Performance Analysis of Stochastic Timed Petri Nets using Linear

Zhen Liu

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Performance Analysis of Stochastic Timed Petri Nets using Linear Programming Approach

Zhen Liu

N° 2642

Août 1995

PROGRAMME 1

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Performance Analysis of Stochastic Timed Petri Nets using Linear Programming Approach*

Zhen Liu

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués

Projet MISTRAL

Rapport de recherche n°2642 — Août 1995 — 36 pages

Abstract: Stochastic timed Petri nets are a useful tool in performance analysis of concurrent systems such as parallel computers, communication networks and flexible manufacturing systems. In general, performance measures of stochastic timed Petri nets are difficult to obtain for problems of practical sizes. In this paper, we provide a method to compute efficiently upper and lower bounds for the throughputs and mean token numbers in general Markovian timed Petri nets. Our approach is based on uniformization technique and linear programming.

Key-words: Stochastic timed Petri net, performance bound, throughput, mean token number, uniformization, linear programming.

(Résumé : tsvp)

*__Correspondence__: Zhen LIU, INRIA, Centre Sophia Antipolis, 2004 route des Lucioles, B.P. 93, 06902 Sophia-Antipolis, France. e-mail: liu@sophia.inria.fr

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)

Analyse de Performance de Réseaux de Petri Stochastiques par la Programmation Linéaire

Résumé : Réseaux de Petri stochastiques deviennent un outil important dans l'analyse de performances des systèmes avec concurrences tels que calculateurs parallèles, réseaux de communication et systèmes de production flexibles. En général, les mesures de performances des réseaux de Petri stochastiques sont difficiles à obtenir pour des problèmes pratiques à cause de leurs tailles. Une nouvelle technique est présentée dans cet article pour le calcul efficace des bornes inférieures et supérieures des débits et des nombres moyens de jetons dans les réseaux de Petri markoviens généraux. L'approche proposée est basée sur la technique d'uniformisation et la programmation linéaire.

Mots-clé : Réseaux de Petri stochastiques, bornes de mesures de performances, débit, nombre moyen de jetons, uniformisation, programmation linéaire.

1 Introduction

Stochastic Timed Petri Nets (STPN) are Petri nets where transitions have firing delays. Since the last decade, they have been receiving increasing interest in the modeling and performance analysis of discrete event systems. Such a tool is particularly useful for modeling systems which exhibit concurrent, asynchronous or nondeterministic behaviors, such as parallel and distributed systems, communication networks and flexible manufacturing systems. The reader is referred to the extensive survey of [36] on theoretical analyses and applications of Petri nets. Applications to the performance evaluation of parallel and distributed machines (hardware components) and parallel and distributed computations (software components) can also be found in [3] and the special issue of *J. of Parallel and Distributed Computing* (Vol. 15, No. 3, July 1992).

Most literature of STPN is on Stochastic Petri Nets (SPN) [29, 35], where transition firing times are mutually independent exponentially distributed random variables, and their extensions: Generalized Stochastic Petri Nets (GSPN) [2] where immediate transitions (i.e. those without firing delay) are allowed, and Extended Stochastic Petri Nets (ESPN) [28] where transitions are allowed to generate random numbers of tokens upon firings. Numerical analysis of such nets is based on the analysis of the embedded Markov chains. Decomposition techniques are proposed, see e.g. [19, 34] and references therein. Analytical solutions exist in product-form for equilibrium distributions for special cases of SPN, see [15] and references their in.

There also exist analyses of stochastic timed Petri nets without Markovian assumptions. Most of them provide performance bounds, see [10, 11, 17, 18, 25]. Others analyze stability conditions [4, 9]. The reader is referred to [5] for a survey on recent results on quantitative analysis of STPN, including approximations and simulations.

Although there exist various quantitative analysis techniques and some software tools (e.g. GreatSPN [23] and SPNP [27]) for STPN, the applications of STPN are most often limited to small size problems. This is mostly due to the time and space complexity of numerical analysis algorithms and of simulations.

In this paper, we provide a new method to compute efficiently upper and lower bounds for linear functions of the throughputs and mean token numbers in general Markovian Petri nets. Our approach is based on uniformization technique and linear programming. The STPN models under consideration are closely related to GSPN models defined in [24], with in addition the possibility of randomly generating tokens upon transition firings.

Uniformization technique is one of the most useful techniques for analyzing continuous time Markov chains [31]. In [32], such a technique was used to establish linear equality constraints among the expectation of state variables in queueing networks. This allowed

the authors to bound the performance measures, both above and below, by solving a linear program. Similar approaches were taken to determine lower bounds on achievable performance of control policies in multiclass queueing networks [13], optimal control policies for Klimov's problem [14], and stability regions of queueing networks and scheduling policies [33]. In these studies, linear or nonlinear programming were used to obtain bounds.

The method of linear programming has already been used in operational analysis for deriving bounds in non-Markovian STPN [17, 18, 25]. Since no statistical assumptions are made on the distributions of firing times, such bounds are usually loose. Several techniques were proposed for the improvement of such bounds in special cases of Petri nets [20, 21].

In our work, we consider Markovian STPN. We show that, like in [32, 13], the Markovian assumption allows us to establish a set of linear equality constraints among the expectation of state variables in the Petri nets, such as token numbers in the places and indicator functions of whether transitions are enabled. More precisely, we analyze the evolution of state variables in steady state and write out evolution equations using the uniformization technique. Taking the quadratic forms of these equations allows us to establish the linear constraints. Exploiting further structural and probabilistic properties of the Petri nets, we obtain an augmenting set of linear equalities and inequalities, some of which are similar to those in [25]. Upper and lower bounds of performance measures are then obtained by solving the linear program.

The paper is organized as follows. In Section 2, we define the STPN models under consideration as well as the notation. In Section 3, we derive the linear equalities based on the uniformization technique. In Section 4, we establish other linear constraints based on the behavioral properties and probabilistic laws. In Section 5, we provide the summary of the linear programming formulation. In Section 6, we present applications of our technique. Finally, in Section 7, we conclude with remarks on the extensions of our results.

2 Notation

A Petri Net can be viewed as a directed graph $\mathcal{N} = (\mathcal{P} \bigcup \mathcal{T}, \mathcal{E})$, where the set of vertices is the union of the set of *places* $\mathcal{P}$ and the set of *transitions* $\mathcal{T}$. The set of arcs $\mathcal{E}$ is composed of two subsets $\mathcal{E}'$ and $\mathcal{E}''$. The arcs of $\mathcal{E}'$ are either of the form (p,t) or of the form (t,p) with $p \in \mathcal{P}$ and $t \in \mathcal{T}$. We shall denote by

- ${}^\bullet p$: the set of transitions that precede place p in $\mathcal{P}$: ${}^\bullet p = \{t \in \mathcal{T} \mid (t,p) \in \mathcal{E}'\}$;
- $p^\bullet$: the set of transitions that follow place p in $\mathcal{P}$: $p^\bullet = \{t \in \mathcal{T} \mid (p,t) \in \mathcal{E}'\}$;
- ${}^\bullet t$: the set of places that precede transition t in $\mathcal{T}$: ${}^\bullet t = \{p \in \mathcal{P} \mid (p,t) \in \mathcal{E}'\}$ and

$t^\bullet$: the set of places that follow transition t in $\mathcal{T}$: $t^\bullet = \{p \in \mathcal{P} \mid (t,p) \in \mathcal{E}'\}$.

The arcs of $\mathcal{E}''$ are *inhibitor arcs* connecting places to transitions. For any $t \in \mathcal{T}$, let ${}^\circ t$ be the set of places from which there is an inhibitor arc, and for any $p \in \mathcal{P}$, let p° be the set of transitions to which there is an inhibitor arc. Denote by $\eta_{p,t}$ the weight of the inhibitor arc from place p to transition t, $p \in {}^\circ t$.

The net $\mathcal{N}$ is *strongly connected* if there is a path from any place/transition to any place/transition.

For all $p \in \mathcal{P}$, define the following set of transitions:

$$\begin{aligned} U_p &= {}^\bullet p \bigcup p^\bullet, \\ V_p &= {}^\bullet p \bigcap p^\bullet, \\ T_{\bullet p} &= {}^\bullet p - p^\bullet, \\ T_{p^\bullet} &= p^\bullet - {}^\bullet p. \end{aligned}$$

Tokens circulate in the Petri Net. This circulation takes place when transitions are fired. When transition $t \in \mathcal{T}$ is fired, $\pi_{p,t}$ tokens are consumed at each place $p \in {}^\bullet t$, and $\sigma_{t,p}$ tokens are created at each place $p \in t^\bullet$. Variables $\pi_{p,t}$ and $\sigma_{t,p}$ are considered as the weights of the arcs of $\mathcal{E}'$.

An example of the Petri net is illustrated in Figure 1. It contains 7 places $\mathcal{P} = \{p_1, p_2, \cdots, p_7\}$ and 7 transitions $\mathcal{T} = \{t_1, t_2, \cdots, t_7\}$. Transitions t_2, t_3, t_7 are immediate transitions. Places p_1 and p_6 have initial marking 1, whereas the others have initial marking 0. There are two inhibitor arcs (p_3, t_5) and (p_4, t_5), represented by arcs ended with a circle.

When the weights of the arcs are upper bounded by 1, $\mathcal{N}$ is called an *ordinary* net, as opposed to *weighted* net.

In this paper, we will consider a more general case where the numbers of tokens created by firing completions are random variables. When transition $t \in \mathcal{T}$ is fired for the n-th time, $\sigma_{t,p}(n)$ tokens are created at each place $p \in t^\bullet$. For all $t \in \mathcal{T}$ $\{\sigma_{t,p}(n),\ p \in t^\bullet\}_{n=1}^\infty$ is assumed to be a sequence of independent and identically distributed (i.i.d.) random variables. The sequences of random variables $\{\sigma_{t_1,p}(n),\ p \in t_1^\bullet\}_{n=1}^\infty$ and $\{\sigma_{t_2,p}(n),\ p \in t_2^\bullet\}_{n=1}^\infty$ are, however, in general dependent for $t_1 \neq t_2$. Let $\sigma_{t,p}$ be the expectation of $\sigma_{t,p}(n)$.

For all $t \in \mathcal{T}$, $\sigma_{t,p_1}(n)$ and $\sigma_{t,p_2}(n)$ can be dependent if $p_1 \neq p_2$. For example, when $\sum_{p \in t^\bullet} \sigma_{t,p}(n) = 1$, transition t creates one token in one of its output places after each

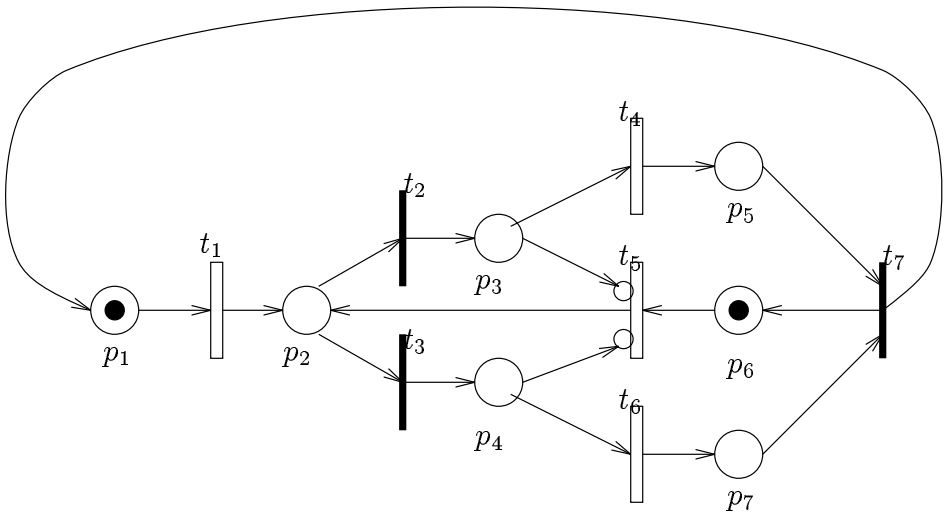

Figure 1: An example of Petri net.

firing. Two cases will be considered: *independent token generation* and *selective token generation*. In the case of independent token generation, we assume that for any $t \in \mathcal{T}$, the sequences of random variables $\{\sigma_{t,p}(n)\}_n$, $p \in t^\bullet$, are assumed to be (statistically) independent. In the case of selective token generation, however, the sequences of random variables $\{\sigma_{t,p}(n)\}_n$, $p \in t^\bullet$, are dependent in such a way that for all n, at most one of the output places has tokens created: $\sum_{p \in t^\bullet} \mathbf{1}_{\sigma_{t,p}(n)>0} \leq 1$, so that $\sigma_{t,p_1}(n)\sigma_{t,p_2}(n) = 0$ for any $p_1 \neq p_2$. A special case of selective token generation is the *routing* mechanism where a token is generated at one and only one of the output places after each firing: $\sum_{p \in t^\bullet} \sigma_{t,p}(n) = 1$ (see below discussions on immediate transitions).

There are two special classes of ordinary Petri nets, referred to as *state machines* and *marked graphs*. A state machine is an ordinary Petri net without inhibitor arcs such that for each transition t, ${}^\bullet t$ is a singleton and $\sum_{p \in t^\bullet} \sigma_{t,p}(n) = 1$, $n = 1, 2, \ldots$. A marked graph is an ordinary Petri net without inhibitor arcs such that for each place p, both ${}^\bullet p$ and $p^\bullet$ are singleton.

Firings of transitions are timed, i.e., each firing takes a certain amount of time before completion. The token consumptions in places of ${}^\bullet t$ and token creations in places of $t^\bullet$ occur simultaneously at the end of a firing of transition t. Throughout the paper we will assume that all the firing times are independent random variables. The firing times of transition $t \in \mathcal{T}$ are i.i.d. random variables of *exponential* distribution with parameter μ_t.

In GSPN framework, Petri nets can have *immediate transitions*, i.e. transitions whose firing times are zero. In this case, immediate transitions have higher firing priorities, see [24]. Using algorithms of [26], these immediate transitions can be eliminated without changing performance behavior of the net.

Of particular interests are immediate transitions which play roles of *synchronization* and/or *routing*. More precisely, in this case, we assume that for any immediate transition t, t is the only output transition of all its input places, i.e. $p^\bullet = \{t\}$ for all $p \in {}^\bullet t$. Further, we assume that for any immediate transition t, ${}^\circ t = \emptyset$, $\pi_{p,t} = 1$, $p \in {}^\bullet t$, and

- either $\sigma_{t,p'}(n) = 1$ a.s., $p' \in t^\bullet$, $n = 1, 2, \ldots$;
- or $\sigma_{t,p'}(n) \leq 1$ a.s., $p' \in t^\bullet$, $n = 1, 2, \ldots$, and for all $p \in {}^\bullet t$, $|{}^\bullet p| = 1$ and $\sigma_{{}^\bullet p,p}(n) = 1$ a.s., $n = 1, 2, \ldots$, where, with a harmless abuse of notation, the index ${}^\bullet p$ denotes the unique transition preceding place p.

In the Appendix, we present a direct transformation technique which removes this kind of immediate transitions without changing the firing behavior of the other transitions.

Thus, we will assume throughout this paper that the Petri net $\mathcal{N}$ has no immediate transition, so that all parameters μ_t are finite.

A transition t is *enabled* to fire when there are at least $\pi_{p,t}$ tokens at each place $p \in {}^\bullet t$ and there are at most $\eta_{p,t} - 1$ tokens at each place $p \in {}^\circ t$. We adopt the *single-server* semantics for the transitions. A firing can start only if the transition is enabled and the previous firing has completed. It is assumed that firings are started as soon as possible. The case of *infinite-server* semantics will be discussed in Section 7.

A firing of transition t is *preempted* when the transition is *disabled* (i.e. at least one place $p \in {}^\bullet t$ has strictly less than $\pi_{p,t}$ tokens, or at least one place $p \in {}^\circ t$ has more than or equal to $\eta_{p,t}$ tokens) before the firing time expires. The firing is *resumed* as soon as the transition becomes enabled. The disabling of a transition is due both to competitions with other transitions having common input places (some tokens in these places can be consumed by other transitions during the firing of the transition), and to token arrivals in input places of inhibitor arcs. The firing mechanism described here is called (cf. [1]) *race* policy with *age memory*. Note that for the case of exponential distributions of firing times, the race policies with or without age memory have stochastically the same performance behavior due to the memorylessness property of exponential distributions. However, in Section 7, when we consider the case where firing times have general distributions, the race policy under consideration will be that with age memory.

The state of the system is characterized by the *marking* $X(\tau) = (X_p(\tau),\ p \in \mathcal{P})$, where $X_p(\tau)$ is the number of tokens in place p at time τ. The process $X(\tau)$ is assumed to be left-continuous so that $X_p(\tau)$ is the number of tokens in place just before time τ. The *initial marking* $M = X(0)$ is the marking at time 0.

Let $\boldsymbol{\pi} = (\pi_{p,t},\ p \in \mathcal{P}, t \in \mathcal{T})$, $\boldsymbol{\sigma} = (\sigma_{t,p},\ t \in \mathcal{T}, p \in \mathcal{P})$, $\boldsymbol{\eta} = (\eta_{p,t},\ p \in \mathcal{P}, t \in \mathcal{T})$, and $\boldsymbol{\mu} = (\mu_t,\ t \in \mathcal{T})$. The Markovian Petri net described above will be denoted by $< \mathcal{N}, M, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \boldsymbol{\mu} >$.

Throughout this paper we will assume that the Petri net is *live*. Moreover, we assume that the net is *stable* in the sense that $X(\tau)$ converges to a stationary variable X (of dimension $|\mathcal{P}|$) when τ goes to infinity. Moreover, we assume that the first and second moments of X are finite, i.e. $E[X_p] < \infty$ and $E[X_p^2] < \infty$ for all $p \in \mathcal{P}$. Under these assumptions it is easy to see (using Hölder's inequality) that for all $p_1, p_2 \in \mathcal{P}$, $E[X_{p_1} X_{p_2}] < \infty$.

Let $e_t(\tau)$ be the indicator function of whether transition t is enabled at time τ (or more precisely, just before time τ):

$$e_t(\tau) = \left(\prod_{p \in {}^\bullet t} \mathbf{1}_{\{X_p(\tau) \geq \pi_{p,t}\}} \right) \cdot \left(\prod_{p \in {}^\circ t} \mathbf{1}_{\{X_p(\tau) < \eta_{p,t}\}} \right).$$

Let e_t be the stationary version of $e_t(\tau)$, and $q_t = E[e_t]$.

Denote by $x_p = E[X_p]$ the mean number of tokens in place $p \in \mathcal{P}$, and $y_{p,t} = E[X_p e_t]$, $t \in \mathcal{T}$. The corresponding vectors are denoted by $\boldsymbol{x} = (x_p, p \in \mathcal{P})$ and $\boldsymbol{y} = (y_{p,t}, p \in \mathcal{P}, t \in \mathcal{T})$. Let $\boldsymbol{q} = (q_t, t \in \mathcal{T})$. Let θ_t the (asymptotic) throughput of transition $t \in \mathcal{T}$, i.e. the number of completed firings of transition t per unit of time, and $\boldsymbol{\theta} = (\theta_t, t \in \mathcal{T})$.

In the sequel, we provide a method of computing upper and lower bounds of $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta})$ for any arbitrarily fixed linear function L. Our approach is based on linear programming. The upper (resp. lower) bound is obtained by maximizing (resp. minimizing) the objective function L under linear constraints.

3 Uniformization and Linear Equalities

We will use the uniformization technique to derive linear equalities between variables $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{q}$ and $\boldsymbol{\theta}$. We will consider the Petri net $\mathcal{N}$ where each transition $t \in \mathcal{T}$ is continuously firing with i.i.d. exponentially distributed firing times of parameter μ_t. When a firing is completed at transition $t \in \mathcal{T}$, there are two possibilities. If t is enabled, then tokens are consumed in places ${}^\bullet t$ and are created in places $t^\bullet$. Otherwise, if t is disabled when the firing is completed, nothing happens, and this firing completion corresponds to a fictive firing completion.

Let $\{\tau_n\}$ be the sequence of time epochs of, real or fictive, firing completions in $\mathcal{N}$. It is clear that $\{\tau_n\}$ is distributed according to a Poisson process with parameter $\mu = \sum_{t\in\mathcal{T}} \mu_t$. Let $\mathcal{F}_{\tau_n}$ denote the σ-field generated by the events up to time τ_n.

Let $A_t(n)$ be the indicator function such that $A_t(n) = 1$ if and only if the n-th, real or fictive, firing completion occurs at transition $t \in \mathcal{T}$. Clearly, $\sum_{t\in\mathcal{T}} A_t(n) = 1$. Moreover, for any $t \in \mathcal{T}$, $\{A_t(n)\}$ is a sequence of i.i.d. random variables, independent of $e_t(\tau_n)$, such that $P(A_t(n) = 1) = \mu_t/\mu$.

Since for any fixed $t \in \mathcal{T}$, the random variables $(\sigma_{t,p}(n),\ p \in t^\bullet)$ are i.i.d. in n, we can assume with no loss of generality that the numbers of tokens created in places $t^\bullet$ at time τ_n are $\sigma_{t,p}(n)$, $p \in t^\bullet$, provided transition t is enabled at time τ_n.

We assume without loss of generality that the system is in steady state so that, owing to PASTA (Poisson process see time average) property (cf. e.g. [6]), $(X(\tau_n), e(\tau_n))$ has the same law as (X, e).

The throughput of transition $t \in \mathcal{T}$ can be computed as follows. In the system, transitions are fired, either really or fictively, at the rate of μ. At each firing completion epoch τ_n, the firing occurs at transition $t \in \mathcal{T}$ with probability μ_t/μ. Therefore, (real or fictive) firing completions occur at transition t at the rate of μ_t. Since these firing completions are independent of e_t, we have

$$\theta_t = \mu_t q_t, \qquad \forall t \in \mathcal{T}. \tag{1}$$

The following evolution equation is essential. For all $p \in \mathcal{P}$, and $n = 0, 1, 2, \ldots$,

$$X_p(\tau_{n+1}) = \begin{cases} X_p(\tau_n), & \text{if} \quad A_t(n) = 1, \quad t \notin U_p, \\ X_p(\tau_n), & \text{if} \quad A_t(n) = 1, \quad t \in U_p, \quad e_t(\tau_n) = 0 \\ X_p(\tau_n) + \sigma_{t,p}(n), & \text{if} \quad A_t(n) = 1, \quad t \in T_{\bullet p}, \quad e_t(\tau_n) = 1 \\ X_p(\tau_n) - \pi_{p,t}, & \text{if} \quad A_t(n) = 1, \quad t \in T_{p^\bullet}, \quad e_t(\tau_n) = 1 \\ X_p(\tau_n) + \sigma_{t,p}(n) - \pi_{p,t}, & \text{if} \quad A_t(n) = 1, \quad t \in V_p, \quad e_t(\tau_n) = 1 \end{cases} \tag{2}$$

Taking the conditional expectation yields

$$\begin{aligned} & E\left[X_p(\tau_{n+1}) | \mathcal{F}_{\tau_n}\right] \\ &= \sum_{t\notin U_p} \frac{\mu_t}{\mu} X_p(\tau_n) + \sum_{t\in U_p} \frac{\mu_t}{\mu} X_p(\tau_n)(1 - e_t(\tau_n)) + \sum_{t\in T_{\bullet p}} \frac{\mu_t}{\mu} \left(X_p(\tau_n) + \sigma_{t,p}\right) e_t(\tau_n) \\ &\quad + \sum_{t\in T_{p^\bullet}} \frac{\mu_t}{\mu} \left(X_p(\tau_n) - \pi_{p,t}\right) e_t(\tau_n) + \sum_{t\in V_p} \frac{\mu_t}{\mu} \left(X_p(\tau_n) + \sigma_{t,p} - \pi_{p,t}\right) e_t(\tau_n) \\ &= X_p(\tau_n) + \sum_{t\in T_{\bullet p}} \frac{\mu_t}{\mu} \sigma_{t,p} e_t(\tau_n) - \sum_{t\in T_{p^\bullet}} \frac{\mu_t}{\mu} \pi_{p,t} e_t(\tau_n) + \sum_{t\in V_p} \frac{\mu_t}{\mu} (\sigma_{t,p} - \pi_{p,t}) e_t(\tau_n) \end{aligned}$$

Thus,

$$E\left[X_p(\tau_{n+1})|\mathcal{F}_{\tau_n}\right] = X_p(\tau_n) + \sum_{t\in {}^\bullet p} \frac{\mu_t}{\mu}\sigma_{t,p}e_t(\tau_n) - \sum_{t\in p^\bullet} \frac{\mu_t}{\mu}\pi_{p,t}e_t(\tau_n) \tag{3}$$

In the steady state, $E[X_p(\tau_{n+1})] = E[X_p(\tau_n)] = E[X_p]$, and $E[e_t(\tau_n)] = E[e_t]$, so that by taking expectation in (3), we obtain the following *flow balance* equalities:

$$\sum_{t\in {}^\bullet p} \mu_t\sigma_{t,p}q_t = \sum_{t\in p^\bullet} \mu_t\pi_{p,t}q_t, \qquad \forall p \in \mathcal{P} \tag{4}$$

Calculating the second moments from (2) yields

$$\begin{aligned}
&E\left[X_p^2(\tau_{n+1})|\mathcal{F}_{\tau_n}\right] \\
&= \sum_{t\notin U_p} \frac{\mu_t}{\mu}X_p^2(\tau_n) + \sum_{t\in U_p} \frac{\mu_t}{\mu}X_p^2(\tau_n)(1-e_t(\tau_n)) + \sum_{t\in T_{{}^\bullet p}} \frac{\mu_t}{\mu}\left(X_p(\tau_n)+\sigma_{t,p}\right)^2 e_t(\tau_n) \\
&\quad + \sum_{t\in T_{p^\bullet}} \frac{\mu_t}{\mu}\left(X_p(\tau_n)-\pi_{p,t}\right)^2 e_t(\tau_n) + \sum_{t\in V_p} \frac{\mu_t}{\mu}\left(X_p(\tau_n)+\sigma_{t,p}-\pi_{p,t}\right)^2 e_t(\tau_n) \\
&= X_p^2(\tau_n) + 2\sum_{t\in T_{{}^\bullet p}} \frac{\mu_t}{\mu}X_p(\tau_n)\sigma_{t,p}e_t(\tau_n) + \sum_{t\in T_{{}^\bullet p}} \frac{\mu_t}{\mu}\sigma_{t,p}^2 e_t(\tau_n) \\
&\quad - 2\sum_{t\in T_{p^\bullet}} \frac{\mu_t}{\mu}X_p(\tau_n)\pi_{p,t}e_t(\tau_n) + \sum_{t\in T_{p^\bullet}} \frac{\mu_t}{\mu}\pi_{p,t}^2 e_t(\tau_n) \\
&\quad + \sum_{t\in V_p} \frac{\mu_t}{\mu}X_p(\tau_n)\left(\sigma_{t,p}-\pi_{p,t}\right)e_t(\tau_n) + \sum_{t\in V_p} \frac{\mu_t}{\mu}\left(\sigma_{t,p}-\pi_{p,t}\right)^2 e_t(\tau_n)
\end{aligned}$$

Thus,

$$\begin{aligned}
&E\left[X_p^2(\tau_{n+1})|\mathcal{F}_{\tau_n}\right] \\
&= X_p^2(\tau_n) + 2\sum_{t\in {}^\bullet p} \frac{\mu_t}{\mu}\sigma_{t,p}X_p(\tau_n)e_t(\tau_n) - 2\sum_{t\in p^\bullet} \frac{\mu_t}{\mu}\pi_{p,t}X_p(\tau_n)e_t(\tau_n) \\
&\quad + \sum_{t\in {}^\bullet p} \frac{\mu_t}{\mu}\sigma_{t,p}^2 e_t(\tau_n) + \sum_{t\in p^\bullet} \frac{\mu_t}{\mu}\pi_{p,t}^2 e_t(\tau_n) - 2\sum_{t\in V_p} \frac{\mu_t}{\mu}\sigma_{t,p}\pi_{p,t}e_t(\tau_n)
\end{aligned} \tag{5}$$

In the steady state, $E[X_p^2(\tau_{n+1})] = E[X_p^2(\tau_n)] = E[X_p^2]$, $E[X_p(\tau_n)e_t(\tau_n)] = E[X_pe_t]$, and $E[e_t(\tau_n)] = E[e_t]$. Hence, by taking expectation in (5), we obtain the following *second moment* condition:

$$\begin{aligned}
&2\sum_{t\in {}^\bullet p} \mu_t\sigma_{t,p}y_{p,t} - 2\sum_{t\in p^\bullet} \mu_t\pi_{p,t}y_{p,t} \\
&= 2\sum_{t\in V_p} \mu_t\sigma_{t,p}\pi_{p,t}q_t - \sum_{t\in {}^\bullet p} \mu_t\sigma_{t,p}^2 q_t - \sum_{t\in p^\bullet} \mu_t\pi_{p,t}^2 q_t, \qquad \forall p \in \mathcal{P}.
\end{aligned} \tag{6}$$

More generally, for any $p_1, p_2 \in \mathcal{P}$, we compute the expectation of the product of numbers of tokens from (2). Assume first that token generations of all transitions $t \in \mathcal{T}$ are statistically independent i.e., random variables $\sigma_{t,p}(n)$, $p \in t^\bullet$, are independent. Then:

$$
\begin{aligned}
&E\left[X_{p_1}(\tau_{n+1})X_{p_2}(\tau_{n+1})|\mathcal{F}_{\tau_n}\right] \\
&= \sum_{t \notin U_{p_1} \bigcup U_{p_2}} \frac{\mu_t}{\mu} X_{p_1}(\tau_n) X_{p_2}(\tau_n) \\
&\quad + \sum_{t \in U_{p_1} \bigcup U_{p_2}} \frac{\mu_t}{\mu} X_{p_1}(\tau_n) X_{p_2}(\tau_n)(1 - e_t(\tau_n)) \\
&\quad + \sum_{t \in T_{\bullet_{p_1}} - U_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1}\right) X_{p_2} e_t(\tau_n) \\
&\quad + \sum_{t \in T_{p_1^\bullet} - U_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) - \pi_{p_1,t}\right) X_{p_2} e_t(\tau_n) \\
&\quad + \sum_{t \in V_{p_1} - U_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1} - \pi_{p_1,t}\right) X_{p_2} e_t(\tau_n) \\
&\quad + \sum_{t \in T_{\bullet_{p_2}} - U_{p_1}} \frac{\mu_t}{\mu} \left(X_{p_2}(\tau_n) + \sigma_{t,p_2}\right) X_{p_1} e_t(\tau_n) \\
&\quad + \sum_{t \in T_{p_2^\bullet} - U_{p_1}} \frac{\mu_t}{\mu} \left(X_{p_2}(\tau_n) - \pi_{p_2,t}\right) X_{p_1} e_t(\tau_n) \\
&\quad + \sum_{t \in V_{p_2} - U_{p_1}} \frac{\mu_t}{\mu} \left(X_{p_2}(\tau_n) + \sigma_{t,p_2} - \pi_{p_2,t}\right) X_{p_1} e_t(\tau_n) \\
&\quad + \sum_{t \in T_{\bullet_{p_1}} \bigcap T_{\bullet_{p_2}}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2}\right) e_t(\tau_n) \\
&\quad + \sum_{t \in T_{\bullet_{p_1}} \bigcap T_{p_2^\bullet}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1}\right) \left(X_{p_2}(\tau_n) - \pi_{p_2,t}\right) e_t(\tau_n) \\
&\quad + \sum_{t \in T_{\bullet_{p_1}} \bigcap V_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2} - \pi_{p_2,t}\right) e_t(\tau_n) \\
&\quad + \sum_{t \in T_{p_1^\bullet} \bigcap T_{\bullet_{p_2}}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2}\right) e_t(\tau_n) \\
&\quad + \sum_{t \in T_{p_1^\bullet} \bigcap T_{p_2^\bullet}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) - \pi_{p_2,t}\right) e_t(\tau_n) \\
&\quad + \sum_{t \in T_{p_1^\bullet} \bigcap V_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2} - \pi_{p_2,t}\right) e_t(\tau_n)
\end{aligned}
$$

$$+ \sum_{t \in V_{p_1} \bigcap T_{\bullet p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1} - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2}\right) e_t(\tau_n)$$

$$+ \sum_{t \in V_{p_1} \bigcap T_{p_2 \bullet}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1} - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) - \pi_{p_2,t}\right) e_t(\tau_n)$$

$$+ \sum_{t \in V_{p_1} \bigcap V_{p_2}} \frac{\mu_t}{\mu} \left(X_{p_1}(\tau_n) + \sigma_{t,p_1} - \pi_{p_1,t}\right) \left(X_{p_2}(\tau_n) + \sigma_{t,p_2} - \pi_{p_2,t}\right) e_t(\tau_n)$$

After some simple algebra, we obtain

$$\begin{aligned}
E\left[X_{p_1}(\tau_{n+1}) X_{p_2}(\tau_{n+1}) | \mathcal{F}_{\tau_n}\right] &= X_{p_1}(\tau_n) X_{p_2}(\tau_n) \\
&\quad + \sum_{t \in {}^\bullet p_2} \frac{\mu_t}{\mu} \sigma_{t,p_2} X_{p_1}(\tau_n) e_t(\tau_n) \\
&\quad - \sum_{t \in p_2^\bullet} \frac{\mu_t}{\mu} \pi_{p_2,t} X_{p_1}(\tau_n) e_t(\tau_n) \\
&\quad + \sum_{t \in {}^\bullet p_1} \frac{\mu_t}{\mu} \sigma_{t,p_1} X_{p_2}(\tau_n) e_t(\tau_n) \\
&\quad - \sum_{t \in p_1^\bullet} \frac{\mu_t}{\mu} \pi_{p_1,t} X_{p_2}(\tau_n) e_t(\tau_n) \\
&\quad + \sum_{t \in {}^\bullet p_1 \bigcap {}^\bullet p_2} \frac{\mu_t}{\mu} \sigma_{t,p_1} \sigma_{t,p_2} e_t(\tau_n) \\
&\quad - \sum_{t \in {}^\bullet p_1 \bigcap p_2^\bullet} \frac{\mu_t}{\mu} \sigma_{t,p_1} \pi_{p_2,t} e_t(\tau_n) \\
&\quad - \sum_{t \in p_1^\bullet \bigcap {}^\bullet p_2} \frac{\mu_t}{\mu} \pi_{p_1,t} \sigma_{t,p_2} e_t(\tau_n) \\
&\quad + \sum_{t \in p_1^\bullet \bigcap p_2^\bullet} \frac{\mu_t}{\mu} \pi_{p_1,t} \pi_{p_2,t} e_t(\tau_n) \qquad (7)
\end{aligned}$$

In the steady state,

$$E[X_{p_1}(\tau_{n+1}) X_{p_2}(\tau_{n+1})] = E[X_{p_1}(\tau_n) X_{p_2}(\tau_n)] = E[X_{p_1} X_{p_2}].$$

Thus, by taking expectation in (7) we obtain the following *population covariance* condition:

$$\begin{aligned}
&\sum_{t \in {}^\bullet p_2} \mu_t \sigma_{t,p_2} y_{p_1,t} - \sum_{t \in p_2^\bullet} \mu_t \pi_{p_2,t} y_{p_1,t} \\
&\quad + \sum_{t \in {}^\bullet p_1} \mu_t \sigma_{t,p_1} y_{p_2,t} - \sum_{t \in p_1^\bullet} \mu_t \pi_{p_1,t} y_{p_2,t}
\end{aligned}$$

$$= - \sum_{t \in {}^\bullet p_1 \bigcap {}^\bullet p_2} \mu_t \sigma_{t,p_1} \sigma_{t,p_2} q_t + \sum_{t \in {}^\bullet p_1 \bigcap p_2^\bullet} \mu_t \sigma_{t,p_1} \pi_{p_2,t} q_t$$
$$+ \sum_{t \in p_1^\bullet \bigcap {}^\bullet p_2} \mu_t \pi_{p_1,t} \sigma_{t,p_2} q_t - \sum_{t \in p_1^\bullet \bigcap p_2^\bullet} \mu_t \pi_{p_1,t} \pi_{p_2,t} q_t, \qquad \forall p_1, p_2 \in \mathcal{P}. \tag{8}$$

Note that when $p_1 = p_2$, relations (6) and (8) are identical.

Assume now that token generations of some transitions are selective, i.e. for some $t \in \mathcal{T}$, $\sum_{p \in t^\bullet} \mathbf{1}_{\sigma_{t,p}(n)>0} \leq 1$ for all $n = 1, 2, \ldots$. Let $\mathcal{T}' \subseteq \mathcal{T}$ be the subset of transitions which have selective token generations. Then, for any $t \in \mathcal{T}'$ and any $p_1 \neq p_2$, $\sigma_{t,p_1}(n)\sigma_{t,p_2}(n) = 0$. Therefore, by a similar computation we obtain:

$$\sum_{t \in {}^\bullet p_2} \mu_t \sigma_{t,p_2} y_{p_1,t} - \sum_{t \in p_2^\bullet} \mu_t \pi_{p_2,t} y_{p_1,t}$$
$$+ \sum_{t \in {}^\bullet p_1} \mu_t \sigma_{t,p_1} y_{p_2,t} - \sum_{t \in p_1^\bullet} \mu_t \pi_{p_1,t} y_{p_2,t}$$
$$= - \sum_{t \in {}^\bullet p_1 \bigcap {}^\bullet p_2 - \mathcal{T}'} \mu_t \sigma_{t,p_1} \sigma_{t,p_2} q_t + \sum_{t \in {}^\bullet p_1 \bigcap p_2^\bullet} \mu_t \sigma_{t,p_1} \pi_{p_2,t} q_t$$
$$+ \sum_{t \in p_1^\bullet \bigcap {}^\bullet p_2} \mu_t \pi_{p_1,t} \sigma_{t,p_2} q_t - \sum_{t \in p_1^\bullet \bigcap p_2^\bullet} \mu_t \pi_{p_1,t} \pi_{p_2,t} q_t, \qquad \forall p_1, p_2 \in \mathcal{P}. \tag{9}$$

Observe that equality (8) can be considered as a special case of (9). Indeed, if $\mathcal{T}' = \emptyset$, then both equalities coincide.

4 Other Constraints

In this section, we derive other linear constraints of variables $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{q}$. Except for (24), the linear constraints established in this section requires no Markovian assumption and holds for general stochastic Petri nets.

4.1 Behavioral Properties

Liveness. Since we assume that the net is live, we have that for any τ, at least one transition is enabled, so that $\sum_{t \in \mathcal{T}} e_t(\tau) \geq 1$. Thus,

$$\sum_{t \in \mathcal{T}} q_t \geq 1. \tag{10}$$

As a consequence,

$$\sum_{t\in\mathcal{T}} y_{p,t} = E\left(X_p \sum_{t\in\mathcal{T}} e_t\right) \geq EX_p = x_p,$$

so that

$$x_p \leq \sum_{t\in\mathcal{T}} y_{p,t}, \qquad \forall p \in \mathcal{P}. \tag{11}$$

Conflicting transitions. For all $t \in \mathcal{T}$, let $\boldsymbol{\pi}(t) = (\pi_{p,t}, p \in \mathcal{P})$ and $\boldsymbol{\eta}(t) = (\eta_{p,t}, p \in \mathcal{P})$, where, by convention, $\pi_{p,t} = 0$ if $p \notin {}^\bullet t$, and $\eta_{p,t} = \infty$ if $p \notin {}^\circ t$. For any pair of transitions $t_1, t_2 \in \mathcal{T}$, if $\boldsymbol{\pi}(t_1) \leq \boldsymbol{\pi}(t_2)$ and $\boldsymbol{\eta}(t_1) \geq \boldsymbol{\eta}(t_2)$ component-wise, then transition t_2 is enabled only if t_1 is enabled, so that $e_{t_1}(\tau) \geq e_{t_2}(\tau)$ for all τ. Hence,

$$q_{t_1} \geq q_{t_2}, \qquad \forall t_1, t_2 \in \mathcal{T}, \quad s.t.\ \boldsymbol{\pi}(t_1) \leq \boldsymbol{\pi}(t_2), \quad \boldsymbol{\eta}(t_1) \geq \boldsymbol{\eta}(t_2). \tag{12}$$

If transitions t_1, t_2 are in equal conflict, i.e. $\boldsymbol{\pi}(t_1) = \boldsymbol{\pi}(t_2)$ and $\boldsymbol{\eta}(t_1) = \boldsymbol{\eta}(t_2)$, then the above relation implies that $q_{t_1} = q_{t_2}$.

Boundedness. For all $p \in \mathcal{P}$, let $b_p \geq 0$ and $B_p \leq \infty$ be the minimum and maximum numbers, respectively, of tokens in place p. Then, trivially,

$$b_p \leq x_p \leq B_p, \qquad \forall p \in \mathcal{P}. \tag{13}$$

As a consequence, for any place $p \in \mathcal{P}$ such that $B_p < \infty$, and any $t \in p^\bullet$, $y_{p,t} = E[X_p e_t] \leq E[B_p e_t] = B_p q_t$, $y_{p,t} \geq E[b_p e_t] = b_p q_t$, so that

$$b_p q_t \leq y_{p,t} \leq B_p q_t, \qquad \forall p \in \mathcal{P}, \quad t \in p^\bullet. \tag{14}$$

The bounds (13) can be extended to a set of places $S \subseteq \mathcal{P}$. Let $b_S \geq 0$ and $B_S \leq \infty$ be the minimum and maximum of total numbers of tokens in places of S. Then, trivially,

$$b_S \leq \sum_{p\in S} x_p \leq B_S, \qquad \forall S \subseteq \mathcal{P}. \tag{15}$$

$$b_S q_t \leq \sum_{p\in S} y_{p,t} \leq B_S q_t, \qquad \forall S \subseteq \mathcal{P}, \quad \forall t \in \mathcal{T}. \tag{16}$$

Cycle population conservation. A special case of (15) is when the subset $S = \{p_1, p_2, \cdots, p_n\}$ of places consists of a cycle, i.e., there is a set of transitions $T = \{t_1, t_2, \cdots, t_n\}$

such that ${}^\bullet p_1 = \{t_n\}$, $p_1^\bullet = \{t_1\}$, ${}^\bullet p_2 = \{t_1\}$, $p_2^\bullet = \{t_2\}$, $\cdots$, ${}^\bullet p_n = \{t_{n-1}\}$, $p_n^\bullet = \{t_n\}$. Since the net is live and stable, the sum of tokens in these places is constant:

$$\sum_{i=1}^{n} X_{p_i} = C_S.$$

Denote by $\mathcal{C}$ any cycle in $\mathcal{N}$, and $C_\mathcal{C}$ the population in $\mathcal{C}$. It then follows

$$\sum_{p\in\mathcal{C}} x_p = C_\mathcal{C}, \qquad \forall \mathcal{C} \in \mathcal{N}, \tag{17}$$

$$\sum_{p\in\mathcal{C}} y_{p,t} = C_\mathcal{C} q_t, \qquad \forall \mathcal{C} \in \mathcal{N}, \quad t \in \mathcal{T}. \tag{18}$$

Reachable markings. Let $C = (C_{p,t})_{|\mathcal{P}|\times|\mathcal{T}|}$ be the incidence matrix such that $C_{p,t} = \sigma_{t,p} - \pi_{p,t}$, where, as usual, $\sigma_{t,p} = 0$ (or $\pi_{p,t} = 0$) if $(t,p) \notin \mathcal{E}$ (or $(p,t) \notin \mathcal{E}$). It is well-known (see e.g. [36]) that any *reachable marking* X from the initial marking M can be written as

$$X^T = M^T + CH, \tag{19}$$

where the superscript ^{T} denotes the transpose operator, and the (column) vector H corresponds to the firing sequence to reach X (or more precisely, the vector of numbers of firings of each transition in order to reach X). Let X in (19) be the random variable of the marking in the stationary regime. Then, by taking expectation in (19) we obtain

$$\boldsymbol{x}^T = M^T + C\boldsymbol{u}^T, \tag{20}$$

where $\boldsymbol{u} = (u_t,\ t \in \mathcal{T})$. Note that $u_t \geq 0$, $t \in \mathcal{T}$, are newly introduced unknown variables. Rewriting (20) in scalar form yields

$$x_p = M_p + \sum_{t\in{}^\bullet p} \sigma_{t,p} u_t - \sum_{t\in p^\bullet} \pi_{t,p} u_t, \qquad \forall p \in \mathcal{P}. \tag{21}$$

4.2 Constraints Derived from Probability Theory

Sample path comparisons. Since for any $t \in \mathcal{T}$, $e_t \leq 1$ almost surely (a.s.), the enabling rate is bounded by one:

$$q_t \leq 1, \qquad \forall t \in \mathcal{T}. \tag{22}$$

For the same reason, we have $X_p e_t \leq X_p$ a.s., so that

$$y_{p,t} \leq x_p, \qquad \forall p \in \mathcal{P}, \quad t \in \mathcal{T}. \tag{23}$$

Another consequence is

$$x_p = EX_p = E\left(X_p \sum_{t\in\mathcal{T}} A_t\right) \geq E\left(X_p \sum_{t\in\mathcal{T}} A_t e_t\right) = \sum_{t\in\mathcal{T}} \frac{\mu_t}{\mu} y_{p,t},$$

so that

$$\mu x_p \geq \sum_{t\in\mathcal{T}} \mu_t y_{p,t}, \qquad \forall p \in \mathcal{P}. \tag{24}$$

According to the relation

$$e_t(\tau) = \left(\prod_{p\in {}^\bullet t} \mathbf{1}_{\{X_p(\tau)\geq \pi_{p,t}\}}\right) \cdot \left(\prod_{p\in {}^\circ t} \mathbf{1}_{\{X_p(\tau)<\eta_{p,t}\}}\right),$$

if $e_t = 1$, then $X_p \geq \pi_{p,t}$ and $X_p \leq \eta_{p,t} - 1$. Therefore,

$$\begin{aligned} (X_p - \pi_{p,t})\, e_t &\geq 0, \qquad \forall t \in \mathcal{T},\ \ p \in {}^\bullet t, \\ (X_p - \eta_{p,t} + 1)\, e_t &\leq 0, \qquad \forall t \in \mathcal{T},\ \ p \in {}^\circ t, \end{aligned}$$

so that, by taking the expectation, we obtain

$$y_{p,t} \geq \pi_{p,t} q_t, \qquad \forall t \in \mathcal{T},\ \ p \in {}^\bullet t, \tag{25}$$

$$y_{p,t} \leq (\eta_{p,t} - 1) q_t, \qquad \forall t \in \mathcal{T},\ \ p \in {}^\circ t. \tag{26}$$

Probabilistic inequalities. According to Chernoff's inequality, we get for all $n \geq 1$,

$$P(X_p \geq n) \leq \frac{x_p}{n} \bigwedge 1, \qquad \forall p \in \mathcal{P} \tag{27}$$

where $\bigwedge$ is the "min" operation.

For bounded places p, $B_p < \infty$, we have for all $n \geq 1$,

$$\begin{aligned} x_p &= E[X_p] = \sum_{i=1}^{B_p} iP(X_p = i) \\ &= \sum_{i=1}^{n-1} iP(X_p = i) + \sum_{i=n}^{B_p} iP(X_p = i) \\ &\leq (n-1) \sum_{i=1}^{n-1} P(X_p = i) + B_p \sum_{i=n}^{B_p} P(X_p = i) \\ &\leq (n-1)P(X_p \leq n-1) + B_p P(X_p \geq n) \\ &= (B_p - n + 1)P(X_p \geq n) + n - 1. \end{aligned}$$

Therefore,

$$P(X_p \geq n) \geq \frac{x_p - n + 1}{B_p - n + 1}\,, \tag{28}$$

or, equivalently,

$$P(X_p < n) \leq 1 - \frac{x_p - n + 1}{B_p - n + 1} = \frac{B_p - x_p}{B_p - n + 1}\,. \tag{29}$$

Consider any transition $t \in \mathcal{T}$ such that all incoming places are bounded, i.e., for all $p \in {}^\bullet t$, $B_p < \infty$. Using the fact that

$$\{e_t = 0\} = \left(\bigcup_{p \in {}^\bullet t} \{X_p < \pi_{p,t}\} \right) \bigcup \left(\bigcup_{p \in {}^\circ t} \{X_p \geq \eta_{p,t}\} \right), \tag{30}$$

we obtain that

$$1 - q_t = P(e_t = 0) \leq \sum_{p \in {}^\bullet t} P(X_p < \pi_{p,t}) + \sum_{p \in {}^\circ t} P(X_p \geq \eta_{p,t}) \leq \sum_{p \in {}^\bullet t} \frac{B_p - x_p}{B_p - \pi_{p,t} + 1} + \sum_{p \in {}^\circ t} \frac{x_p}{\eta_{p,t}}\,,$$

where the last inequality comes from relations (27,29). Hence, we obtain an *enabling lower bound*:

$$q_t \geq 1 - \sum_{p \in {}^\bullet t} \frac{B_p - x_p}{B_p - \pi_{p,t} + 1} - \sum_{p \in {}^\circ t} \frac{x_p}{\eta_{p,t}}\,, \qquad \forall t \in \mathcal{T}, \quad s.t.\ \forall p \in {}^\bullet t,\ B_p < \infty. \tag{31}$$

Applying again Chernoff's inequality to (30) yields

$$\begin{aligned}
q_t &= P\left(\sum_{p \in {}^\bullet t} \mathbf{1}_{\{X_p \geq \pi_{p,t}\}} + \sum_{p \in {}^\circ t} \mathbf{1}_{\{X_p < \eta_{p,t}\}} \geq |\,{}^\bullet t| + |\,{}^\circ t| \right) \\
&\leq \frac{1}{|\,{}^\bullet t| + |\,{}^\circ t|} \left(\sum_{p \in {}^\bullet t} E\mathbf{1}_{\{X_p \geq \pi_{p,t}\}} + \sum_{p \in {}^\circ t} E\mathbf{1}_{\{X_p < \eta_{p,t}\}} \right) \\
&= \frac{1}{|\,{}^\bullet t| + |\,{}^\circ t|} \left(\sum_{p \in {}^\bullet t} P(X_p \geq \pi_{p,t}) + \sum_{p \in {}^\circ t} P(X_p < \eta_{p,t}) \right) \\
&\leq \frac{1}{|\,{}^\bullet t| + |\,{}^\circ t|} \left(\sum_{p \in {}^\bullet t} \left(\frac{x_p}{\pi_{p,t}} \bigwedge 1 \right) + \sum_{p \in {}^\circ t} \frac{B_p - x_p}{B_p - \eta_{p,t} + 1} \right)
\end{aligned}$$

where the last inequality comes from relations (27,29). Thus, we obtain an *enabling upper bound*:

$$q_t \leq \frac{1}{|\,{}^\bullet t| + |\,{}^\circ t|} \left(\sum_{p \in {}^\bullet t} \left(\frac{x_p}{\pi_{p,t}} \bigwedge 1 \right) + \sum_{p \in {}^\circ t} \frac{B_p - x_p}{B_p - \eta_{p,t} + 1} \right), \quad \forall t \in \mathcal{T}, \quad s.t.\ \forall p \in {}^\circ t,\ B_p < \infty. \tag{32}$$

Note that in (32), the "min" operator "$\bigwedge$" is nonlinear. However, linear inequalities can be generated by taking either operand of any of the "min" operators.

Consider now an arbitrary bounded place p with bound B_p. Then for any $t \in \mathcal{T}$,

$$\begin{aligned} B_p q_t &= B_p P(e_t = 1) = B_p \sum_{i=0}^{B_p} P(X_p = i,\ e_t = 1) \\ &= \sum_{i=0}^{B_p} iP(X_p = i,\ e_t = 1) + \sum_{i=0}^{B_p} (B_p - i)P(X_p = i,\ e_t = 1) \\ &\leq \sum_{i=0}^{B_p} iP(X_p = i,\ e_t = 1) + \sum_{i=0}^{B_p} (B_p - i)P(X_p = i) \\ &= y_{p,t} + B_p - x_p. \end{aligned}$$

Thus,

$$x_p \leq y_{p,t} + B_p(1 - q_t), \qquad \forall t \in \mathcal{T},\ \ p \in \mathcal{P}\ s.t. B_p < \infty. \tag{33}$$

Similarly,

$$\begin{aligned} q_t &= \sum_{i=0}^{B_p} P(X_p = i,\ e_t = 1) \\ &= \sum_{i=0}^{B_p} iP(X_p = i,\ e_t = 1) + \sum_{i=0}^{B_p} (1 - i)P(X_p = i,\ e_t = 1) \\ &= y_{p,t} + P(X_p = 0,\ e_t = 1) + \sum_{i=1}^{B_p} (1 - i)P(X_p = i,\ e_t = 1) \\ &\geq y_{p,t} + \sum_{i=1}^{B_p} (1 - i)P(X_p = i) \\ &= y_{p,t} + P(X_p \geq 1) - x_p \\ &\geq y_{p,t} + \frac{x_p}{B_p} - x_p, \end{aligned}$$

where the last inequality comes from (28). Thus,

$$(1 - \frac{1}{B_p})x_p \geq y_{p,t} - q_t, \qquad \forall t \in \mathcal{T},\ \ p \in \mathcal{P}\ s.t. B_p < \infty. \tag{34}$$

Single entry transitions. For all $t \in \mathcal{T}$, if ${}^\bullet t = \{p\}$ and $t^\circ = \emptyset$, then

$$x_p = \sum_{i=1}^{\infty} iP(X_p = i) = \sum_{i=\pi_{p,t}}^{\infty} iP(X_p = i,\ e_t = 1) + \sum_{i=1}^{\pi_{p,t}-1} iP(X_p = i) \leq y_{p,t} + \pi_{p,t} - 1,$$

so that

$$x_p \leq y_{p,t} + \pi_{p,t} - 1, \qquad t \in \mathcal{T}, \quad {}^\bullet t = \{p\}, \quad {}^\circ t = \emptyset. \tag{35}$$

Little's Law. According to Little's law (see e.g. [40]), for all $p \in \mathcal{P}$,

$$x_p = \lambda_p R_p,$$

where λ_p is the input rate tokens at place p, and R_p is the mean token sojourn time at place p. Since R_p is lower bounded by the minimum firing times of output transitions of p, we obtain

$$x_p \geq \lambda_p \frac{1}{\sum_{t \in p^\bullet} \mu_t} = \left(\sum_{t \in {}^\bullet p} \mu_t \sigma_{t,p} q_t \right) \cdot \frac{1}{\sum_{t \in p^\bullet} \mu_t},$$

or, equivalently,

$$x_p \sum_{t \in p^\bullet} \mu_t \geq \sum_{t \in {}^\bullet p} \mu_t \sigma_{t,p} q_t, \qquad \forall p \in \mathcal{P}. \tag{36}$$

4.3 Subnet Throughputs

Like in [21], we derive bounds on throughputs of transitions by comparing throughputs of $\mathcal{N}$ with those in the subnets (when they are considered in isolation) of $\mathcal{N}$. We will consider in particular two special classes of subnets: *strongly connected state machines* (SCSM) and *strongly connected marked graphs* (SCMG).

Let $\mathcal{N} = (\mathcal{P} \bigcup \mathcal{T}, \mathcal{E})$ be an arbitrary Petri net, and $\mathcal{N}' = (\mathcal{P}' \bigcup \mathcal{T}', \mathcal{E}')$ a subnet of $\mathcal{N}$, i.e., $\mathcal{P}' \subseteq \mathcal{P}$, $\mathcal{T}' \subseteq \mathcal{T}$, and $\mathcal{E}'$ is a restriction of $\mathcal{E}$ on $\{\mathcal{P}' \bigcup \mathcal{T}'\} \times \{\mathcal{P}' \bigcup \mathcal{T}'\}$. Assume that the transitions of $\mathcal{T}'$ (resp. arcs of $\mathcal{E}'$, places of $\mathcal{P}'$) have the same sequences of firing times (resp. weights, initial markings) in both nets. Assume further that none of the places of $\mathcal{P}'$ is connected with transitions of $\mathcal{T} - \mathcal{T}'$ in the original net $\mathcal{N}$ by non-inhibitor arcs, i.e. in $\mathcal{N}$, there is no $(t,p) \in (\mathcal{T} - \mathcal{T}') \times \mathcal{P}'$ such that $t \in {}^\bullet p \cup p^\bullet$.

Let θ'_t denote the throughput of transition $t \in \mathcal{T}'$ when the subnet $\mathcal{N}'$ is considered in isolation. The following theorems show that under some conditions, the throughputs of these transitions of the $\mathcal{N}'$ are upper bounds of the throughputs of the same transitions in the original net.

Theorem 1 *If $\mathcal{N}'$ is a strongly connected marked graph, then for any transition t in $\mathcal{N}'$, $\theta_t \leq \theta'_t$.*

Proof. Due to the fact that in the original net $\mathcal{N}$, none of the places of $\mathcal{P}'$ is connected with transitions of $\mathcal{T} - \mathcal{T}'$ by non-inhibitor arcs, the subnet is connected with the rest of

the system only through transitions of $\mathcal{T}'$. As $\mathcal{N}'$ is a strongly connected marked graph, no transitions in $\mathcal{N}'$ are in conflict. Moreover, in $\mathcal{N}$, the firing mechanism is race policy with age memory. Thus, for any transition $t \in \mathcal{T}'$, the only effect that tokens in places of ${}^\bullet t - \mathcal{P}'$ have is delaying the firings of t of $\mathcal{T}'$. Thus, by the monotonicity property of marked graphs [11], we conclude that $\theta_t \leq \theta'_t$ for all $t \in \mathcal{T}'$. ∎

Theorem 2 *Assume that $\mathcal{N}'$ is a strongly connected state machine such that for any two transitions t_1 and t_2 of $\mathcal{N}'$, t_1 and t_2 are in conflict in $\mathcal{N}'$ implies that t_1 and t_2 are in equal conflict in $\mathcal{N}$, i.e. ${}^\bullet t_1 = {}^\bullet t_2$. Then for any transition t in $\mathcal{N}'$, $\theta_t \leq \theta'_t$.*

Proof. The proof is similar to that of Theorem 1. Note first that the subnet $\mathcal{N}'$ is connected with the rest of the system only through transitions of $\mathcal{T}'$ or inhibitor arcs. Under the assumption of the theorem, any two transitions which are in (equal) conflict in $\mathcal{N}'$ are in equal conflict in $\mathcal{N}$. Moreover, in $\mathcal{N}$, the firing mechanism is race policy with age memory. Thus, in $\mathcal{N}$, tokens in the places of $\cup_{t\in\mathcal{T}'} {}^\bullet t - \mathcal{P}'$ will not change the winners of firing races among transitions of $\mathcal{T}'$. In other words, the only effect that tokens in places of ${}^\bullet t - \mathcal{P}'$ have is delaying the firings of transitions of $\mathcal{T}'$. Thus, by the monotonicity property of state machines [8], we conclude that $\theta_t \leq \theta'_t$ for all $t \in \mathcal{T}'$. ∎

Note that the above two theorems hold for any arbitrarily fixed sequences of firing times. No Markovian assumption is needed.

Efficient computational algorithms for computing the throughput of state machines have been proposed in the queueing literature, see e.g. [16, 22, 38, 39].

The computation of throughput of SCMGs has been investigated in [11], where various computable upper bounds were proposed. Exact value of the throughput can be obtained by simulation using matrix multiplications in the $(\max, +)$ algebra, see [7].

5 Summary of the Linear Programming Formulation

Theorem 3 *Let $< \mathcal{N}, M, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \boldsymbol{\mu} >$ be an arbitrary Markovian timed Petri net, and $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta})$ an arbitrary linear function defined on the nonnegative state variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta}$ of the net. Let α and β be the solutions of the linear programming problems*

$$\alpha = \min L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta}) \tag{37}$$

$$\beta = \max L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta}) \tag{38}$$

such that the linear constraints of Table (1) are satisfied, where $u_t \geq 0$ *for all* $t \in \mathcal{T}$. *Then,*

$$\alpha \leq L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{\theta}) \leq \beta. \tag{39}$$

Recall that in Table 1, inequalities containing the operator "$\bigwedge$" represent any inequalities generated by taking either operand of any of the "min" operators.

6 Applications

In this section we illustrate applications of the above techniques to the performance analyses. We shall consider two applications, one in manufacturing system, another in parallel computing. Unless otherwise stated, the numerical results are obtained without linear inequalities pertaining to boundedness of subsets of places and subnet throughputs.

6.1 Production Line

The first example is concerned with a production line with infinite supply, see Figure 2-(a). In the example, there are four servers, represented by circles. The first server has an infinite-capacity buffer with an infinite number of production requirements, represented by small dashed circles. The other three servers have finite-capacity buffers: 3, 2 and 4. For $i = 1, 2, 3$, server i starts a service only when the downstream buffer $i+1$ has at least one empty room. This corresponds to the so-called *blocking before service*.

The corresponding Petri net model is depicted in Figure 2-(b), where transitions represent the servers and the initial markings of the places on the bottom represent the buffer capacities.

We assume that the service times at server i are i.i.d. exponentially distributed with parameter μ_i, $i = 1, 2, 3, 4$. The objective function in this problem is the total throughput $\theta_1 + \theta_2 + \theta_3 + \theta_4$. The numerical results are presented and compared with the exact values.

In the experimentation, we have carried out computations for five sets of parameters of μ_i's. The lower and upper bounds are given in the columns "l.b.", "u.b.1" and "u.b.2", whereas the exact values are provided in the column "exact". The upper bounds in column "u.b.1" are obtained by further using subnet throughput constraints. In columns "o.l.b." and "o.u.b.", we also present the bounds computed by linear programming approch based on the linear constraints of Section 4 without Markovian assumption (which implies in particular that the linear equalities of Section 3 are not used).

Table 1: Summary of Linear Constraints

throughput	$\theta_t = \mu_t q_t$	$\forall t \in \mathcal{T}$
flow balance	$\sum_{t \in {}^\bullet p} \ \mu_t \sigma_{t,p} q_t = \sum_{t \in p^\bullet} \ \mu_t \pi_{p,t} q_t$	$\forall p \in \mathcal{P}$
second moment	$2 \sum_{t \in {}^\bullet p} \ \mu_t \sigma_{t,p} y_{p,t} - 2 \sum_{t \in p^\bullet} \ \mu_t \pi_{p,t} y_{p,t}$ $= 2 \sum_{t \in V_p} \ \mu_t \sigma_{t,p} \pi_{p,t} q_t$ $- \sum_{t \in {}^\bullet p} \ \mu_t \sigma_{t,p}^2 q_t - \sum_{t \in p^\bullet} \ \mu_t \pi_{p,t}^2 q_t$	$\forall p \in \mathcal{P}$
population covariance	$\sum_{t \in {}^\bullet p_2} \ \mu_t \sigma_{t,p_2} y_{p_1,t} - \sum_{t \in p_2^\bullet} \ \mu_t \pi_{p_2,t} y_{p_1,t}$ $+ \sum_{t \in {}^\bullet p_1} \ \mu_t \sigma_{t,p_1} y_{p_2,t} - \sum_{t \in p_1^\bullet} \ \mu_t \pi_{p_1,t} y_{p_2,t}$ $= - \sum_{t \in {}^\bullet p_1 \bigcap {}^\bullet p_2 - \mathcal{T}'} \ \mu_t \sigma_{t,p_1} \sigma_{t,p_2} q_t$ $+ \sum_{t \in {}^\bullet p_1 \bigcap p_2^\bullet} \ \mu_t \sigma_{t,p_1} \pi_{p_2,t} q_t$ $+ \sum_{t \in p_1^\bullet \bigcap {}^\bullet p_2} \ \mu_t \pi_{p_1,t} \sigma_{t,p_2} q_t$ $- \sum_{t \in p_1^\bullet \bigcap p_2^\bullet} \ \mu_t \pi_{p_1,t} \pi_{p_2,t} q_t$	$\forall p_1, p_2 \in \mathcal{P}$
liveness	$\sum_{t \in \mathcal{T}} \ q_t \geq 1$	
	$x_p \leq \sum_{t \in \mathcal{T}} y_{p,t}$	$\forall p \in \mathcal{P}$
conflicting transitions	$q_{t_1} \geq q_{t_2}$	$\forall t_1, t_2 \in \mathcal{T}, \ \ s.t.$ $\boldsymbol{\pi}(t_1) \leq \boldsymbol{\pi}(t_2), \ \boldsymbol{\eta}(t_1) \geq \boldsymbol{\eta}(t_2)$
	$q_{t_1} = q_{t_2}$	$\forall t_1, t_2 \in \mathcal{T}, \ \ s.t.$ $\boldsymbol{\pi}(t_1) = \boldsymbol{\pi}(t_2), \ \boldsymbol{\eta}(t_1) = \boldsymbol{\eta}(t_2)$
boundedness	$b_p \leq x_p \leq B_p$	$\forall p \in \mathcal{P}$
	$b_p q_t \leq y_{p,t} \leq B_p q_t$	$\forall p \in \mathcal{P}, \ t \in p^\bullet$
	$b_S \leq \sum_{p \in S} \ x_p \leq B_S$	$\forall S \subseteq \mathcal{P}$
	$b_S q_t \leq \sum_{p \in S} \ y_{p,t} \leq B_S q_t$	$\forall S \subseteq \mathcal{P}, \ t \in \mathcal{T}$
	$x_p \leq y_{p,t} + B_p(1 - q_t)$	$\forall t \in \mathcal{T}, \ p \in \mathcal{P}$
	$(1 - \frac{1}{B_p}) x_p \geq y_{p,t} - q_t$	$\forall t \in \mathcal{T}, \ p \in \mathcal{P}$
cycle population	$\sum_{p \in \mathcal{C}} \ x_p = C_{\mathcal{C}}$	$\forall \mathcal{C} \in \mathcal{N}$
	$\sum_{p \in \mathcal{C}} \ y_{p,t} = C_{\mathcal{C}} q_t$	$\forall \mathcal{C} \in \mathcal{N}, \ t \in \mathcal{T}$
reachable marking	$x_p = M_p + \sum_{t \in {}^\bullet p} \ \sigma_{t,p} u_t - \sum_{t \in p^\bullet} \ \pi_{t,p} u_t$	$\forall p \in \mathcal{P}$
sample path comparisons	$q_t \leq 1$	$\forall t \in \mathcal{T}$
	$y_{p,t} \leq x_p$	$\forall p \in \mathcal{P}, \ \ t \in \mathcal{T}$
	$\mu x_p \geq \sum_{t \in \mathcal{T}} \ \mu_t y_{p,t}$	$\forall p \in \mathcal{P}$
	$x_p \leq y_{p,t} + \pi_{p,t} - 1$	$\forall t \in \mathcal{T}, \ s.t. \ {}^\bullet t = \{p\}, \ \ {}^\circ t = \emptyset$
	$y_{p,t} \geq \pi_{p,t} q_t$	$\forall t \in \mathcal{T}, \ \ p \in {}^\bullet t$
	$y_{p,t} \leq (\eta_{p,t} - 1) q_t$	$\forall t \in \mathcal{T}, \ \ p \in {}^\circ t$
enabling bound	$q_t \geq 1 - \sum_{p \in {}^\bullet t} \frac{B_p - x_p}{B_p - \pi_{p,t} + 1} - \sum_{p \in {}^\circ t} \frac{x_p}{\eta_{p,t}}$	$\forall t \in \mathcal{T}, \ s.t. \ \forall p \in {}^\bullet t, \ B_p < \infty$
	$q_t \leq \frac{1}{\lvert {}^\bullet t \rvert + \lvert {}^\circ t \rvert} \left(\sum_{p \in {}^\bullet t} \left(\frac{x_p}{\pi_{p,t}} \bigwedge 1 \right) + \sum_{p \in {}^\circ t} \frac{B_p - x_p}{B_p - \eta_{p,t} + 1} \right)$	$\forall t \in \mathcal{T}, \ s.t. \ \forall p \in {}^\circ t, \ B_p < \infty$
Little's Law	$x_p \cdot \left(\sum_{t \in p^\bullet} \ \mu_t \right) \geq \sum_{t \in {}^\bullet p} \ \mu_t \sigma_{t,p} q_t$	$\forall p \in \mathcal{P}$
Subnet throughput	$\theta_t \leq \theta'_t$	$\forall t \in \mathcal{T}'$ s.t. $\mathcal{N}'$ is SCMG
	$\theta_t \leq \theta'_t$	$\forall t \in \mathcal{T}'$ s.t. $\mathcal{N}'$ is SCSM

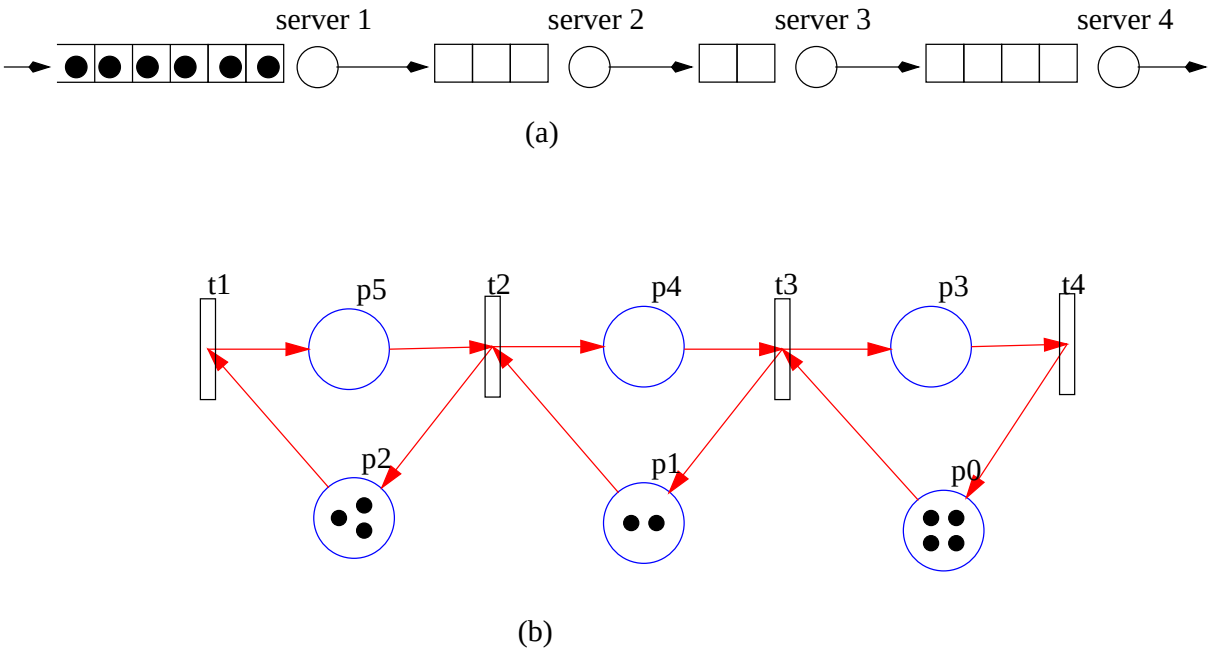

Figure 2: (a): Example of production line. (b): The corresponding Petri net model.

Table 2: Bounds on the throughput of the production line

Case	μ_1	μ_2	μ_3	μ_4	l.b.	exact	u.b.1	u.b.2	o.l.b.	o.u.b.
1	1	1.25	2	0.5	1.165	1.951	1.994	2.000	0.930	2.000
2	1	1.25	2	2.5	1.829	2.978	3.306	3.529	1.481	4.000
3	1	1.25	1.25	2.5	1.581	2.873	3.306	3.333	1.333	4.000
4	1	1.25	1.25	1	1.359	2.757	3.306	3.333	1.111	4.000
5	1.111	1.111	1.111	1.111	1.350	2.667	2.963	2.963	1.111	4.444

Recall that the Petri net is a marked graph so that according to [11] the throughput is increasing in the firing rates of transitions. Such a fact is clearly shown in the column "exact" for cases 1, 2, 3 and 4. It is worthwhile noticing that the lower and upper bounds in the columns "l.b.", "u.b.1" and "u.b.2" also reflect such monotonicity.

6.2 Cyclic Execution

Consider now performance analysis of a parallel computing system. Parallel programs are represented by directed acyclic graphs, referred to as *task graphs*, where vertices correspond to tasks of a parallel program, and directed edges correspond to *precedence relations* between tasks: a task can start execution only when all its predecessors have completed execution. The tasks are assigned to the parallel processors for execution according to some predefined rules.

In our example, parallel programs have the same structure, given by the task graph in Figure 3-(a). These programs differ only in the running times of tasks which are independently and exponentially distributed random variables, with parameters $\mu_1, \mu_2, \cdots, \mu_6$ for tasks $1, 2, \cdots, 6$. There are three identical processors. Tasks 1 and 2 are assigned to processor 1, tasks 3 and 4 to processor 2, and tasks 5 and 6 to processor 3.

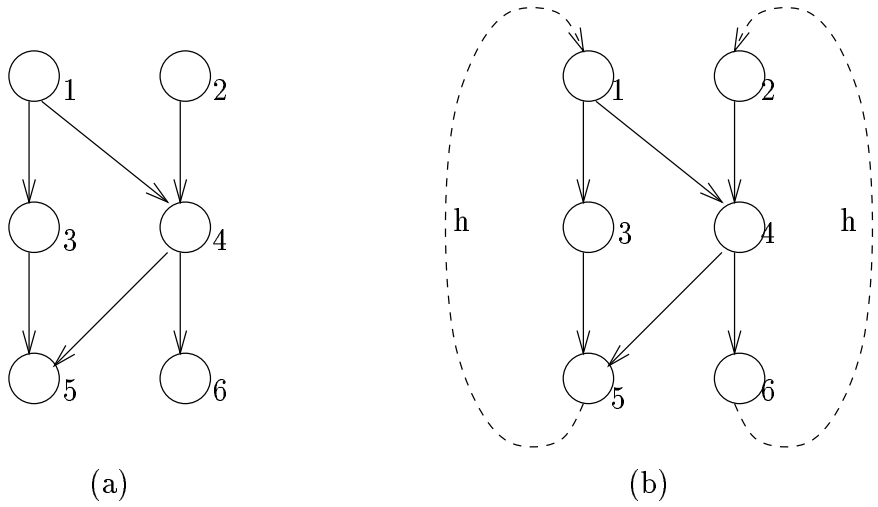

Figure 3: (a) Task graph of parallel programs. (b) Cyclic execution of the parallel program.

On each processor, different instantiations of the same task are executed according to the rule first come first serve (FCFS). i.e., task i of the n-th arrived program can start execution only after task i of program $n-1$ completes. Different tasks assigned to the same processor are, however, executed according to the processor sharing (PS) discipline. In our example, since only two different tasks are assigned to each processor, the processor

is shared by at most two tasks. A parallel program is considered completed if all its tasks finish their execution.

We consider cyclic execution of the task graph, cf. Figure 3-(b). The cyclic execution is defined in such a way that task $t1$ (resp. task $t2$) of program $n+h$ can start execution only after task $t5$ (resp. $t6$) of program n completes execution, $n = 1, 2, \ldots$ The number h is referred to as the *height* of the cyclic execution in the literature [30].

The representation of this parallel computing system by STPN is illustrated in Figure 4. The initial marking of place $p1$ (the same for place $p2$) corresponds to the height h.

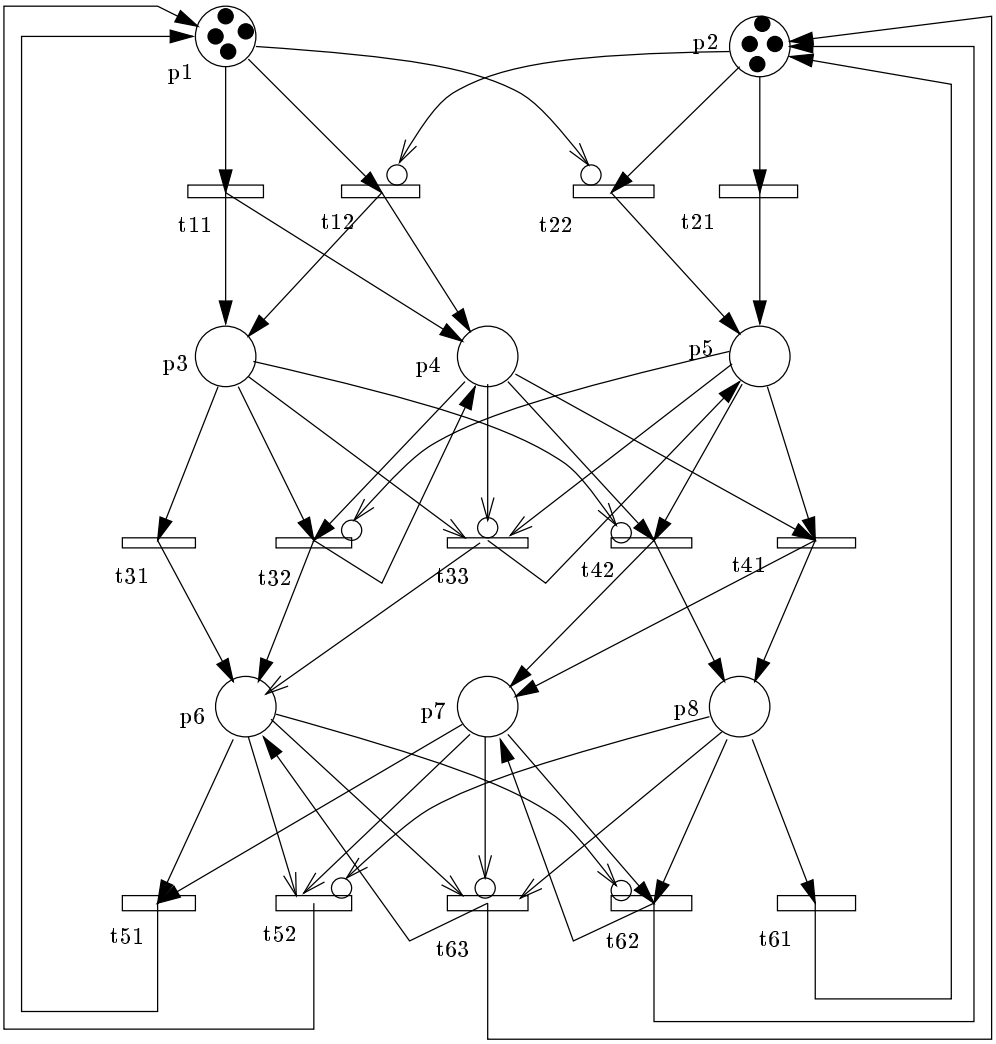

Figure 4: Petri net representation of the cyclic execution

Inhibitor arcs are used to model the PS mechanism. Transitions $t11$ and $t12$ (resp. t21 and t22, t31 and t32 and t33, t41 and t42, t51 and t52, t61 and t62 and t63) represent tasks 1 (resp. 2, 3, 4, 5, 6). Firing times of transitions t11 and t12 (resp. t21 and t22,

t31 and t32 and t33, t41 and t42, t51 and t52, t61 and t62 and t63, are exponentially distributed with parameter $\mu_1/2$ (resp. $\mu_2/2$, $\mu_3/2$, $\mu_4/2$, $\mu_5/2$, $\mu_6/2$). Two or three transitions are used for each task in order to represent situations whether the execution of a task is shared with others. Note that transitions t32 and t33 (resp. t62 and t63) are never enabled simultaneously. The use of two additional transitions for task 3 (resp. task 6) is due to the fact that in each program, task 4 (resp. task 5) is allowed to start execution only when both tasks 1 and 2 (resp. tasks 3 and 4) have completed execution. Thus, the execution of task 3 (resp. task 6) is not shared with task 4 (resp. task 5) if only task 1 or only task 2 (resp. only task 3 or only task 4) completes execution.

The objective function in this problem is still the total throughput, i.e. the sum of transition throughputs. It is easy to see that this total throughput is equal to six times the throughput of the parallel system in terms of the number of programs completed per unit of time.

In Table 3, we provide numerical results for five different sets of parameters with fixed height $h = 4$. The lower and upper bounds are given in the columns "l.b." and "u.b.", whereas in columns "o.l.b." and "o.u.b.", we also present the bounds computed by linear programming approch based on the linear constraints of Section 4 without Markovian assumption (which implies in particular that the linear equalities of Section 3 are not used).

Table 3: Bounds on the throughput of the cyclic execution

Case	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	l.b.	u.b.	o.l.b.	o.u.b.
1	2	2	2	2	2	2	1.161	4.8	1	12
2	2	2	1	1	3	3	0.980	3	0.818	6
3	2	1	2	1	3	3	0.970	2.4	0.818	6
4	2	1	1	2	3	3	0.964	3	0.818	6
5	1	3	3	1	2	2	0.955	3	0.818	6

In Figures 5 and 6, we provide the curves of the bounds as functions of the the height of the cyclic execution in Cases 1 and 2.

7 Conclusions and Extensions

In this paper, we have established performance bounds for Markovian STPN by taking a linear programming approach. We first provided a set of linear equality constraints among the expectation of state variables in the Petri nets, such as token numbers in the places and indicator functions of transition enabling. We further obtained an augmenting set of

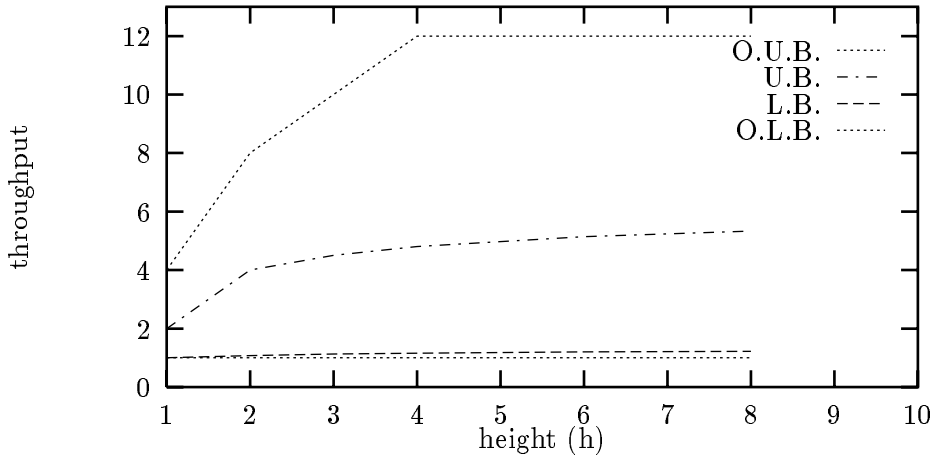

Figure 5: Bounds as functions of height of the cyclic execution for Case 1

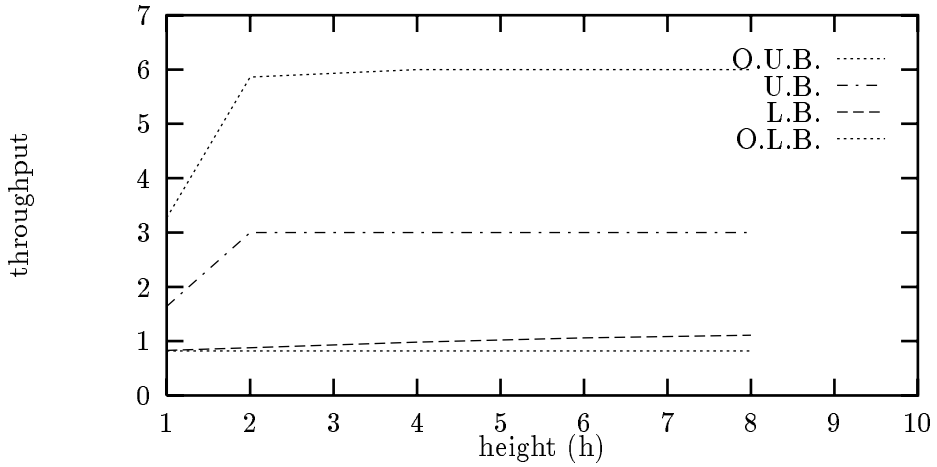

Figure 6: Bounds as functions of height of the cyclic execution for Case 2

linear equalities and inequalities by exploiting structural and probabilistic properties of the Petri nets. These linear constraints allowed us to compute upper and lower bounds of performance measures by solving the linear program. We have applied this method to performance analyses of a manufacturing system and a parallel system.

The constraints derived in Section 4 are not restricted to exponentially distributed firing times. These inequalities can also be used in operational analysis of timed Petri nets.

In Theorems 1 and 2, we compared throughputs of transitions in a net and those in a SCMG or a SCSM subnet (when it is considered in isolation). Similar inequalities can be obtained using monotonicity results [11, 8, 12] for other subnets.

Throughout this paper, the transitions have single-server semantics. Our analysis can be extended immediately to STPN with bounded marking and *infinite-server* transitions. Indeed, in such a case, each infinite-server transition can be replaced by K single-server transitions, where K is the upper bound of the token numbers in the places. More precisely, we replace each infinite-server transition t by K single-server transitions $t_1, t_2, \cdots, t_K$ in such a way that $\pi_{p,t_k} = k\pi_{p,t}$, $k = 1, 2, \cdots, K$, and $\sigma_{t_k,p} = \sigma_{t,p}$. An example of such a transformation is illustrated in Figure 7, where $K = 3$.

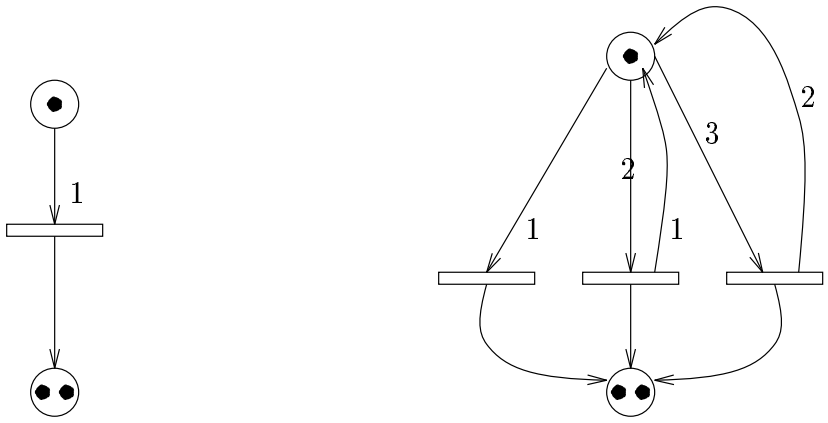

Figure 7: Transformation from infinite-server transition to single-server transitions. (a) an infinite-server transition. (b) equivalent single-server transitions.

In our analysis, we assumed that the firing times of each transition are i.i.d. exponential random variables with a fixed parameter. It is simple to extend the results to the case of *marking dependent firing rates*, i.e., the firing rate of a transition depends on the marking of input places, provided the number of different firing rates is bounded. As an example, consider a transition t with a single input place. Let μ_t^1, μ_t^2 and μ_t^3 be the firing rates of

transition t when there are one token, two tokens, and more than 3 tokens in the input place. We replace the transition by three transitions t_1, t_2 and t_3 with firing rates μ_t^1, μ_t^2 and μ_t^3, respectively, in such a way that at any time at most one of the transitions is enabled, see Figure 8. The set of outgoing places are the same as that of t: $\sigma_{t_k,p} = \sigma_{t,p}$, $k = 1, 2, 3$.

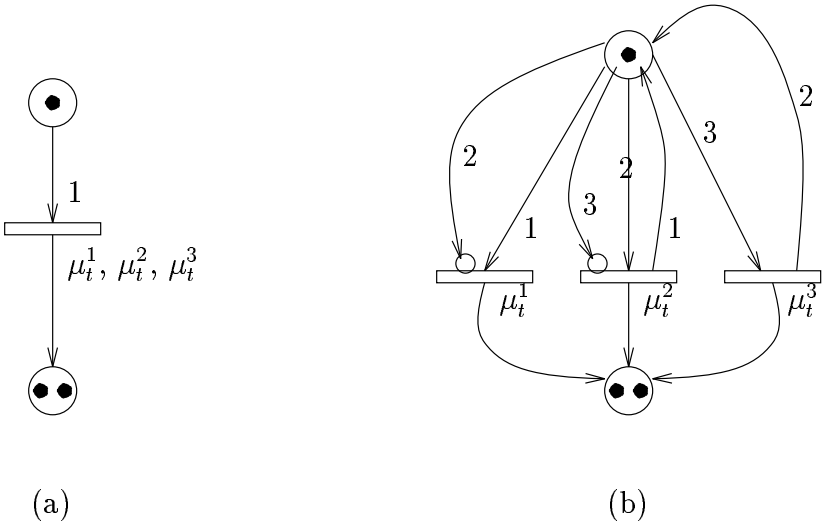

Figure 8: Transformation from marking-dependent firing rate to marking-independent firing rates. (a) transition with 3 firing rates. (b) equivalent transitions with fixed firing rates.

Our approach can be extended to the case that firing times have *phase-type distributions* [37]. A phase-type distribution can be considered as the distribution of the time that a token passes through an ordinary Markovian state machine with a single source and a single sink transitions. The firing times have exponential distributions for all transitions except the source and the sink which are immediate transitions. The sink transition represents the absorbing state. Let transition t have a phase-type distribution which is represented by a Markovian state machine $\mathcal{N}'$ with source transition t_0 and sink transition t_s. We replace transition t in the original net by the subnet $\mathcal{N}'$ as follows. For any $p \in \mathcal{P}$ and any transition $t' \neq t_s$ of $\mathcal{N}'$ (including t_0 and excluding t_s), let $\pi_{p,t'} = \pi_{p,t}$, $\eta_{p,t'} = \eta_{p,t}$, and $\sigma_{t',p} = \pi_{p,t}$. For any $p \in \mathcal{P}$, $\sigma_{t_s,p} = \sigma_{t,p}$. Moreover, $\pi_{t_0,p_0} = \sigma_{t_s,p_0} = 1$, where p_0 is a new place with initial marking 1. An example of the construction is illustrated in Figure 9 for an Erlang distribution with 3 stages. Recall that the firing mechanism under consideration is race policy with age memory. The reader can therefore easily check that when transitions whose firing times have phase-type distributions are thus replaced by corresponding Markovian state machines, we obtain a stochastically equivalent STPN with exponential firing times.

The performance measures considered in the paper are mostly the throughputs of transitions and the expectations of X_p and $X_p e_t$. The same approach can be used to

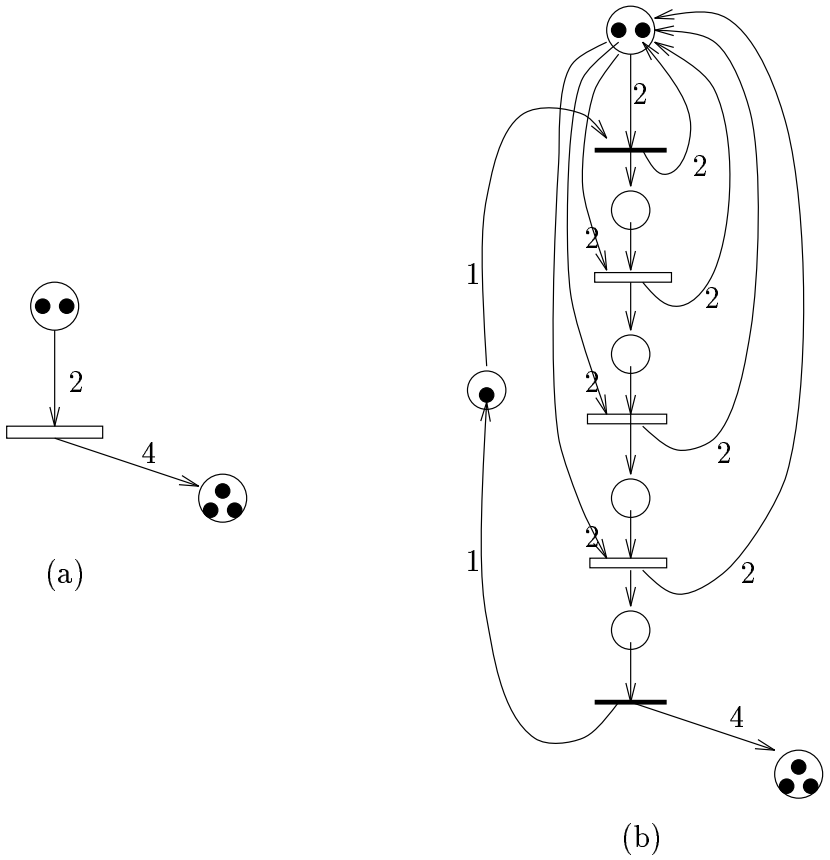

Figure 9: Transformation from phase-type firing times to exponential firing times. (a) transition with 3-stages-Erlang-distribution firing times. (b) transitions with exponential firing times.

obtain linear equalities among *higher moments* of the token numbers. More precisely, for any $m \geq 1$, linear equalities can be established for variables $E[X_p^k]$ and $E[X_p^k e_t]$, where $k = 1, 2, \cdots, m$, by equating $E\left[E\left[X_p^{m+1}(\tau_{n+1})|\mathcal{F}_{\tau_n}\right]\right]$ and $E\left[X_p^{m+1}(\tau_n)\right]$. Similarly, for any $m \geq 1$, linear equalities can be established for variables $E[X_{p_1}^{k_1} X_{p_2}^{k_2}]$, $E[X_{p_1}^{k_1} e_t]$ and $E[X_{p_2}^{k_2} e_t]$, where $k_1, k_2 = 1, 2, \cdots, m$, by equating $E\left[E\left[X_{p_1}^{m+1}(\tau_{n+1}) X_{p_2}^{m+1}(\tau_{n+1})|\mathcal{F}_{\tau_n}\right]\right]$ and $E\left[X_{p_1}^{m+1}(\tau_n)[X_{p_2}^{m+1}(\tau_n)\right]$.

Finally, we remark that when the weights of the STPN are real numbers, all our analyses go through straightforwardly and the same results hold.

8 Appendix: Elimination of Immediate Transitions

We present here a direct transformation technique which removes immediate transitions playing roles of synchronization and/or routing. We assume that for any such immediate transition t, t is the only output transition of all its input places, i.e. $p^\bullet = \{t\}$ for all $p \in {}^\bullet t$. Further, we assume that for any immediate transition t, ${}^\circ t = \emptyset$, $\pi_{p,t} = 1$, $p \in {}^\bullet t$, and

- either $\sigma_{t,p'}(n) = 1$ a.s., $p' \in t^\bullet$, $n = 1, 2, \ldots$;
- or $\sigma_{t,p'}(n) \leq 1$ a.s., $p' \in t^\bullet$, $n = 1, 2, \ldots$, and for all $p \in {}^\bullet t$, $|{}^\bullet p| = 1$ and $\sigma_{{}^\bullet p,p}(n) = 1$ a.s., $n = 1, 2, \ldots$, where, with a harmless abuse of notation, the index ${}^\bullet p$ denotes the unique transition preceding place p.

We show below that this kind of immediate transitions can be removed from the net without changing the firing behavior of the other transitions. Consider a net $\mathcal{N} = (\mathcal{P} \bigcup \mathcal{T}, \mathcal{E})$ with initial marking M and weights $\boldsymbol{\pi}, \boldsymbol{\sigma}$. Let t_0 be an immediate transition, and ${}^\bullet t_0 = \{p^1, p^2, \cdots, p^h\}$, $t_0^\bullet = \{p_1, p_2, \cdots, p_k\}$. Without loss of generality, we assume that $\min_{p \in {}^\bullet t_0} M_p = 0$.

We construct a new net $\widetilde{\mathcal{N}} = (\widetilde{\mathcal{P}} \bigcup \widetilde{\mathcal{T}}, \widetilde{\mathcal{E}})$ with initial marking $\widetilde{M}$ and weights $\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{\sigma}}$. The key idea is to create a place p_j^i for each pair of input place p^i and output place p_j of transition t_0. The set of input transitions of p_j^i is the union of ${}^\bullet p^i$ and ${}^\bullet p_j$. The set of output transitions of p_j^i is $p_j^\bullet$. Such a transformation is illustrated in Figure 10, where transition t_0 is an immediate transition.

The mathematical definition of $\widetilde{\mathcal{N}}$ is as follows.

$$\widetilde{\mathcal{T}} \;=\; \mathcal{T} - \{t_0\},$$

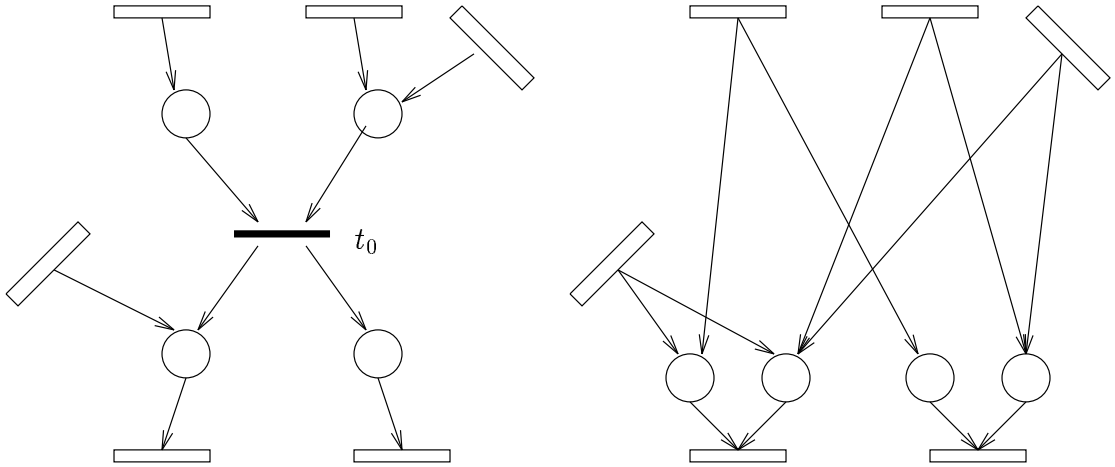

Figure 10: Removal of Immediate Transitions. (a): A subnet containing an immediate transition. (b): The subnet without immediate transition.

$$\begin{aligned}\widetilde{\mathcal{P}} &= \mathcal{P} - {}^\bullet t_0 - t_0^\bullet + \{p_j^i \mid 1 \le i \le h,\ 1 \le j \le k\},\\ \widetilde{\mathcal{E}} &= \mathcal{E} - \{(u,v) \mid u \in {}^\bullet t_0 \cup t_0^\bullet,\ \text{or } v \in {}^\bullet t_0 \cup t_0^\bullet\} + E\end{aligned}$$

where E is defined by

$$\begin{aligned}E &= \{(t,p_j^i) \mid (t,p^i) \in \mathcal{E},\ 1 \le i \le h,\ 1 \le j \le k\}\\ &\quad \bigcup \{(t,p_j^i) \mid (t,p_j) \in \mathcal{E},\ 1 \le i \le h,\ 1 \le j \le k\}\\ &\quad \bigcup \{(p_j^i,t) \mid (p_j,t) \in \mathcal{E},\ 1 \le i \le h,\ 1 \le j \le k\}.\end{aligned}$$

The initial marking $\widetilde{M}$ and weights $\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{\sigma}}$ are defined accordingly:

$$\widetilde{M}_p = \begin{cases} M_p, & p \in \mathcal{P} - {}^\bullet t_0 - t_0^\bullet;\\ M_{p^i} + M_{p_j}, & p = p_j^i,\ 1 \le i \le h,\ 1 \le j \le k.\end{cases}$$

$$\tilde{\pi}_{p,t}(n) = \begin{cases} \pi_{p,t}(n), & (p,t) \in \mathcal{E},\ p \in \mathcal{P} - {}^\bullet t_0 - t_0^\bullet;\\ \pi_{p_j,t}(n), & (p_j,t) \in \mathcal{E},\ p = p_j^i,\ 1 \le i \le h,\ 1 \le j \le k.\end{cases}$$

$$\tilde{\sigma}_{t,p}(n) = \begin{cases} \sigma_{t,p}(n), & (t,p) \in \mathcal{E},\ p \in \mathcal{P} - {}^\bullet t_0 - t_0^\bullet;\\ \sigma_{t,p^i}(n)\sigma_{t_0,p_j}(n), & (t,p^i) \in \mathcal{E},\ p = p_j^i,\ 1 \le i \le h,\ 1 \le j \le k;\\ \sigma_{t,p_j}(n), & (t,p_j) \in \mathcal{E},\ p = p_j^i,\ 1 \le i \le h,\ 1 \le j \le k.\end{cases}$$

It is easily seen that if the sequences of the firings times are the same for the same transitions in $\mathcal{N}$ and $\widetilde{\mathcal{N}}$, their firing commencement and completion times are identical. The detailed proof can be done by induction and is left to the interested reader.

Acknowledgements: The author is very grateful to Dr Alain JEAN-MARIE for constructive comments and for efficient help in the computation of numerical results.

References

[1] M. Ajmone Marsan, G. Balbo, A. Bobbio, G. Chiola, G. Conte, A. Cumani, "The Effect of Execution Policies on the Semantics and Analysis of Stochastic Petri Nets", *IEEE Trans. Software Eng.*, Vol 15, pp. 832-846, 1989.

[2] M. Ajmone Marsan, G. Balbo, and G. Conte. "A Class of Generalized Stochastic Petri Nets for the Performance Analysis of Multiprocessor Systems", *ACM Transactions on Computer Systems*, Vol. 2, No. 1, May 1984.

[3] M. Ajmone Marsan, G. Balbo, G. Conte, *Performance Models of Multiprocessor Systems*, Cambridge, MA, MIT Press, 1986.

[4] F. Baccelli, "Ergodic Theory of Stochastic Petri Nets", *The Annals of Probability*, Vol. 20, pp. 375-396, 1992.

[5] F. Baccelli, G. Balbo, R.J. Boucherie, J. Campos, and G. Chiola. "Annotated Bibliography on Stochastic Petri Nets", In O.J. Boxma and G.M. Koole (editors), *Performance Evaluation of Parallel and Distributed Systems — Solution Methods.* CWI Tract 105 & 106, CWI, Amsterdam, 1994.

[6] F. Baccelli, P. Bremaud, *Elements of Queueing Theory*, Springer-Verlag, Berlin, 1994.

[7] F. Baccelli, M. Canales, "Parallel Simulation of Stochastic Petri Nets Using Recursive Equations", *ACM-Tomacs*, Vol. 3, pp. 20-41, 1993.

[8] F. Baccelli, G. Cohen, B. Gaujal, "Recursive Equations and Basic properties of Timed Petri Nets", *Journal of Discrete Event Systems*, Vol. 1, pp. 415-439, 1992.

[9] F. Baccelli and B. Gaujal. "Stationary regime and stability of free-choice Petri nets", In *Proceedings of the 11-th International Conference on Analysis and Optimisation of Systems*, Juan les Pins, France, June 1994. Springer Verlag.

[10] F. Baccelli and P. Konstantopoulos. "Estimates of Cycle Times in Stochastic Petri Nets", *Lecture Notes in Control and Information Sciences*, Springer Verlag, Vol. 177, pp. 1–21, 1992.

[11] F. Baccelli, Z. Liu, "Comparison Properties of Stochastic Decision Free Petri Nets", *IEEE Trans. on Automatic Control*, Vol. 37, pp. 1905-1920, 1992.

[12] F. Baccelli, Z. Liu, M. Silva, "Global and Local Monotonicities of Stochastic Petri Nets", in preparation.

[13] D. Bertsimas, I. Paschalidis, J. Tsitsiklis, "Optimization of Multiclass Queueing Networks: Polyhedral and Nonlinear Characterizations of Achievable Performance", to appear in *Annals of Applied Probability*.

[14] D. Bertsimas, I. Paschalidis, J. Tsitsiklis, "Branching Bandits and Klimov's Problem: Achievable Region and Side Constraints", preprint.

[15] R.J. Boucherie. "A Characterisation of Independence for Competing Markov Chains with Applications to Stochastic Petri Nets", *Proceedings of PNPM*, 1993.

[16] J. P. Buzen, "A computational Algorithm for Closed Queueing Networks with Exponential Servers." *Comm. ACM*, Vol. 14, pp. 527-531, 1973.

[17] J. Campos, G. Chiola, and M. Silva. "Ergodicity and Throughput Bounds of Petri Nets with Unique Consistent Firing Count Vector", *IEEE Transactions on Software Engineering*, Vol. 17, pp. 117–125, February 1991.

[18] J. Campos, G. Chiola, and M. Silva. "Properties and performance bounds for closed free choice synchronized monoclass queueing networks", *IEEE Transactions on Automatic Control*, Vol. 36, pp. 1368–1382, December 1991.

[19] J. Campos, J. M. Colom, H. Jungnitz, and M. Silva. "A General Iterative Technique for Approximate Throughput Computation of Stochastique Marked Graphs", In *Proceedings of the 5th International Workshop on Petri Nets and Performance Models*, pp. 138-147, Toulouse, France, October 1993. IEEE-Computer Society Press.

[20] J. Campos, B. Sánchez, and M. Silva. "Throughput Lower Bounds for Markovian Petri Nets: Transformation Techniques", In *Proceedings of the 4th International Workshop on Petri Nets and Performance Models*, pages 322–331, Melbourne, Australia, December 1991. IEEE-Computer Society Press.

[21] J. Campos, M. Silva, "Embedded Product-form Queueing Networks and the Improvement of Performance Bounds for Petri Net Systems", *Performance Evaluation*, Vol. 18, pp. 3-19, 1993.

[22] K. M. Chandy and C. H. Sauer, "Computational Algorithms for Product Form Queueing Networks." *Comm. ACM*, Vol. 23, pp. 573-583, 1980.

[23] G. Chiola. "GreatSPN 1.5 Software Architecture", In *Proc.* 5^{th} *Int. Conf. Modeling Techniques and Tools for Computer Performance Evaluation*, Torino, Italy, February 1991.

[24] G. Chiola, M. Ajmone Marsan, G. Balbo, and G. Conte. "Generalized Stochastic Petri Nets: A Definition at the Net Level and Its Implications", *IEEE Transactions on Software Engineering*, Vol. 19, pp. 89–107, February 1993.

[25] G. Chiola, C. Anglano, J. Campos, J. M. Colom, and M. Silva. "Operational Analysis of Timed Petri Nets and Application to the Computation of Performance Bounds", In *Proceedings of the 5th International Workshop on Petri Nets and Performance Models*, pages 128–137, Toulouse, France, October 1993. IEEE-Computer Society Press.

[26] G. Chiola, S. Donatelli, G. Franceschinis, "GSPN versus SPN: What is the Actual Role of Immediate Transitions?", In *Proceedings of the 4rd International Workshop on Petri Nets and Performance Models*, pages 20–31, Melbourne, Australia, December 1991. IEEE-Computer Society Press.

[27] G. Ciardo, J. Muppala, and K.S. Trivedi. "SPNP: Stochastic Petri net package", In *Proc. 3rd Intern. Workshop on Petri Nets and Performance Models*, Kyoto, Japan, December 1989. IEEE-CS Press.

[28] J.B. Dugan, K.S. Trivedi, R.M. Geist, and V.F. Nicola. "Extended Stochastic Petri Nets: Applications and Analysis", In *Proc. PERFORMANCE '84*, Paris, France, December 1984.

[29] G. Florin and S. Natkin. "Les réseaux de Petri stochastiques", *Technique et Science Informatiques*, Vol. 4, February 1985.

[30] C. Hanen, A. Munier, "Cyclic scheduling on parallel processors: an overview", In *Scheduling Theory and Its Applications*, P. Chretienne et al. (Eds.), J. Wiley, to appear, 1995.

[31] J. Keilson, *Markov Chain Models / Rarity and Exponentiality*, Springer-Verlag, 1979.

[32] S. Kumar, P. R. Kumar, "Performance Bounds for Queueing Networks and Scheduling Policies", Technical Report, Coordinated Science Laboratory, Univ. of Illinois, 1992.

[33] P. R. Kumar, S. P. Meyn, "Stability of Queueing Networks and Scheduling Policies", Technical Report, Coordinated Science Laboratory, Univ. of Illinois, 1993.

[34] Y. Li, C. M. Woodside, "Complete Decomposition of Stochastic Petri Nets Representing Generalized Service Networks", *IEEE Transactions on Computers*, Vol. 44, pp. 577-592, 1995.

[35] M.K. Molloy. "Performance Analysis using Stochastic Petri Nets", *IEEE Transaction on Computers*, Vol. 31, pp.913–917, September 1982.

[36] T. Murata, "Petri Nets: Properties, Analysis and Applications", *Proc. of the IEEE*, Vol. 77, pp. 541-580, 1989.

[37] M. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins, Baltimore, Md. 1981.

[38] M. Reiser and H. Kobayashi, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms." *IBM J. Res. Dev.*, Vol. 19, pp. 283-294, 1975.

[39] M. Reiser and S. S. Lavenberg, "Mean-value analysis of closed multichain queueing networks." *J. ACM*, Vol. 27, pp. 313-322, 1980.

[40] S. Stidham, "A Last Word on $L = \lambda W$", *Oper. Res.*, Vol. 22, pp.417-421, 1974.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399