs is computationally intensive. We then opted to study the transient analysis of subclasses of DSPNs using bounding and approximation techniques instead. By not seeking “exact” results, we are able to study very large models, since, essentially, only stationary and transient analysis of CTMCs is required. More specifically, for the class of DSPNs having a single persistent deterministic transition that can become enabled only in a given state i^* :

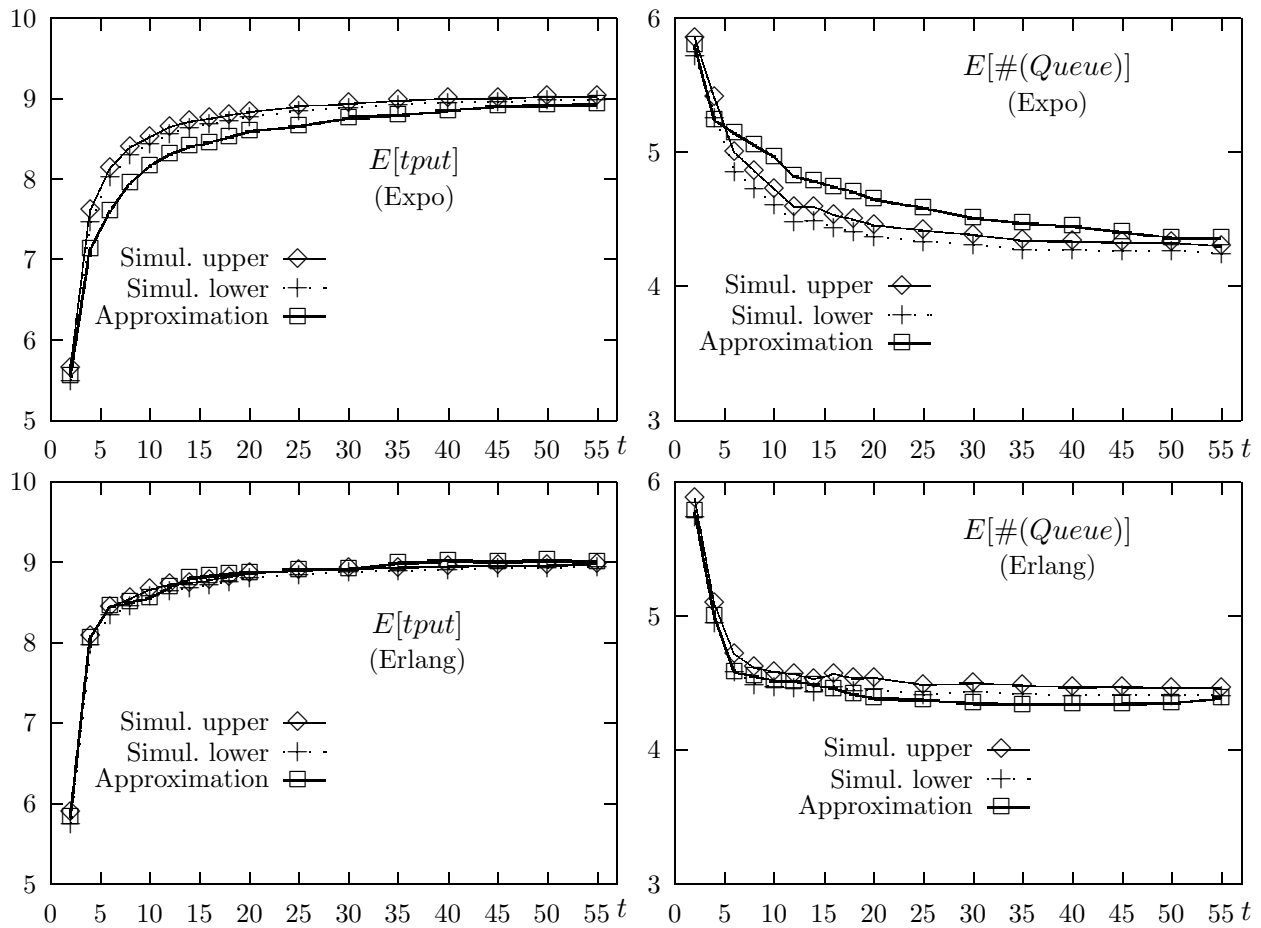


Figure 7: Numerical results for an exponential or Erlang failure process (second model).

- We proved that the model defines an underlying renewal process.
- We defined upper and lower bounds for the average sojourn time in each state and for the accumulated reward up to time t ; while our lower bound can be overly pessimistic for short transient times, one can use the bound information to gauge how close the DSPN is to steady state, since our bounds coincide in the limit.
- We defined a heuristic approximation guaranteed to fall within our bounds, which appears to be quite accurate in practice.

Then, for a more general class of DSPNs having a single persistent deterministic transition and iid renewal times:

- We defined an iterative heuristic approximation which requires multiple transient analyses of CTMCs, and appears to be quite accurate in practice as well.

In conclusion, it is our belief that the use of CTMC techniques to compute approximate or exact transient measures on a DSPN is very promising, and our contribution is a first step in this direction.

Acknowledgments

We sincerely thanks the anonymous referees whose careful reading and helpful suggestions allowed us to improve this paper. We also acknowledge the National Science Foundation which, through the contract 96-SC-NSF-1011 from the Center for Advanced Computing and Communication, partially supported this work.

References

- [1] 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.
- [2] E. Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, 1975.
- [3] 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.
- [4] H. Choi, V. G. Kulkarni, and K. S. Trivedi. Markov regenerative stochastic Petri nets. *Perf. Eval.*, 20(1-3):337–357, 1994.
- [5] 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.
- [6] G. Ciardo, R. German, and C. Lindemann. A characterization of the stochastic process underlying a stochastic Petri net. *IEEE Trans. Softw. Eng.*, 20(7):506–515, July 1994.
- [7] G. Ciardo and C. Lindemann. Analysis of deterministic and stochastic Petri nets. In *Proc. 5th Int. Workshop on Petri Nets and Performance Models (PNPM'93)*, pages 160–169, Toulouse, France, Oct. 1993. IEEE Comp. Soc. Press.
- [8] R. German. New results for the analysis of deterministic and stochastic Petri nets. In *Proc. IEEE International Computer Performance and Dependability Symposium (IPDS'95)*, pages 114–123, Erlangen, Germany, 1995. IEEE Comp. Soc. Press.

- [9] R. German, D. Logothetis, and K. S. Trivedi. Transient analysis of Markov regenerative stochastic Petri nets: a comparison of approaches. In *Proc. 6th Int. Workshop on Petri Nets and Performance Models (PNPM'95)*, pages 103–112, Durham, NC, Oct. 1995. IEEE Comp. Soc. Press.
- [10] A. Heindl and R. German. A fourth-order algorithm with automatic stepsize control for the transient analysis of DSPNs. In *Proc. 7th Int. Workshop on Petri Nets and Performance Models (PNPM'97)*, pages 60–69, St. Malo, France, June 1997. IEEE Comp. Soc. Press.
- [11] G. Li. A Stochastic Petri Nets Simulator Engine, 1998. MS thesis, Department of Computer Science, College of William and Mary.
- [12] C. Lindemann. An improved numerical algorithm for calculating steady-state solutions of Deterministic and Stochastic Petri Net models. *Perf. Eval.*, 8, 1993.
- [13] J. K. Muppala and K. S. Trivedi. Numerical transient analysis of finite Markovian queueing systems. In I. V. Basava and U. N. Bhat, editors, *Queueing and Related Models*, pages 262–284. Oxford University Press, Oxford, UK, 1992.
- [14] T. Murata. Petri Nets: properties, analysis and applications. *Proc. of the IEEE*, 77(4):541–579, Apr. 1989.
- [15] R. Nelson. *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag, 1995.
- [16] A. L. Reibman and K. S. Trivedi. Numerical transient analysis of Markov models. *Computers and Operations Research*, 15(1):19–36, 1988.
- [17] S. M. Ross. Approximations in renewal theory. *Probability in the Engineering and Informational Sciences*, (1):163–174, 1987.
- [18] S. M. Ross. *Introduction to probability models*. Academic Press, Sand Diego, CA, 1993.