Efficient approximate transient analysis for a class 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. We present a new method to compute bounds and an approximation on the average state sojourn times for a special class of deterministic and stochastic Petri nets (DSPNs). In addition to the idea of the subordinated Markov chain traditionally used for the stationary solution of DSPNs, our algorithm makes use of concepts from renewal theory. An application to a finite-capacity queue with a server subject to breakdowns is included.

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 [7, 8, 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 [13].

However, the exact transient study of DSPNs is much more difficult. In [4, 5], 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 [9], and further improved through the use of automatic stepsize control [11]. In [10], the two approaches are compared in terms of memory requirements, efficiency, and accuracy.

In this paper, we present instead 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 s, and d is "persistent", that is, once it becomes enabled, it can become disabled only because of its own firing.

Our idea is to treat this stochastic process as a renewal process. By computing the expected number of renewal cycles up to time t and the average sojourn time in each state during a renewal cycle, we can then compute bounds on the average total sojourn time in each state.

The paper is organized as follows. Section 2 reviews renewal theory and DSPN terminology. Section 3, provides the theoretical results required for our approach. Section 4 describes the step-by-step computational method we propose. A complete example is illustrated in Section 5. Finally, concluding remarks are listed in Section 6.

2. Background

We briefly review the essential concepts of renewal theory and DSPNs. For further information, the reader can consult [3, 15] for the former, and [2, 5, 8] for the latter.

2.1. Renewal theory

Definition 2.1 If the sequence of nonnegative random variables $\{T_1, T_2, \ldots\}$ is independent and identically distributed, the counting process $\{N(t) = \max\{n \in \mathbb{N} : T_1 + \cdots + T_n \leq t\} : t \geq 0\}$ 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, independently 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,

$$S_0 = 0, \quad \text{and} \quad \forall n \geq 1, \; S_n = \sum_{k=1}^{n} T_k$$

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

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n+1\} \\ &= P\{S_n \leq t\} - P\{S_{n+1} \leq t\}. \end{aligned}$$

Since the random variables $\{T_1, T_2, \ldots\}$ are independent and have a common distribution F, S_n is distributed as 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 $m(t)$:

$$\begin{aligned} m(t) &= E[N(t)] = \sum_{n=1}^{\infty} P\{N(t) \geq n\} \\ &= \sum_{n=1}^{\infty} P\{S_n \leq t\} = \sum_{n=1}^{\infty} F_n(t). \qquad \square \end{aligned}$$

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 it at the instants where a deterministic transition becomes enabled or upon entering any state where no deterministic transition is enabled.

For each deterministic transition d and each marking s where d becomes enabled, a continuous-time Markov chain (CTMC) is defined, to model the evolution of the DSPN while d is enabled. Each such "subordinated Markov chain" (SMC) $\{\hat{X}(t) : t \geq 0\}$, with state space $\hat{\mathcal{S}}$, is studied in the transient at time τ, the firing time of transition d. For every state $i \in \hat{\mathcal{S}}$, we compute $\hat{\pi}_i(\tau)$, the probability of being in i at time τ, and $\hat{\sigma}_i(\tau)$, the amount of time spent in i up to time τ. Note that the firing of another transition can disable d; the SMC reaches an absorbing state in this case.

Then, a discrete time Markov chain (DTMC), the so-called "embedded Markov chain" (EMC) with state space $\tilde{\mathcal{S}}$ is defined. For states where no deterministic transition is enabled, this is the classical embedding of a CTMC into a DTMC: if state i goes to state j with rate $\lambda_{i,j}$ in the CTMC, state i goes to state j with probability $\lambda_{i,j} \cdot h_i$ in the EMC, where the expected holding time h_i is simply the inverse of the sum of the rates leaving state i toward any other state. For a state s enabling a deterministic transition d, instead, the EMC contains a transition from state s to state j with probability $\sum_{i \in \hat{\mathcal{S}}} \hat{\pi}_i(\tau) \cdot \mathbf{P}_{i,j}$, where matrix $\mathbf{P} \in \mathbb{R}^{\hat{\mathcal{S}} \times \tilde{\mathcal{S}}}$, whose rows are probability vectors, translates the SMC state i into the state j of the EMC. If i is a state reached by disabling d, then j is the same as i, that is, $\mathbf{P}_{i,j} = 1$ iff $i = j$; otherwise, $\mathbf{P}_{i,j}$ is the probability that the DSPN reaches state j by firing d in state i. The expected holding time h_s in state s for the EMC is defined as $\sum_{\mathcal{E}(d,i)} \hat{\sigma}_i(\tau)$, where the predicate $\mathcal{E}(d, i)$ is true iff d is enabled in the state $i \in \hat{\mathcal{S}}$ of the SMC. Thus, the state space $\tilde{\mathcal{S}}$ of the EMC contains only states where a deterministic transition becomes enabled and states where no deterministic transition is enabled. The limiting behavior of this EMC is then obtained by computing, for all $i \in \tilde{\mathcal{S}}$, the stationary probability, p_i, if the EMC is ergodic, or the expected number of visits to each transient state until absorption, n_i, otherwise.

Finally, the analogous limiting quantities for the original

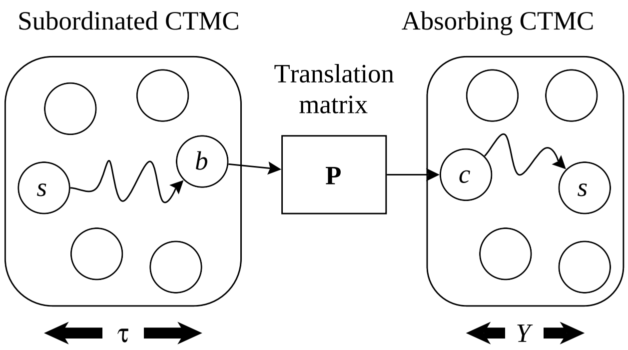

Figure 1. A depiction of our class of MRGPs.

process are obtained as:

$$\lim_{t\to\infty}\pi_i(t) = \frac{p_i \cdot h_i}{\sum_{j\in\tilde{\mathcal{S}}} p_j \cdot h_j} \quad \text{and} \quad \lim_{t\to\infty}\sigma_i(t) = n_i \cdot h_i.$$

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. Our approach

We now 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 a given marking s. 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 s 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 $b \in \hat{\mathcal{S}}$ be the state immediately before the firing.
4. The firing of d in b causes the DSPN to reach state $c \in \tilde{\mathcal{S}}$ with probability $\mathbf{P}_{b,c}$.
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 s, which is considered absorbing for the AMC. Let Y be the time to reach state s starting from state c (this random variable depends on c of course).

$T = \tau + Y$ 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 $\hat{\mathcal{S}} \cap \tilde{\mathcal{S}} = \{s\}$; $\tilde{\pi}_i(t)$ is the probability of state $i \in \tilde{\mathcal{S}}$ and $\tilde{\sigma}_i(t)$ is the expected sojourn time in state $i \in \tilde{\mathcal{S}} \setminus \{s\}$ 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, possibly with mass at zero (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 s and is persistent. Then, the number of visits to state s up to time t constitutes a renewal process.

Proof: Assume that the DSPN just entered state s at time 0, that is, $X(0) = s$ (if this is not the case, we obtain a delayed renewal process). Then, if we let $S_0 = 0$ and S_n be the time at which state s 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 $\{N(t) = max\{n \in \mathbb{N} : S_n = T_1 + \cdots + T_n \leq t\} : t \geq 0\}$ is a renewal process. □

Definition 3.1 [16] A random variable $K \in \mathbb{N}$ is a stopping time for the sequence of independent random variables $\{A_1, A_2, \ldots\}$ iff the event $\{K = n\}$, for any $n \in \mathbb{N}$, is independent of $\{A_{n+1}, A_{n+2}, \ldots\}$. □

Lemma 3.1 $N(t) + 1$ is a stopping time for the sequence $\{T_1, T_2, \ldots\}$.

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 $\{T_1, T_2, \ldots, T_n\}$ and $\{T_{n+1}, T_{n+2}, \ldots\}$. $\Box$

Lemma 3.2 For $i \in \mathcal{S}$, let T_r^i be the sojourn time in state i during the r^{th} renewal cycle. Then $N(t)+1$ is a stopping time for the sequence $\{T_1^i, T_2^i, \ldots\}$.

Proof: For any $n > 0$, the random variables $\{T_{n+1}^i, T_{n+1}^i, \ldots\}$ are independent of $\{T_1, \ldots, T_n\}$, so $N(t)+1 = n$ is independent of $\{T_{n+1}^i, T_{n+2}^i, \ldots\}$. $\Box$

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

Proof: Since $N(t)+1$ is a stopping time for the sequence $\{T_1^i, T_2^i, \ldots\}$, we can apply Wald's equation [16] and obtain

$$E\left[\sum_{n=1}^{N(t)+1} T_n^i\right] = E[T^i] \cdot E[N(t)+1] = E[T^i] \cdot (m(t)+1). \;\Box$$

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

Theorem 3.2 The average sojourn time up to time t in state i is bounded by: if $i \in \hat{\mathcal{S}}$, $E[T^i] \cdot (m(t)+1) - \tau \le \sigma_i(t) \le E[T^i] \cdot (m(t)+1)$; if $i \in \tilde{\mathcal{S}} \setminus \{s\}$, $E[T^i] \cdot (m(t)+1) - (E[T_{N(t)+1}] - \tau) \le \sigma_i(t) \le E[T^i] \cdot (m(t)+1)$.

Proof: We have

$$E\left[\sum_{n=1}^{N(t)} T_n^i\right] \le \sigma_j(t) < E\left[\sum_{n=1}^{N(t)+1} T_n^i\right].$$

By Lemma 3.3, we also have

$$E\left[\sum_{n=1}^{N(t)+1} T_n^i\right] = E[T^i] \cdot (m(t)+1),$$

which proves the upper bound. For the lower bound, observe that

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

and that $T_{N(t)+1}^i$ is at most τ for $i \in \hat{\mathcal{S}}$, and at most $T_{N(t)+1} - \tau$ for $i \in \tilde{\mathcal{S}} \setminus \{s\}$. $\Box$

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 a weighted sum: $r(t) = \sum_{i \in \mathcal{S}} \sigma_i(t) \cdot \rho_i$. Then, the following bounds on r 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: $\left(\sum_{i \in \mathcal{S}} E[T^i] \cdot \rho_i\right) \cdot (m(t)+1) - \tau \cdot \max_{i \in \hat{\mathcal{S}}}\{\rho_i\} - (E[T_{N(t)+1}] - \tau) \cdot \max_{i \in \tilde{\mathcal{S}} \setminus \{s\}}\{\rho_i\} \le r(t) \le \left(\sum_{i \in \mathcal{S}} E[T^i] \cdot \rho_i\right) \cdot (m(t)+1)$.

Proof: For the upper bound, consider that

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

For the lower bound, we have

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

And, by the same argument as in Theorem 3.2,

$$\sum_{i \in \hat{\mathcal{S}}} E[T_{N(t)+1}^i] \cdot \rho_i \le \tau \cdot \max_{i \in \hat{\mathcal{S}}}\{\rho_i\} \quad \text{and}$$

$$\sum_{i \in \tilde{\mathcal{S}} \setminus \{s\}} E[T_{N(t)+1}^i] \cdot \rho_i \le (E[T_{N(t)+1}] - \tau) \cdot \max_{i \in \tilde{\mathcal{S}} \setminus \{s\}}\{\rho_i\}. \;\Box$$

Theorems 3.2 and 3.3 tell us that if we can compute the average sojourn time during a generic renewal cycle 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 $i \in \tilde{\mathcal{S}} \setminus \{s\}$, $E[T^i]$ is simply $\tilde{\sigma}(\infty)_i = \lim_{t \to \infty} \tilde{\sigma}_i(t)$, a quantity that can be easily computed by solving the nonhomogenous 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 s removed). See [6] 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 $F_n(t) = P\{T_1 + \ldots + T_n \leq t\}$ requires the computation of an n-dimensional integral. Ross [15] proposed an efficient algorithm which requires as inputs only one-dimensional integrals.

Theorem 3.4 (Ross [15]) For a renewal process where the renewal interval has pdf $f(x)$, define

$$m_r = \int_0^\infty \sum_{k=0}^{r-1} (1 + m_{r-k}) \frac{e^{-\lambda x}(\lambda x)^k}{k!} f(x) dx,$$

where $r = 1, 2, \ldots, n, \lambda = n/t$. Then, if $m(t)$ is continuous at t, m_n converge to $m(t)$ as n goes to infinity. $\square$

This theorem tells us that if we know the distribution $F(x)$ 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 $t \to \infty$). 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 truncated at t, $(1 - e^{-\lambda t})/\lambda$, and $E[T_{N(t)+1}] = (2 - e^{-\lambda t})/\lambda$, less than $2/\lambda$, but still larger than the average length of a typical renewal interval, $1/\lambda$.

Taking into account the fact that the age is no greater than t is computationally difficult unless the distribution F is analytically known and tractable. However, we can always ignore this restriction and compute an upper bound on $E[T_{N(t)+1}]$ by using the relation [14, p. 269]

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

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 $\tilde{\mathcal{S}} \setminus \{s\}$, $\sigma_i(t)$ 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 $i \in \tilde{\mathcal{S}} \setminus \{s\}$ is analogous), exactly one of these three events must occur:

- e_1: In the last renewal cycle (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 cycle, the DSPN visits state i, but, at time t, it is not in state i.

Conditioning on these mutually exclusive events, we have $E[T^i_{N(t)+1}] = E[T^i_{N(t)+1}|e_1] \cdot \Pr\{e_1\} + E[T^i_{N(t)+1}|e_2] \cdot \Pr\{e_2\} + E[T^i_{N(t)+1}|e_3] \cdot \Pr\{e_3\} = 0 \cdot \Pr\{e_1\} + 2E[T^i] \cdot \Pr\{e_2\} + E[T^i] \cdot \Pr\{e_3\} \leq 2E[T^i]$. By the same argument as in Theorem 3.2, we obtain the result. $\square$

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, $A(t) = t - \sum_{n=1}^{N(t)} T_n$, 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,

$$\begin{aligned}
\sigma_i(t) &= E[T_1^i + T_2^i + \cdots + T^i_{N(t)} + A^i(t)] \\
&= \sum_{n=0}^{N(t)} E[T_1^i + \cdots + T^i_{N(t)} | N(t) = n] \cdot \Pr\{N(t) = n\} \\
&\quad + \sum_{n=0}^{N(t)} E[A^i(t) | N(t) = n] \cdot \Pr\{N(t) = n\}
\end{aligned}$$

We can approximate the first summation with:

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

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

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

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 s, 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 $i \in \hat{\mathcal{S}}$,

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

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

$$\tilde{\sigma}_i(\infty) \cdot (m(t)+1) - (E[T_{N(t)+1}] - \tau)$$
$$\le m(t) \cdot \tilde{\sigma}_i(\infty) + 0 \le$$
$$m(t) \cdot \tilde{\sigma}_i(\infty) + \tilde{\sigma}_i(a-\tau) \le \tilde{\sigma}_i(\infty) \cdot (m(t)+1).$$

4. Computational algorithm

Our approach hinges on the ability to obtain the distribution $F(x)$ of the renewal interval T. Since the length of this interval is $T = \tau + Y$, where τ is a constant, we know that $F(x) = 0$ for $x < \tau$. For $x \ge \tau$, instead, $F(x)$ is the probability that the AMC is absorbed in state s by time $x - \tau$, given that it is started with the initial probability distribution $\tilde{\pi}(0)$ previously defined. Thus, computing $F(x)$ numerically only requires us to perform a transient analysis on the AMC to compute the probability of being in state s at times $0, t_1, t_2, \ldots, t_{MAX}$, appropriately chosen so that this discretization of $F(x)$ is a good approximation of its continuous behavior (in particular $F(t_{MAX})$ should be very close to one). The detailed steps required for our bounding approach are then:

1. From the model, compute the reachability graph $(\mathcal{R}, \mathcal{A})$. If there is only one deterministic transition d, with duration τ, d becomes enabled only upon entering a unique state s, and d is persistent, continue; otherwise stop.
2. Define the SMC and the AMC from the model, and compute their state spaces $\hat{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ and infinitesimal generators $\hat{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$.
3. Perform an instantaneous and cumulative transient analysis the SMC: starting from the initial state s, 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 $[0, \tau]$, $\hat{\sigma}_i(\tau)$.
4. 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, $\tilde{\pi}_j(0) = \sum_{i \in \hat{\mathcal{S}}} \hat{\pi}_i(\tau) \cdot \mathbf{P}_{i,j}$.
5. Compute the probability $\Pr\{Y \le y\}$ that the AMC in state s at times $y = 0, t_1, t_2, \ldots, t_{MAX}$, then define a discretization of F using the values
$$F(x) = \begin{cases} 0 & \text{if } x < \tau, \\ \Pr\{Y \le y\} & \text{if } x = \tau + y, \text{ for} \\ & y \in \{0, t_1, \ldots, t_{MAX}\}. \end{cases}$$
6. Compute the expected sojourn time $\tilde{\sigma}_i(\infty)$ in each transient state $i \in \tilde{\mathcal{S}} \setminus \{s\}$ of the AMC until absorption, starting from the initial distribution $\tilde{\pi}(0)$, by solving the linear system $\tilde{\sigma}(\infty) \cdot \tilde{\mathbf{Q}}^* = -\tilde{\pi}(0)^*$.
7. For each time t of interest, compute an approximation of the average number of renewal cycles $m(t)$ using Ross's method.
8. Compute $E[T_{N(t)+1}] = E[T^2]/E[T]$. While $E[T^2]$ can be obtained from the discretization of $F(x)$, $E[T]$ is even simpler: $E[T] = \tau + \sum_{i \in \tilde{\mathcal{S}} \setminus \{s\}} \tilde{\sigma}_i(\infty)$.
9. 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 $a = t - m(t) \cdot E[T]$.

5. Example

We now present an application of our techniques. Consider the DSPN shown in Fig. 2, modeling a finite-capacity queue where the server is subject to breakdowns. We use the notation $\#(p, i)$ to indicate the number of tokens in place p when the marking is i and $\lambda(t, i)$ to indicate the firing rate for transition t when the marking is i (the marking is omitted if the "current state" is intended).

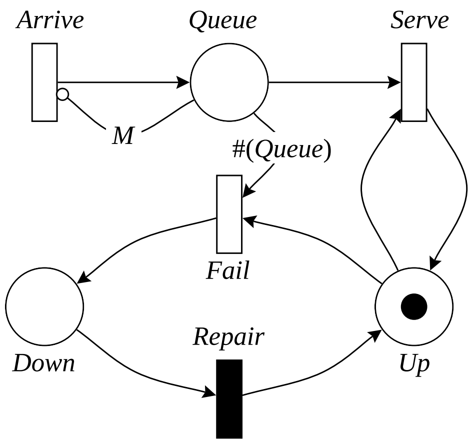

Figure 2. An example of our class of DSPNs.

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 $Fail$, whose firing flushes the contents of place $Queue$ (the input arc from $Queue$ to $Fail$ has a "variable cardinality". The repair of the server is the only activity with a deterministic duration. The conditions for applying our results 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 $Fail$. 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.

- Transition $Serve$ can have an arbitrary PH distribution, again reset by the firing of transition $Fail$. This is reasonable, since it 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, hence this allows 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 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, since the SMC would consist of the single absorbing marking $\{Down\}$.

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

- The expected customer throughput:

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

- 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. 3 shows the corresponding underlying process obtained 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. 4, where the component "P" of the state indicates the phase of the Erlang distribution.

We observe that, in our example, the distribution $F(x)$ of the renewal time 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.

Hence, we can compute analytically the value of $E[T_{N(t)+1}]$. If we ignore that the age of the renewal interval is bounded by t, we can simply assume the worst-case scenario (longest $T_{N(t)+1}$) that the DSPN is in the AMC at time t:

$$\lim_{t\to\infty} E[T_{N(t)+1}] \leq \lim_{t\to\infty} E[T_{N(t)+1} | X(t) \in \tilde{\mathcal{S}} \setminus \{s\}],$$

which results in

$$\lim_{t\to\infty} E[T_{N(t)+1}] \leq \tau + \frac{2}{\phi}$$

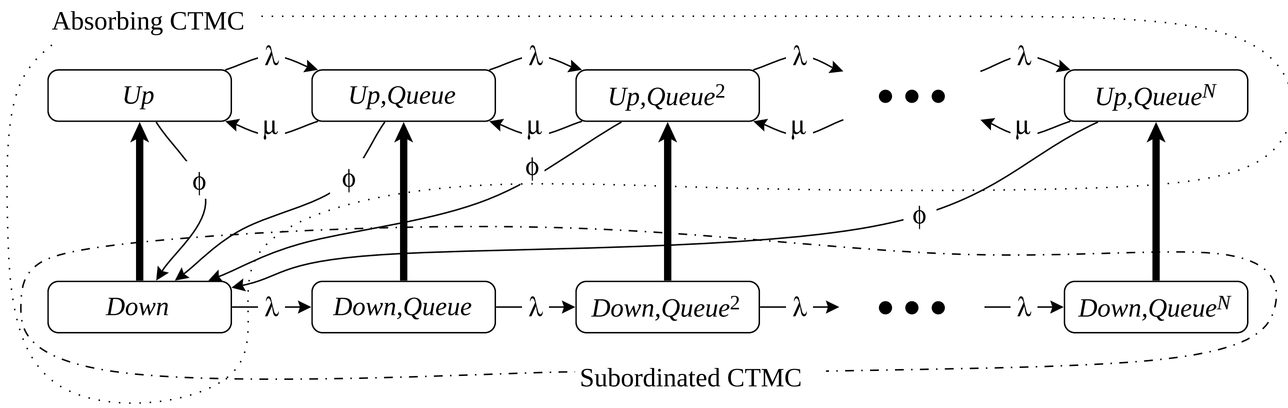

Figure 3. The process underlying our DSPN (exponential failure process).

if the failure process is exponential and

$$\lim_{t\to\infty} E[T_{N(t)+1}] \le \tau + \frac{K+1}{\phi}$$

if it is Erlang. To compute the actual value of $E[T_{N(t)+1}]$ for a finite t, we need to take into account the fact that the renewal interval is bounded by t. Again by assuming the worst-case scenario (which, for the case of an Erlang failure process, assumes that the DSPN is in the first phase of the Erlang distribution at time t, not just in any state of the AMC), we obtain

$$E[T_{N(t)+1}] \le \tau + \frac{2 - e^{-\phi(t-\tau)}}{\phi}$$

if the failure process is exponential while, if the failure process is Erlang, we obtain

$$E[T_{N(t)+1}] \le \tau + \frac{K + 1 - e^{-\phi(t-\tau)}}{\phi}.$$

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

We used these bounds on $E[T_{N(t)+1}]$ to compute our results, reported in Fig. 5. 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, and the value obtained using our heuristic approximation. 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 cycle, $T_{N(t)+1}$, 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). It should be possible to obtain tighter bounds by explicitly considering the evolution of the reward values during the last renewal cycle, for example by computing the maximum of the accumulated rewards conditioning on i being the state of the DSPN at time t, for each $i \in \mathcal{S}$. However, this is computationally prohibitive for realistic systems, as it requires to solve the model $O(|\mathcal{S}|)$ times.

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 a subclass of DSPNs using bounding and approximation techniques instead. By not seeking “exact” results, we are able to study very large models, since, es-

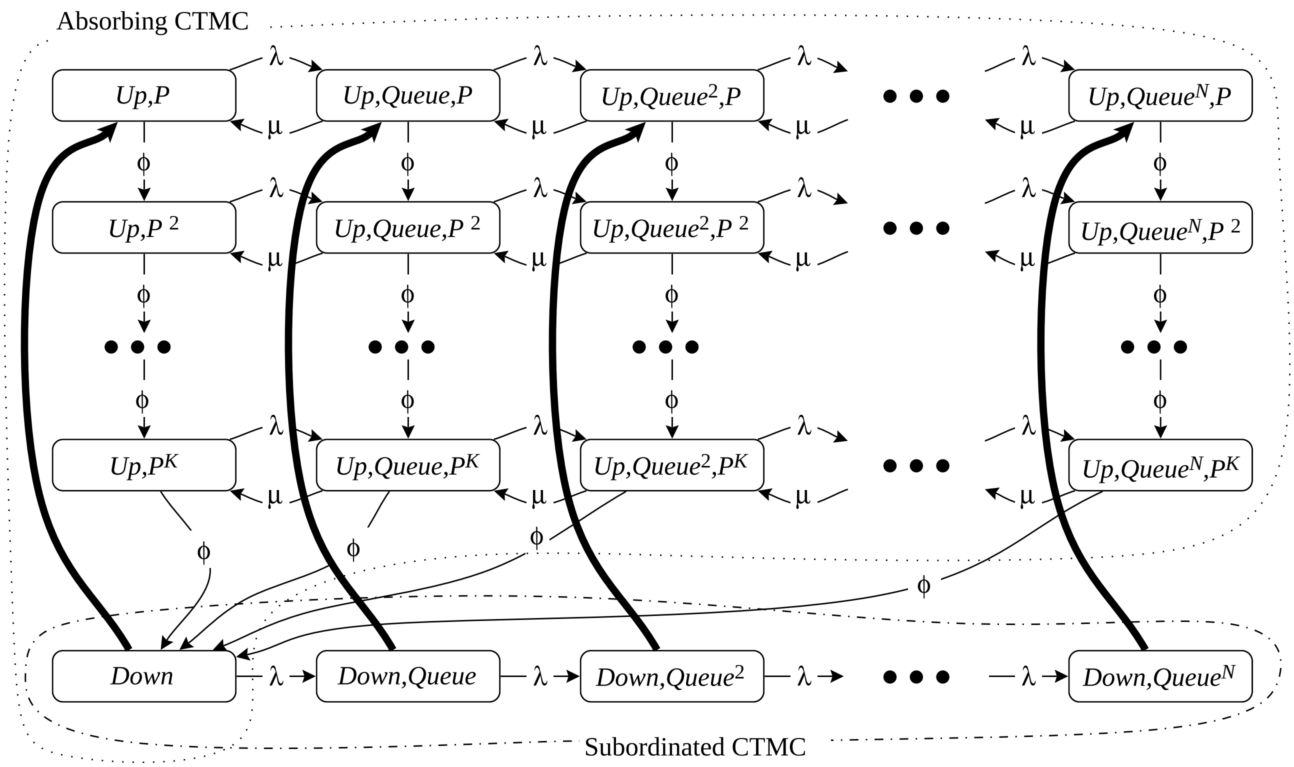

Figure 4. The process underlying our DSPN (Erlang failure process).

sentially, only stationary and transient analysis of CTMCs is required. More specifically:

1. We proved that the subclass of DSPNs we considered defines a renewal process.

2. 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.

3. We defined a heuristic approximation guaranteed to fall within our bounds, which appears to be quite accurate in practice.

We believe that the idea of using CTMC techniques to compute approximate or exact transient measures on a DSPN is very promising, and that our contribution is just a step in this direction.

References

[1] M. Ajmone Marsan and G. Chiola. On Petri nets with deterministic and exponentially distributed firing times. In *Proc. 7th European Workshop on Application and Theory of Petri Nets*, Oxford, England, June 1986.

[2] 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.

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

[4] 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.

[5] 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.

[6] G. Ciardo, A. Blakemore, P. F. J. Chimento, J. K. Muppala, and K. S. Trivedi. Automated generation

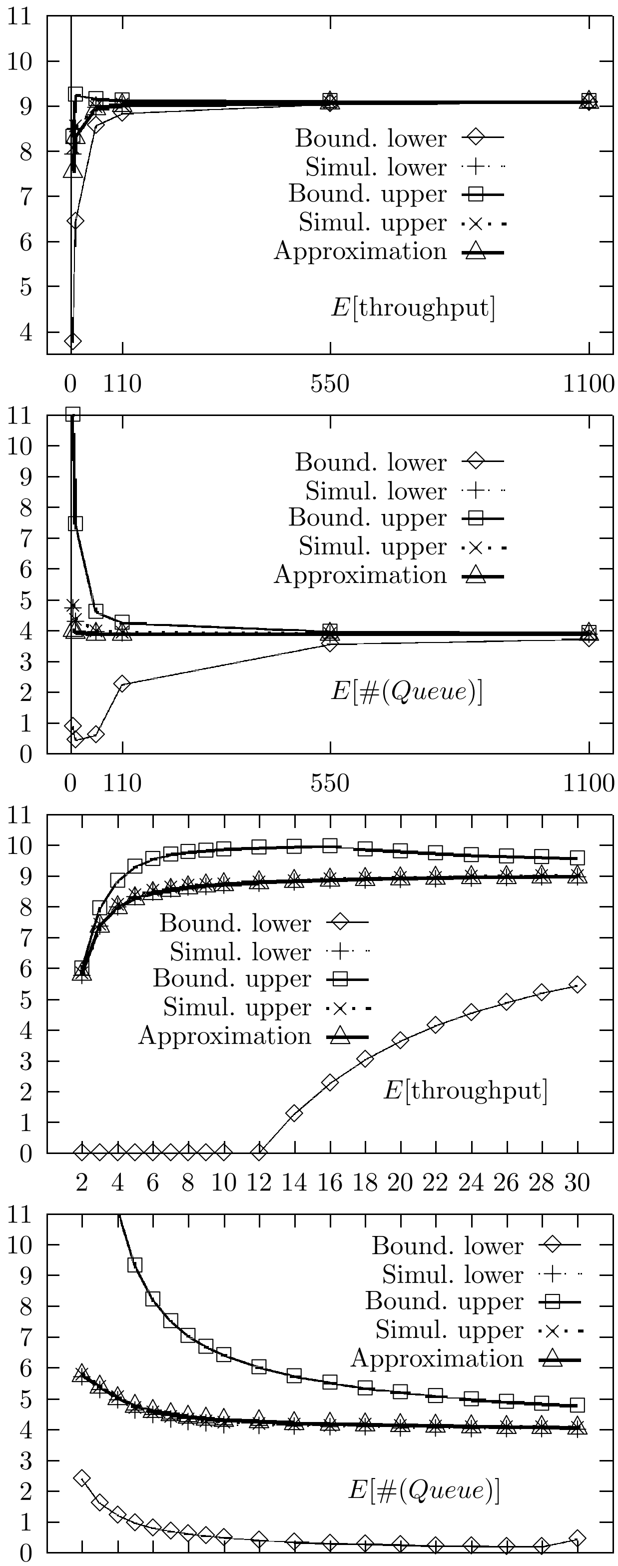

Figure 5. Numerical results for an exponential (top) or Erlang (bottom) failure process.

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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] A. Heindl and R. German. A fourth-order algortithm 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, Saint Malo, France, June 1997. IEEE Comp. Soc. Press.

[12] C. Lindemann. An improved numerical algorithm for calculating steady-state solutions of Deterministic and Stochastic Petri Net models. *Perf. Eval.*, 8, 1993.

[13] T. Murata. Petri Nets: properties, analysis and applications. *Proc. of the IEEE*, 77(4):541–579, Apr. 1989.

[14] R. Nelson. *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag, 1995.

[15] S. M. Ross. Approximations in renewal theory. *Probability in the Engineering and Informational Sciences*, (1):163–174, 1987.

[16] S. M. Ross. *Introduction to probability models*. Academic Press, Sand Diego, CA, 1993.