Tight Polynomial Bounds for Steady-State Performance of Marked Graphs

J. Campos*
Dpto. de Ingeniería Eléctrica e Informática
Universidad de Zaragoza, Spain.

G. Chiola†
Dipartimento di Informatica
Università di Torino, Italy.

J.M. Colom*
Dpto. de Ingeniería Eléctrica e Informática
Universidad de Zaragoza, Spain.

M. Silva*
Dpto. de Ingeniería Eléctrica e Informática
Universidad de Zaragoza, Spain.

Abstract

The problem of computing both upper and lower bounds for the steady-state performance of timed and stochastic Marked Graphs is studied. In particular, Linear Programming problems defined on the incidence matrix of the underlying Petri nets are used to compute tight (i.e., reachable) bounds for the throughput of transitions for live and bounded Marked Graphs with time associated with transitions. These bounds depend on the initial marking and the mean values of the delays but not on the probability distributions (thus including both the deterministic and the stochastic cases). Connections between results and techniques typical of qualitative and quantitative analysis of Petri models are stressed.

1 Introduction

One of the main problems in the actual use of timed and stochastic Petri net models for the performance evaluation of large systems is the explosion of the computational complexity of the analysis algorithms. Exact performance results are usually obtained from the numerical solution of a Markov chain, whose dimension is given by the size of the state space of the model. Simplified methods of computational complexity polynomial on the size of Petri net description would provide a striking break-through in this field, but results of this type are unlikely to be achievable for the computation of exact performances for general models.

In this paper we study the possibility of obtaining (upper and lower) *bounds* on the steady-state performance of Marked Graphs (MG), a well known subclass of Petri nets that allow only concurrency and synchronization but no choice. In particular we study the throughput of transitions, defined as the average number of firings per unit time. From this quantity, applying Little's formula it is possible to derive all average performance estimates of the model. Under these restrictions we will show results that can be computed in polynomial time on the size of the net model, and that depend only on the mean values and not on the higher moments of the probability distributions of the random variables that describe the timing of the system. Our approach differs from that of Molloy [Mol85] in that we analyze a net with a given initial marking instead of computing a limiting behaviour for increasing number of tokens. In a sense we have taken a complementary approach with respect to that of Bruell and Ghanta [BG85], that assumed exponentially distributed random delays and bounded the effect of synchronization. Our independence of the probability distribution can be viewed as a useful generalization of the performance results, since higher moments of the delays are usually unknown for real cases, and difficult to estimate and assess. Moreover we show that both upper and lower bounds, computed by means of proper Linear Programming problems, are tight, in the sense that for any MG model it is possible to define families of MG models with stochastic timings, such that the steady-state performances of the timed PN models are arbitrarily close to either bound.

In a Petri net there is an obvious relation between the concepts of steady-state behaviour and that of repeatable firing sequences: sequences of transitions that are repeatable only a finite number of times cannot contribute to the steady-state performance of the model. In the particular case of safe MGs, this relation can be made evident by looking at the occurrence nets that describe their behaviours [WT85].

Figure 1 depicts an example of a live and safe MG. It is easily seen that only sum and "max" operators are needed to compute the performance: indeed the actual cycle time in this example is the random variable $\gamma = \tau_1 + \max(\tau_2, \tau_3) + \tau_4$ (where τ_i denotes the firing delay of transition ti), therefore the average cycle time is $\Gamma =$

$$E[\tau_1] + E[\max(\tau_2, \tau_3)] + E[\tau_4] = \theta_1 + E[\max(\tau_2, \tau_3)] + \theta_4$$

(where θ_i denotes the average firing delay of transition ti). The idea is that of computing fast bounds based only on the knowledge of the first moments of

*This work has been supported by project PA86-0028 of the Spanish Comisión Interministerial de Ciencia y Tecnología (CICYT).

†This work was performed while G. Chiola was visiting the ETSII Zaragoza, with the financial support of the Caja de Ahorros de la Inmaculada de Zaragoza.

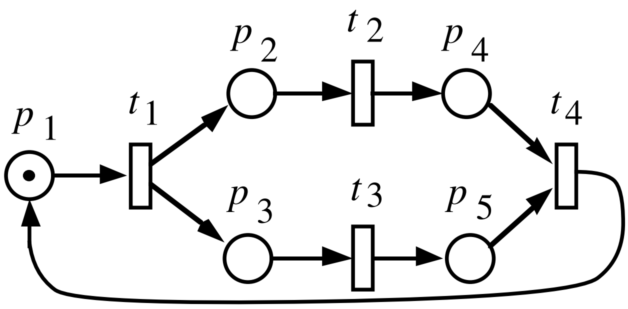


Figure 1: Example of a safe MG.

PDF: the sum is independent of the probability distribution (for linearity); since for non-negative variables $x_i \leq \max_i(x_i) \leq \sum_i x_i$, $E[\max_i(x_i)]$ can be bounded by $\max_i(E[x_i]) \leq E[\max_i(x_i)] \leq \sum_i E[x_i]$. Therefore for the net in Figure 1 we can write:

$$\theta_1 + \max(\theta_2, \theta_3) + \theta_4 \leq \Gamma \leq \theta_1 + \theta_2 + \theta_3 + \theta_4$$

The paper is organized as follows. Section 2 contains a discussion of the implications that the introduction of a timing semantics has on the behaviour of a Petri net model. In particular the concepts of enabling bound and weak ergodicity are defined. Sections 3 and 4 present the upper and lower bounds, respectively. Section 5 contains some concluding remarks and considerations on possible extensions of the work. In the Appendix the notation and some classical results for Petri nets are collected.

2 Stochastic interpretation of nets

In the original definition, Petri nets did not include the notion of time, and tried to model only the logical behaviour of systems by describing the causal relations existing between otherwise unrelated events. This approach showed its power in the specification and analysis of concurrent systems in a non-interleaved way, i.e. in a primitive way independent of the concept of time. Nevertheless the introduction of timing specification is essential if we want to use this class of models for an evaluation of the performance of distributed systems.

2.1 Timing and firing process

Since Petri nets are bipartite graphs, historically there have been two ways of introducing the concept of time in them, namely, associating a time interpretation with either places [Sif78] or transitions [Ram74]. Moreover, in the case of timed transition models, two different firing rules have been defined: single phase (atomic) firing, and three phase (start-firing with deletion of the input tokens, delay, and end-firing with creation of the output tokens). In summary, using timed place or timed transition models with three phase firing we can define a policy for conflict resolution independent of the time specification but we cannot model preemption; on the other hand, using timed transition models with single phase firing we can model preemption but we cannot define conflict resolution policies indepent of the timing specification (unless introducing the concept of immediate transitions [AMBC84] that adds one degree of freedom, but is still part of the timing specification).

In any case, from the analysis of the different choices, it follows that the only effect of these different timing interpretations on the performance evaluation of a model is due to the different implications that the choices have on the resolution of conflicts. Since in the context of this work we are considering a class of nets in which there are no conflicts, we do not have to fully choose one particular interpretation: we only say that we consider timed transition MGs.

2.2 Single versus multiple server semantics: enabling bound

Another possible source of confusion in the definition of the timed interpretation of a PN model is the concept of "degree of enabling" of a transition (or re-entrance). In the case of timing associated with places, it seems quite natural to define an unavailability time which is independent of the total number of tokens already present in the place, an this can be interpreted as an "infinite server" policy from the point of view of queueing theory. In the case of time associated with transitions, it is less obvious a-priori whether a transition enabled k times in a marking should work at conditional throughput 1 or k times that it would work in the case it was enabled only once. In the case of Stochastic PNs with exponentially distributed firing times associated with transitions, the usual implicit hypothesis is to have "single server" semantics (see, e.g., [FN85a], [Mol82]), and the case of "multiple server" is handled as a case of firing rate dependent on the marking; unfortunately this trick cannot work in the case of more general probability distributions, and in particular cannot be used in the case of deterministic timings. This is the reason why people working with deterministic timed transition PN prefer an infinite server semantics (see, e.g., [Zub85], [RP84], [HV87]). Of course an infinite server transition can always be constrained to a "k–server" behaviour by just reducing its enabling bound to k, as we will see later.

Therefore the infinite server semantics appears to be the most general one, and for this reason it will be adopted in this work. However this generality of the infinite server assumption will be payed in terms of complexity of the algorithms for the computation of performance bounds. Indeed the performance of a model with infinite server semantics does depend on the maximum degree of enabling of the transitions. For this reason we introduce here a concept of degree of enabling of a transition: the enabling bound $E(t)$. It allows us to generalize the classical concept of enabling of a transition (a generalization of the concept of liveness of a transition is presented in [CCS89]).

DEFINITION 2.1 *Enabling bound. Let $\langle \mathcal{N}, M_0 \rangle$ be a marked PN, $\forall t \in T$*

$$\begin{array}{rll} E(t) = & \max & k \\ & s.t. & \exists M \in R(\mathcal{N}, M_0) \ : \ M \geq kPRE[t] \end{array}$$

From the above definition it appears clear how it is possible to obtain a k–server transition from an infinite server one: adding one place that is both input and output (with multiplicity 1) for that transition and marking it with k tokens.

The definition above refers to a behavioural property that depends on the reachability graph of a PN. Since we are looking for computational techniques at the structural level, we can also introduce the structural counterpart of the concept. Structural net theory has been developed from two complementary points of view: graph theory [Bes86] and mathematical programming (or more specifically linear programming and linear algebra) [SC87]. Let us introduce our structural definition from the mathematical programming point of view; essentially in this case the reachability condition is substituted by the weaker (linear) constraint that markings satisfy the net state equation: $M = M_0 + C\vec{\sigma}$, with $M, \vec{\sigma} \geq 0$.

DEFINITION 2.2 *Structural enabling bound:* $\forall t \in T$

$$\begin{array}{rll} SE(t) = & \max & k \\ & s.t. & \exists \vec{\sigma} \geq 0, M = M_0 + C\vec{\sigma} \geq kPRE[t] \end{array}$$

Note that, the definition of structural enabling bound reduces to the formulation of a Linear Programming problem [Mur83] using matrix C (the incidence matrix of the PN). For MGs the behavioural concepts always collapse into the structural ones. In fact, according to Theorem A.1, for any MG $M \in R(M_0)$ iff $M = M_0 + C\vec{\sigma} \wedge \vec{\sigma} \geq 0$. This allows an efficient computation of enabling bound based on the Linear Programming problem that characterizes the structural enabling bound.

In case of non-strongly connected MGs it is possible to obtain $SE(t) = \infty$ for some transition t; this creates no harm: because of the assumption of an infinite server semantics, it only implies that the timing of that transition does not affect the steady-state performance of the model.

2.3 Ergodicity, measurability, bounds

In order to be able to speak about steady-state performance we have to assume that some kind of "average behaviour" can be estimated on the long run of the system we are studying. The usual assumption in this case is that the system models must be (strongly) *ergodic* (see definitions of ergodicity in [FN85b]). This assumption is very strong and difficult to verify in general; moreover, it creates problems when we want to include the deterministic case as a special case of a stochastic model (see [CCS89]). Thus we introduce the concept of *weak ergodicity* that allows the estimation of long run performances also in the case of deterministic models.

DEFINITION 2.3 *The marking process of a stochastic marked net is weakly ergodic (or measurable in long run) iff the following limit exists:*

$$\overline{M} \stackrel{\text{def}}{=} \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau M(s)ds < \vec{\infty}$$

The firing process of a stochastic marked net is weakly ergodic (or measurable in long run) iff the following limit exists:

$$\vec{\sigma}^* \stackrel{\text{def}}{=} \lim_{\tau \to \infty} \frac{\vec{\sigma}(\tau)}{\tau} < \vec{\infty}$$

According to the above definitions and the properties listed in the Appendix, strong connectivity and boundedness have equivalent meaning for live MGs. From Properties A.2, A.3, the following result follows:

COROLLARY 2.1 *Strongly connected live MGs have weakly ergodic marking process.*

3 Upper bounds for stochastic strongly connected MGs

In this and the next Sections, performance bounds for strongly connected (and thus structurally bounded) MGs are presented. Strong connectivity of a graph is a well known problem of polynomial time complexity.

3.1 Upper bound for the steady state throughput

Let us take into account just the first moments of the probability density functions associated with transitions. In the following, let θ_i be the mean value of the random variable associated with the firing of transition t_i, and D the diagonal matrix with elements θ_i, $i = 1, ..., m$.

The limit expected firing count vector per time unit is

$$\vec{\sigma}^* = \lim_{\tau \to \infty} \frac{\vec{\sigma}(\tau)}{\tau} \tag{1}$$

and the mean time between two consecutive firings of a selected transition, t_i,

$$\Gamma_i = \frac{1}{\vec{\sigma}^*_i} \tag{2}$$

Then the components of $Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i$ (where $\vec{\sigma}^*$ has been normalized for having the i^{th} component equal 1) represent the product of the number of tokens reserved for firing the transitions and the mean length of time that these tokens reside in each place between two consecutive firings of t_i.

Let $\overline{M}$ be the limit vector of the average number of tokens in all places (i.e. $\overline{M} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau M(s)ds$). Then, provided that the previous limit exists, $\overline{M} \cdot \Gamma_i$ is the vector of products of the mean number of tokens and the length of one cycle and we have:

$$\overline{M} \cdot \Gamma_i \geq Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i \tag{3}$$

From this inequality, the minimum cycle time associated with transition t_i, Γ_i^{min}, can be derived. We take into account that Γ_i^{min} must be such that inequality (3) holds and for some place p_j the equality is reached:

$$\Gamma_i^{min} = \frac{Pre(j) \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i}{\overline{M}(j)} \tag{4}$$

Since the vector $\overline{M}$ is unknown, (4) cannot be solved. Making the product with a P-semiflow Y for any reachable marking M:

$$Y^T \cdot M_0 = Y^T \cdot M = Y^T \cdot \overline{M} \tag{5}$$

Now, from (3) and (5):

$$Y^T \cdot M_0 \cdot \Gamma_i \geq Y^T \cdot Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i \tag{6}$$

And the minimum cycle time in steady state is:

$$\Gamma_i^{min} = \max_{Y \in \{\text{P-semiflow}\}} \frac{Y^T \cdot Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i}{Y^T \cdot M_0} \tag{7}$$

Of course, an upper bound for the throughput of t_i is $\frac{1}{\Gamma_i^{min}}$.

Let us formulate the previous lower bound for the cycle time in terms of a particular class of optimization problems that we introduce now.

DEFINITION 3.1 *A fractional programming problem is an optimization problem of the form* [Mur83]*:*

$$\begin{array}{lll} \max & f(x) = & \dfrac{c^T \cdot x + \alpha}{d^T \cdot x + \beta} \\ \textit{subject to} & & A \cdot x = b; \quad x \geq 0 \\ & & A \in R^{v \times u}; \quad b \in R^v \\ & & c, d \in R^u; \quad \alpha, \beta \in R \end{array}$$

A fractional programming problem is said to be homogeneous if coefficients α and β are equal to zero.

THEOREM 3.1 *For any net, the minimum cycle time associated with transition t_i can be computed by the following homogeneous fractional programming problem:*

$$\begin{array}{rll} \Gamma_i^{min} = & \max & \dfrac{Y^T \cdot Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i}{Y^T \cdot M_0} \\ & s.t. & Y^T \cdot C = 0 \\ & & Y \geq 0, \quad \vec{1}^T \cdot Y > 0 \end{array} \tag{FPP1}$$

PROOF: Just notice that the optimum can be always reached with an elementary P-semiflow. If Y^* gives us the optimal solution, Z^*, and it is not an elementary P-semiflow, then

$$Y^* = \sum_{i=1}^{k} (a_i Y_i) \tag{8}$$

with Y_i elementary P-semiflows and $a_i > 0, \sum_{i=1}^{k} a_i = 1$ (i.e. $Z^* = \sum_{i=1}^{k}(a_i Y_i^T \cdot Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i)$). Then $a_i = \frac{1}{k}\ \forall i$, all $Y_i^T \cdot Pre \cdot D \cdot \vec{\sigma}^* \cdot \Gamma_i$ must be equal (if not, the maximum of them would give a larger value than Z^*) and all Y_i are optimal solutions of the problem.
Q.E.D.

Because for MGs $\vec{\sigma}^* \cdot \Gamma_i = \vec{1}$, writing $\theta = D \cdot \vec{1}$ the following can be directly stated:

COROLLARY 3.1 *The minimum cycle time for strongly connected MGs can be obtained by solving the following fractional programming problem:*

$$\begin{array}{rll} \Gamma^{min} = & \max & \dfrac{Y^T \cdot Pre \cdot \theta}{Y^T \cdot M_0} \\ & s.t. & Y^T \cdot C = 0 \\ & & Y \geq 0, \quad \vec{1}^T \cdot Y > 0 \end{array} \tag{FPP2}$$

THEOREM 3.2 *The minimum cycle time for live strongly connected MGs can be obtained by solving the following linear programming problem:*

$$\begin{array}{rll} \Gamma^{min} = & \max & Y^T \cdot Pre \cdot \theta \\ & s.t. & Y^T \cdot C = 0, Y^T \cdot M_0 = 1 \\ & & Y \geq 0 \end{array} \tag{LPP1}$$

PROOF: The problem (FPP2) can be rewritten into a linear problem just taking into account that if we write:

$$\begin{array}{rll} \Gamma^{min} = & \max & \dfrac{Y^T \cdot Pre \cdot \theta}{q} \\ & \text{s.t.} & Y^T \cdot C = 0, \quad q = Y^T \cdot M_0 \\ & & Y \geq 0, \quad \vec{1}^T \cdot Y > 0 \end{array}$$

Then, because $Y^T \cdot M_0 > 0$ (guaranteed for live MGs), we can change $\frac{Y}{q}$ by Y and obtain the problem (LPP1) in which $\vec{1}^T \cdot Y > 0$ is redundant ($Y^T \cdot M_0 = 1 \Longrightarrow \vec{1}^T \cdot Y > 0$) and can be removed.
Q.E.D.

It is well known that the simplex method for the solution of linear programming problems gives good results in practice, even if it has exponential theoretical worst case complexity. In any case an algorithm of polynomial theoretical worst case complexity can be found in [Kar84].

Theorem 3.2 shows that the problem of finding an upper bound for the steady-state throughput in a strongly connected stochastic MG can be solved looking at the cycle times associated with each P-semiflow (cycles for MGs) of the net, considered in isolation. These cycle times can be computed making the summation of the average firing times of all the transitions involved in the P-semiflow, and dividing by the number of tokens present in it.

The above bound is the same that has been obtained for strongly connected deterministic MGs by other authors (see for example [Ram74], [RH80], [Mur85]), but here it is considered in a practical LPP form. For these nets, the reachability of this bound has been shown ([Ram74], [RH80]). Since deterministic timing is just a particular case of stochastic timing, the reachability of the bound is assured for our purposes. Even more, the next result shows that the previous bound cannot be improved only on the base of the knowledge of the coefficients of variation for the transition firing times.

THEOREM 3.3 *For strongly connected MGs with arbitrary values of mean and variance for transition firing times, the bound for the cycle time obtained from (FPP2) cannot be improved.*

PROOF: Let σ_i^2 the arbitrary variance associated with transition t_i. We know from [Ram74] that for deterministic timing the bound is reached. Let θ_i be the average firing time associated with transition t_i. Then there exists a sequence of families of m distributions with means θ_i and variances σ_i^2, $i = 1, ..., m$, for which the cycle time tends to the one obtained from (FPP2).

Consider the family (for varying values of the parameter $0 \leq \alpha < 1$):

$$X_{\theta_i,\sigma_i}(\alpha) = \begin{cases} \theta_i \alpha & \text{with probability } 1-\epsilon_i \\ \theta_i(\alpha + \frac{1-\alpha}{\epsilon_i}) & \text{with probability } \epsilon_i \end{cases}$$

where

$$\epsilon_i = \frac{\theta_i^2(1-\alpha)^2}{\theta_i^2(1-\alpha)^2 + \sigma_i^2}$$

Now, taking α closer to 1 for the previous family of random variables, the cycle time tends to the bound given by (FPP2). This is because only "max" and sum operators are needed to compute the cycle time and the previous family of random variables behaves closer to deterministic variables when α tends to 1, i.e.

$$\lim_{\alpha \to 1} E[\max(X_{\theta_i,\sigma_i}(\alpha), X_{\theta_j,\sigma_j}(\alpha))] = \max(\theta_i, \theta_j)$$

and, of course, $\forall\ 0 \leq \alpha < 1$

$$E[X_{\theta_i,\sigma_i}(\alpha) + X_{\theta_j,\sigma_j}(\alpha)] = \theta_i + \theta_j$$

Q.E.D.

A polynomial computation of the minimal cycle time for deterministic timed strongly connected MGs was proposed in [Mag84], solving the following linear programming problem:

$$\begin{aligned} \Gamma^{min} = \quad & \min \gamma \\ & \text{s.t.} \quad -C \cdot z + \gamma M_0 \geq Post \cdot \theta \qquad (LPP2) \\ & \qquad \gamma \geq 0, \quad z \geq 0 \end{aligned}$$

To investigate the relationship between (LPP1) and (LPP2) let us consider the *dual* problem of (LPP2):

$$\begin{aligned} \Gamma^{min} = \quad & \max \quad Y^T \cdot Post \cdot \theta \\ & \text{s.t.} \quad Y^T \cdot C \leq 0, \quad Y^T \cdot M_0 \leq 1 \qquad (DPP2) \\ & \qquad Y \geq 0 \end{aligned}$$

Since MGs are consistent nets, the restriction $Y^T \cdot C \leq 0$ of (DPP2) can be substituted by $Y^T \cdot C = 0$ (i.e. the restriction of (LPP1)). For all Y such that $Y^T \cdot C = 0$, $Y^T \cdot Post = Y^T \cdot Pre$. Now, for live nets, $\forall Y \in N^n$, $Y \neq 0$ such that $Y^T \cdot C = 0$ then $Y^T \cdot M_0 \geq 1$. Thus the restriction $Y^T \cdot M_0 \leq 1$ of (DPP2) can be substituted by $Y^T \cdot M_0 = 1$ for live nets (i.e. the restriction of (LPP1)).

Then for live strongly connected MGs, the general problem (FPP1) takes the linear form (LPP1) which is equivalent to (LPP2) formulated in [Mag84] for deterministic systems. For non-live nets the problem (FPP1) has unbounded optimal solution (see Theorem 3.4). This can be easily understood since non live nets have a null throughput (infinite minimal cycle time).

3.2 Interpretation and derived results

Linear programming problems give an easy way to derive results and interpret them. Just looking at the objective function of the problem (FPP2) the following monotonicity property is obtained: the optimum value for the minimum cycle time decreases if θ decreases or if M_0 increases.

PROPERTY 3.1 *Let $\mathcal{N}$ be a strongly connected MG and θ the mean times vector.*

i) *For a fixed θ, if $M_0' \geq M_0$ (i.e. increasing the number of initial resources) then the throughput upper bound of $\langle \mathcal{N}, M_0', \theta \rangle$ is larger than or equal to the one of $\langle \mathcal{N}, M_0, \theta \rangle$ (i.e. $\Gamma^{min'} \leq \Gamma^{min}$).*

ii) *For a fixed M_0, if $\theta' \leq \theta$ (i.e. for faster resources) then the throughput upper bound of $\langle \mathcal{N}, M_0', \theta' \rangle$ is larger than or equal to the one of $\langle \mathcal{N}, M_0, \theta \rangle$ (i.e. $\Gamma^{min'} \leq \Gamma^{min}$).*

The next property is strongly related to the reversibility of live MGs.

LEMMA 3.1 *For live strongly connected MGs, the bound obtained with the problem (LPP1) does not change for any reachable marking.*

PROOF: Let us consider the minimum cycle time for a marking $M = M_0 + C \cdot \vec{\sigma}$ in terms of a linear programming problem:

$$\begin{aligned} \Gamma^{min} = \quad & \max \quad Y^T \cdot Pre \cdot \theta \\ & \text{s.t.} \quad Y^T \cdot C = 0, \quad Y \geq 0 \\ & \qquad M = M_0 + C \cdot \vec{\sigma} \\ & \qquad Y^T \cdot M = 1 \\ & \qquad M \geq 0, \quad \vec{\sigma} \geq 0 \end{aligned}$$

Since $Y^T \cdot M = Y^T \cdot M_0$ this problem is equivalent to:

$$\begin{array}{rll} \Gamma^{min} = & \max & Y^T \cdot Pre \cdot \theta \\ & \text{s.t.} & Y^T \cdot C = 0, \quad Y \geq 0 \\ & & M = M_0 + C \cdot \vec{\sigma} \\ & & Y^T \cdot M_0 = 1 \\ & & M \geq 0, \quad \vec{\sigma} \geq 0 \end{array}$$

Since the restrictions $M = M_0 + C \cdot \vec{\sigma}$, $M \geq 0$ and $\vec{\sigma} \geq 0$ do not affect the solution, they can be removed without changing the optimum of this problem with respect to the one of (LPP1).
Q.E.D.

The next is a result on the complexity of the verification of liveness for MGs. It has been recently pointed out in [ES89] by using quite different arguments and techniques. Here liveness is characterized by the finiteness of the cycle time.

THEOREM 3.4 *Liveness of a strongly connected MG can be decided in polynomial time.*

PROOF: We know that for strongly connected MGs, liveness and deadlock-freeness coincide. Then for deciding liveness of a strongly connected MG it is enough to study the finiteness of the optimal value of (LPP1).

For strongly connected MGs, the optimal value of (LPP1) is a lower bound for the cycle time. If this optimal value is infinite the cycle time is unbounded so the net is non live. If the optimal value of (LPP1) is finite, since it is reachable for some (deterministic [Ram74] as well as stochastic) timing (cfr. Theorem 3.3), the net must be live.
Q.E.D.

COROLLARY 3.2 *The problem (LPP1) has unbounded solution iff $\exists Y \geq 0, Y \neq 0$ such that $Y^T \cdot M_0 = 0$ and $Y^T \cdot C = 0$*

This result has the following topological interpretation: the problem (LPP1) has unbounded solution iff there exists an unmarked circuit in the strongly connected MG.

4 Lower bounds for strongly connected MGs

4.1 Basic result for 1–live MG

A trivial lower bound in steady-state performance for a live PN with a unique repetitive firing count vector [CCS89] is of course given by the sum of the firing times of all the transitions weighted by the firing count vector itself. Since the net is live all transition must be fireable, and the sum of all firing times multiplied by the number of occurrences of each transition in the (unique) average cycle of the model corresponds to any complete sequentialization of all the activities represented in the model. This lower bound is always reached in an MG consisting of a single loop of transitions and containing a single token in one of the places, independently of the higher moments of the PDFs (this observation can be trivially confirmed by the computation of the upper bound, which in this case gives the same value).

To improve this trivial lower bound let us first consider the case of 1–enabled MGs (i.e. strongly connected MG in which $E(t) = 1$ for all transitions t). Of course live and safe MGs are guaranteed to be 1–enabled, but the result that we are going to present apply to more general cases. If we specify only the mean values of the transition firing times and not the higher moments, we may always find stochastic models whose steady-state throughput is arbitrarily close to the trivial lower bound, independently of the topology of the MG (only provided that it is 1–enabled). Let us give a formal proof of this (somewhat counter-intuitive) result.

LEMMA 4.1 $\forall \mu \geq 0$ *there exists a family of random variables $x^i_\mu(\epsilon)$ with expected value $E[x^i_\mu(\epsilon)] = \mu$ $\forall 0 < \epsilon \leq 1$, $\forall i \geq 0$ and with coefficient of variation ranging from 0 to ∞ for decreasing values of $0 < \epsilon \leq 1$, and fixed values of $i > 0$. This family is defined as:*

$$x^i_\mu(\epsilon) = \begin{cases} 0 & \textit{with probability } 1 - \epsilon^i \\ \frac{\mu}{\epsilon^i} & \textit{with probability } \epsilon^i \end{cases}$$

PROOF: $E[x^i_\mu(\epsilon)] = \mu$ is straightforward to compute. $E[(x^i_\mu(\epsilon))^2] = \frac{\mu^2}{\epsilon^i}$ implies that the coefficient of variation is 0 for $\epsilon = 1$, and that it tends to ∞ as $\epsilon \to 0$ provided that $i > 0$.

THEOREM 4.1 *For any live and safe MG with a specification of the mean firing times θ_j for each $t_j \in T$ it is possible to assign PDFs to the transition firing times such that the average cycle time is $\Gamma = \sum_j \theta_j - O(\epsilon)$ $\forall 0 < \epsilon \leq 1$, independently of the topology of the net (and thus independently of the potential maximum degree of parallelism intrinsic in the MG). (We use here the notation $O(f(x))$ to indicate any function $g(x)$ such that $\lim_{x \to 0} \frac{g(x)}{f(x)} \leq k \in I\!R$.)*

PROOF: by construction, we will show that the association of the family of random variables $x^{j-1}_{\theta_j}(\epsilon)$ with each transition $t_j \in T$ yields exactly the cycle time Γ claimed by the theorem. To give the proof we will consider a sequence of models ordered by the index of transitions, in which the q-th model of the sequence has transitions $t_1, t_2, \ldots t_q$ timed with the random variables $x^{j-1}_{\theta_j}(\epsilon)$, and all other transitions immediate (firing in zero time); the $|T|$-th model in the sequence represents an example of reachability of the lower bound, independent of the net topology. Now we will prove by induction that the q-th model in the sequence has a cycle time $\Gamma_q = \sum_{j=1}^q \theta_j - O(\epsilon)$

Base: $q = 1$: trivial since the repetitive cycle that constitute the steady–state behaviour of the MG contains only one (single-server) deterministic transition with average firing time $\Gamma_1 = \theta_1$.

Induction step: $q > 1$: taking the limit $\epsilon \to 0$, the newly timed transition t_q will fire most of the times with time zero, thus normally not disturbing the behaviour of the other timed transition, and not contributing to the computation of the cycle time, that will be just $\Gamma_{q-1} = \sum_{j=1}^{q-1} \theta_j - O(\epsilon)$ (as in the case of model $q-1$) with probability $1 - \epsilon^{q-1}$. On the other hand, the newly timed transition has a (very small) probability ϵ^{q-1} of delaying its firing of a time $\frac{\theta_q}{\epsilon^{q-1}}$, which is at least order of $\frac{1}{\epsilon}$ bigger than any other firing time in the cycle, so that in this case all other transitions will wait for the firing of t_q after having completed their possible current firings in a time which is $O(\epsilon)$ lower than the firing time of t_q itself (i.e., $\frac{\theta_q}{\epsilon^{q-1}} = \frac{\Gamma_{q-1}}{O(\epsilon)}$). Therefore we obtain that $\Gamma_q = (1-\epsilon^{q-1})\Gamma_{q-1} + \epsilon^{q-1}(\frac{\theta_q}{\epsilon^{q-1}} - O(\epsilon)) = \sum_{j=1}^{q} \theta_j - O(\epsilon)$. Q.E.D.

Until now we have shown that the trivial sum of the average firing times of all transitions in the net constitutes a tight (reachable) lower bound for the performance of a live and safe MG (or more generally of a 1–enabled strongly connected MG, but otherwise independently of the topology) in which only the mean values and neither the PDFs nor the higher moments are specified for the transition firing times. Let us now extend this result to the more general case of k–enabled strongly connected MGs, and see whether we can derive some reachable lower bound.

An intuitive idea could be to try to derive a lower bound for MG containing transitions with enabling bound $k \geq 1$ (remember that for MGs $E(t) = SE(t)$) by taking the algorithm used for the computation of the upper bound in the case of non-safe MG, and substitute in it the "max" operator with the sum of the firing times of all transitions involved. After some manipulation to avoid counting more than once the contribution of the same transition, one can arrive at the formulation of the following value for the maximum cycle time.

THEOREM 4.2 *For any live and bounded MG with a specification of the mean firing times θ_j for each $t_j \in T$ it is not possible to assign PDFs to the transition firing times such that the average cycle time is greater than*

$$\Gamma_{max} = \sum_j \frac{\theta_j}{E(t_j)}$$

independently of the topology of the net (and thus independently of the potential maximum degree of parallelism intrinsic in the MG).

PROOF: we give in the following a proof of this results by constructing some auxiliary MG models. These auxiliary models are obtained by adding structural constraints on the firing of the transitions with respect to the original one, in such a way that the performance may only remain the same or decrease, and than verify that their maximum cycle time is $\Gamma \leq \Gamma_{max}$.

4.1.1 Construction to demonstrate the Theorem

LEMMA 4.2 *Any strongly connected MG can be constraint to contain a main cycle including all transitions, without changing their enabling bound. This main cycle contains a number of tokens equal to the maximum of the enabling bounds among all transitions. In addition there are other minor cycles that preserve the enabling bounds for transitions with bound lower than the maximum. The idea behind this constrain is to introduce a structural sequentialization between all transitions, thus potentially reducing the degree of parallelism between the activities modelled by the transitions. In other words from the partial order given by the initial MG structure we try to derive a total order without changing the enabling bound.*

PROOF: To construct an MG of the desired form we can apply the following iterative procedure that interleaves two non-disjoint cycles into a single one. Since the MG is strongly connected each node belongs to at least one cycle; moreover, since the original MG is finite and each cycle cannot contain the same node more than once, this cycle interleaving procedure must terminate after a finite number of iterations. To reduce the number of cycles, implicit places created after each iteration can be removed. The iteration step is the following:

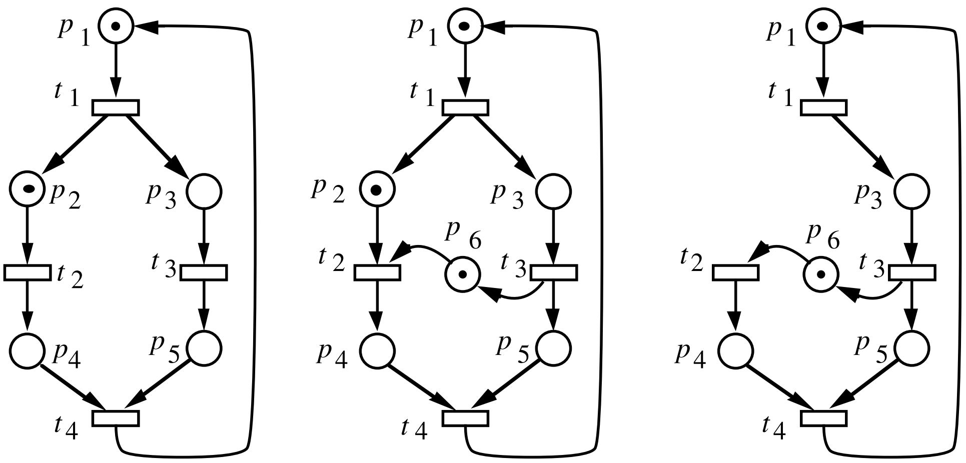
1. take two arbitrary non-disjoint cycles (unless the MG already contains a main cycle including all nodes, there always exists such a pair of cycles because the MG is strongly connected);

2. combine them in a single cycle in such a way that the partial order among transitions given by the two original cycles is substituted by a compatible but otherwise arbitrary total order. This combination can be obtained by adding new places that are connected as input for a transition of one cycle and output for a transition of the other cycle that we decide must follow in the sequence determined by the new cycle we are creating;

3. mark the new places added in such a way that the new cycle contains the same number of tokens as the maximum of the number of tokens in the two original cycles.

Consider as an example of the application of this iterative step the net depicted in Figure 2a. This net contains only two cycles, namely t1,t2,t4, and t1,t3,t4; we can then add either the cycle t1,t2,t3,t4 or t1,t3,t2,t4; Figure 2b depicts the resulting net in case we choose to add the second cycle. In this case only place p6 (from t3 to t2) needs to be added to obtain the longer cycle, and it should be marked with one token, so that the new cycle comprising places p1,p3,p6,p4 contains two tokens, as the original cycle p1,p2,p4 (while the other original cycle p1,p3,p5 contained only one).

The above procedure is applied iteratively until all transitions are constrained into a single main cycle. In



a) Original net with t_2 and t_3 concurrent.

b) Transformed net: t_2 and t_3 sequentialized and p_2 made implicit.

c) Elimination of implicit places. Main loop: p_1, p_3, p_6, p_4. Minor cycle: p_1, p_3, p_5.

Figure 2: Example of structural sequentialization.

our example, we need not iterate the procedure since we already have obtained a cycle containing all transitions of the MG. At this point we can identify and eliminate the implicit places that have been created during the cycles interleaving procedure. We obtain then an MG composed by one main cycle containing $N_M = \max_{t \in T} E(t)$ tokens that connects all transitions, and a certain number of minor cycles containing less tokens than N_M that maintain the enabling bound of the other transitions. In our example we can easily see that place p2 become implicit in Figure 2b, so that it can be eliminated, finally leading ourselves to the MG depicted in Figure 2c.
Q.E.D.

It is evident that the MG transformed by applying the above Lemma has a cycle time which is greater than or equal to the cycle time of the original one, since some additional constraints have been added to the enabling of transitions: thus the cycle time of the transformed MG is a lower bound for the performance of the original one. Now if $N_M = 1$ in the above lemma, we re-find the lower bound of Theorem 4.1. In the case of $N_M > 1$ we can show that the cycle time of the transformed net cannot exceed Γ_{max} as follows.

PROOF for Theorem 4.2: without loss of generality, assume that transitions in the net resulting from the application of Lemma 4.2 are partitioned in two classes S_2 and S_1, with enabling bounds $K_2 = N_M > 1$ and $K_1 < N_M$, respectively (the proof is easily extended to the case of more than two classes). Construct a new model containing only K_1 tokens in the main cycle; at this point all transitions behave as K_1–servers, so that the cycle time is given by the sum of the firing times of all transitions, divided by the total number of customers in the main loop K_1; moreover the delay time for the transitions belonging to class S_1 is simply given by $D_1 = \sum_{t_j \in S_1} \theta_j$. Now if we increase the number of tokens in the main loop from K_1 to K_2 the delay time of S_1 cannot increase, so that the contribution of S_1 to the cycle time cannot exceed D_1 for each of the first K_1 tokens. Under the hypothesis that the throughput of the system is given by the inverse of Γ_{max} (i.e., assuming $X = \frac{1}{\Gamma_{max}}$), the average number of tokens of the main loop computed using Little's formula cannot exceed $N_1 = XD_1$, therefore the average number of tokens available to fire transitions in S_2 cannot be lower than

$$N_2 = K_2 - N_1 = K_2 \frac{\frac{K_2 - K_1}{K_1} \sum_{t_j \in S_1} \theta_j + \sum_{t_j \in S_2} \theta_j}{\sum_{t_j \in S_2} \theta_j + \frac{K_2}{K_1} \sum_{t_j \in S_1} \theta_j}$$

On the other hand, we need only

$$N_2 = XD_2 = K_2 \frac{D_2}{\sum_{t_j \in S_2} \theta_j + \frac{K_2}{K_1} \sum_{t_j \in S_1} \theta_j}$$

tokens to sustain throughput X in subnet S_2, so that we are assuming a delay in S_2

$$D2 \leq \frac{K_2 - K_1}{K_1} \sum_{t_j \in S_1} \theta_j + \sum_{t_j \in S_2} \theta_j$$

Now we claim that this is the actual maximum delay because the first K_1 tokens can proceed at the maximum speed in the whole net, thus experiencing only delay $\sum_{t_j \in S_2} \theta_j$ in subnet S2, while the remaining $K_2 - K_1$ tokens can also queue up for travelling through S_1, thus experiencing an additional delay of $\frac{1}{K_1} \sum_{t_j \in S_1} \theta_j$ each.
Q.E.D.

4.2 Reachability of the lower bound

The lower bound in performance given by the computation of Γ_{max} as defined in Theorem 4.2 can be shown to be reachable for any MG topology and for some assignement of PDF to the firing delay of transitions, exploiting the reachability of the trivial bound shown in Theorem 4.1 for 1–enabled MGs.

THEOREM 4.3 *For any strongly connected MG with a specification of the mean firing times θ_j for each $t_j \in T$, and for all $0 < \epsilon \leq 1$, it is possible to assign PDFs to the transition firing times such that the average cycle time is:*

$$\Gamma_{max} = \sum_j \frac{\theta_j}{E(t_j)} - O(\epsilon)$$

independently of the topology of the net (and thus independently of the potential maximum degree of parallelism intrinsic in the MG).

PROOF: by construction, in a very similar way than in the case of Theorem 4.1. The only technical difference is that now, without any loss of generality, we assume first of all to enumerate transitions in non-increasing order of enabling bound, i.e., rename the transitions in such a way that $\forall t_i, t_j \in T,\ i > j \implies E(t_i) \leq E(t_j)$. Then, as in the case of Theorem 4.1, we will show that the association of the family of random variables $x_{\theta_j}^{j-1}(\epsilon)$ with each transition $t_j \in T$ yields exactly the cycle time Γ_{max} claimed by the theorem. To give the proof we will consider a sequence of models ordered by the index of transitions, in which the q–th model of the sequence has transitions $t_1, t_2, \ldots t_q$ timed with the random variables $x_{\theta_j}^{j-1}(\epsilon)$, and all other transitions immediate (firing in zero time); the $|T|$-th model in the sequence represents the resulting model that is expected to provide the example of reachability of the lower bound. Now we will prove by induction that the q-th model in the sequence has a cycle time

$$\Gamma_q = \sum_{j=1}^{q} \frac{\theta_j}{E(t_j)} - O(\epsilon)$$

Base: $q = 1$: trivial since the repetitive cycle that constitute the steady–state behaviour of the MG contains only one ($E(t_1)$–server) deterministic transition with average firing time $\Gamma_1 = \theta_1 / E(t_1)$.

Induction step: $q > 1$: taking the limit $\epsilon \to 0$, each server of the newly timed transition t_q will fire most of the times with time zero, thus normally not disturbing the behaviour of the other timed transition, and not contributing to the computation of the cycle time, that will be just $\Gamma_q = \sum_{j=1}^{q-1} \frac{\theta_j}{E(t_j)} - O(\epsilon)$ (as in the case of model $q-1$) with probability $1 - \epsilon^{q-1}$. On the other hand, each of the servers of the newly timed transition has a (very small) probability ϵ^{q-1} of delaying its firing of a time $\frac{\theta_q}{\epsilon^{q-1}}$, which is at least order of $\frac{1}{\epsilon}$ bigger than any other firing time in the cycle. Now if $E(t_q) = 1$, then the proof is completed, since also $\forall j > q\ E(t_j) = 1$ by hypothesis, and we reduce to the induction step of the proof of Theorem 4.1. Instead if $E(t_q) > 1$ then we can consider $E(t_q)$ consecutive firings of t_q, and compute the average firing time as the total time to fire $E(t_q)$ times the transition, divided by $E(t_q)$. Now if we consider m consecutive firings of instances of transition t_q we obtain an average delay:

$$\sum_{j=m-1}^{0} (1 - \epsilon^{q-1})^j \epsilon^{(q-1)(m-j)} \frac{(m-j)\theta_q}{\epsilon^{(q-1)}} = \theta_q (1 + O(\epsilon))$$

Therefore the average cycle time of the q–th model will be

$$\Gamma_q = (1 - O(\epsilon^{q-1}))\Gamma_{q-1} + \frac{\theta_q}{E(t_q)}(1 + O(\epsilon)) = \sum_{j=1}^{q} \frac{\theta_j}{E(t_j)} - O($$

Q.E.D.

4.3 A polynomial algorithm to compute the lower bound

First of all we recall that in the case of live MGs the enabling bound equals the structural enabling bound for each transition; thus we present a characterization of the problem of the determination of the structural enabling bound in terms of a Linear Programming problem, which is known to be solvable in polynomial time.

For any transition $t \in T$, the computation of the structural enabling bound $SE(t)$ can be formulated in terms of the following LPP:

$$\begin{aligned} SE(t) = \quad & \max \quad \alpha \\ & s.t. \quad M = M_0 + C\vec{\sigma} \\ & \qquad\ \ M \geq \alpha PRE[t] \\ & \qquad\ \ M \geq 0, \vec{\sigma} \geq 0 \end{aligned}$$

by definition. Then we can observe that the vector M is redundant in the system of linear inequalities, so that we can remove it, obtaining:

$$\begin{aligned} SE(t) = \quad & \max \quad \alpha \\ & s.t. \quad M_0 + C\vec{\sigma} \geq \alpha PRE[t] \\ & \qquad\ \ M_0 + C\vec{\sigma} \geq 0, \sigma \geq 0 \end{aligned}$$

Alternatively, we can switch to the dual LPP:

$$\begin{aligned} SE(t) = \quad & \min \quad Y^T M_0 \\ & s.t. \quad Y^T C \leq 0 \\ & \qquad\ \ Y^T PRE[t] = 1 \\ & \qquad\ \ Y^T \geq 0 \end{aligned}$$

Now we can recall that strongly connected MGs are consistent nets with a single minimal T-semiflow which is the vector $\vec{1}$, so that the constraint $\vec{\sigma} \geq 0$ can be relaxed in the primal problem. The effect on the dual problem of this relaxation is the transformation of the first constraint into $Y^T C = 0$. In other words, the dual problem for the computation of $SE(t)$ can be rewritten as follows:

$$\begin{aligned} SE(t) = \quad & \min \quad Y^T M_0 \\ & s.t. \quad Y^T C = 0 \\ & \qquad\ \ Y^T PRE[t] = 1 \\ & \qquad\ \ Y^T \geq 0 \end{aligned}$$

This LPP is less complex to solve with the simplex algorithm than the original dual problem because it involves the introduction of fewer slack variables.

Because of the minimization requirement, the optimum of the objective function is always achieved with elementary P-semiflows Y. In case of MGs, these elementary P-semiflows can only be elementary cycles, so that we can give the following interpretation of the dual LPP in net terms: the enabling bound for a transition t of a strongly connected MG is given by the minimum number of tokens contained in any cycle of places containing transition t. In a non-strongly connected MG there can be no such cycle, so that this number can be infinite.

As final remarks we can state the following:

a) Liveness for a strongly connected MG can be a byproduct of a more general (polynomial complexity) computation:

$$\langle \mathcal{N}, M_0 \rangle \text{ is a live MG } \iff \forall t \in T \quad SE(t) > 0.$$

b) If the MG is known to be live for M_0, and $\exists t \in T$ such that $SE(t) = 1$, then $\forall t' \in T$ belonging to the same cycle denoted by Y in the corresponding LPP, $SE(t') = 1$.

5 Conclusions

In this paper we have addressed the problem of computing upper and lower bounds for the throughput of systems modelled by means of strongly connected stochastic MGs. Both bounds can be computed by means of proper Linear Programming problems on the incidence matrix of the net, whose solution is known to be of worst case theoretical polynomial complexity. As a by-product, we can characterize the liveness of a MG in terms of non-null throughput for all its transitions, so that we obtained an alternative proof of a recently obtained result on the polynomiality of the liveness problem for MGs [ES89]. This shows an example of possible interleaving between qualitative and quantitative analysis for timed and stochastic Petri nets.

The upper bound on throughput for MGs was first proposed by Ramchandani in 1974, and then re-discovered and/or re-interpreted by many others, in the framework of the study of the exact performance of timed Petri nets with deterministic timing. The contributions given by this paper in this sense are three: an alternative reformulation in terms of Linear Programming problems; the proof that this case represents an upper bound in performance independently of the probability distribution also in the framework of stochastic Petri nets; the proof that the upper bound is reachable not only by deterministic but also by stochastic models, with arbitrary values of coefficient of variations.

The lower bounds on throughput presented in this paper as well as the concept of enabling bound for transitions are new results. The lower bound in throughput consisting in the inverse of the sum of the firing times of all transitions divided by their respective enabling bounds reduces to the trivial sequentialization of all transitions in the case of safe nets, but has been shown to be reachable with some probability distribution when the coefficient of variation increases. The concept of enabling bound generalizes the usual one of enabling for a transition, and provides another example of possible interleaving between qualitative and quantitative analysis for timed and stochastic Petri nets.

This work can be extended in two directions: by considering classes of Petri nets behaviourally “similar” to MGs, as done in the companion paper [CCS89], or by removing some behavioural restriction. Work is still in progress for the case of unbounded MGs, for which the extension of the results presented in this paper is not trivial. In particular, some “trivial” extensions suggested by many authors that studied the case of deterministic bounded MGs (like, e.g., [Mag84]) appear not to work in the case of unbounded MGs.

References

[AMBC84] 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*, 2(1), May 1984.

[Bes86] E. Best. Structure theory of Petri nets: the free-choice hiatus. In *Advanced Course on Petri Nets*, L.N.C.S. Springer-Verlag, Berlin, September 1986.

[BG85] S.C. Bruell and S. Ghanta. Throughput bounds for generalized stochastic petri net models. In *Proc. Int. Workshop on Timed Petri Nets*, IEEE-CS Press, Torino, Italy, July 1985.

[Bra83] G. W. Brams. *Réseaux de Petri: Théorie et Pratique. T.1. théorie et analyse.* Masson, Paris, 1983.

[BV84] E. Best and K. Voss. Free-choice systems have home states. *Acta Informatica*, (21):89–100, 1984.

[CCS89] J. Campos, G. Chiola, and M. Silva. Properties and steady-state performance bounds for Petri nets with unique repetitive firing count vector. In *Proc. of 3rd Intern. Workshop PNPM'89*, IEEE-CS Press, Kyoto, Japan, December 1989.

[CHEP71] F. Commoner, A. W. Holt, S. Even, and A. Pnueli. Marked directed graphs. *Journal of Computer and System Science*, (5):511–523, 1971.

[ES89] J. Esparza and M. Silva. Liveness of bounded free-choice nets is decidable in polynomial time. Report, Dpto. Ingeniería Eléctrica e Informática, Universidad de Zaragoza, February 1989.

[FN85a] G. Florin and S. Natkin. Les réseaux de Petri stochastiques. *Technique et Science Informatiques*, 4(1):143–160, 1985.

[FN85b] G. Florin and S. Natkin. Les réseaux de Petri stochastiques. Thesis de Doctorat d'Etat, Université P. et M. Curie, Paris, 1985.

[HV87] M. A. Holliday and M. Vernon. A generalized timed Petri net model for performance analysis. *IEEE Transactions on Software Engineering*, 13(12):1297–1310, December 1987.

[Kar84] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, (4):373–395, 1984.

[Mag84] J. Magott. Performance evaluation of concurrent systems using Petri nets. *Information Processing Letters*, (18):7–13, 1984.

[Mol82] M. K. Molloy. Performance analysis using stochastic Petri nets. *IEEE Transaction on Computers*, 31(9):913–917, September 1982.

[Mol85] M.K. Molloy. Fast bounds for stochastic petri nets. In *Proc. Int. Workshop on Timed Petri Nets*, IEEE-CS Press, Torino, Italy, July 1985.

[Mur77] T. Murata. Circuit theoretic analysis and synthesis of marked graphs. *IEEE Transactions on Circuits and Systems*, 24(7), July 1977.

[Mur83] K. G. Murty. *Linear Programming.* John Wiley & Sons, 1983.

[Mur85] T. Murata. Use of resource-time product concept to derive a performance measure of timed Petri nets. In *Proc. 1985 Midwest Symposium Circuits and Systems*, Louisville, U.S.A. August 1985.

[Ram74] C. Ramchandani. *Analysis of Asynchronous Concurrent Systems by Petri Nets.* PhD thesis, M.I.T. Cambridge, Mass. February 1974.

[RH80] C. V. Ramamoorthy and G. S. Ho. Performance evaluation of asynchronous concurrent systems using Petri nets. *IEEE Transactions on Software Engineering*, 6(5):440–449, September 1980.

[RP84] R. R. Razouk and C. V. Phelps. Performance analysis using timed Petri nets. In *Proc. International Conference on Parallel Processing*, pages 126–129, August 1984.

[SC87] M. Silva and J. M. Colom. On the computation of structural synchronic invariants in P/T nets. In G. Rozenberg, editor, *Advances in Petri Nets 1988*, pages 386–417, L.N.C.S. Springer-Verlag, Berlin, 1987.

[Sif78] J. Sifakis. Use of Petri nets for performance evaluation. *Acta Cybernetica*, 4(2):185–202, 1978.

[WT85] R. P. Wiley and R. R. Tenney. Performance evaluation of decision-free stochastic timed Petri nets. In *Proc. 24th Conf. Decision and Control*, Ft. Lauderdale, FL. December 1985.

[Zub85] W. M. Zuberek. Performance evaluation using timed Petri nets. In *Proc. International Workshop on Timed Petri Nets*, pages 272–278, IEEE-CS Press, Torino,Italy, July 1985.

A Petri net definitions and notation

A.1 Net structure

A *place/transition* net $\mathcal{N}$ is a 4-tuple $\mathcal{N} = \langle P, T, Pre, Post\rangle$, where

- P is the set of *places* ($|P|$ =n),
- T is the set of *transitions* ($P \cap T = \emptyset$) ($|T|$ =m),
- Pre ($Post$) is the pre- (post-) incidence function representing the input (output) arcs $Pre : P \times T \rightarrow I\!N$ ($Post : P \times T \rightarrow I\!N$)

The *pre-* and *post-set* of a transition $t \in T$ are defined respectively as ${}^\bullet t = \{p | Pre(p,t) > 0\}$ and $t^\bullet = \{p | Post(p,t) > 0\}$. The *pre-* and *post-set* of a place $p \in P$ are defined respectively as ${}^\bullet p = \{t | Post(p,t) > 0\}$ and $p^\bullet = \{t | Pre(p,t) > 0\}$.

The *incidence matrix* of the net $C = [c_{ij}]$ ($1 \leq i \leq$n, $1 \leq j \leq$m) is defined by $c_{ij} = Post(p_i,t_j) - Pre(p_i,t_j)$. Similarly the pre- and post-incidence matrices are defined as $PRE = [a_{ij}]$ and $POST = [b_{ij}]$, where $a_{ij} = Pre(p_i,t_j)$ and $b_{ij} = Post(p_i,t_j)$.

A.2 Token game

A function $M : P \rightarrow I\!N$ is called a *marking.* A marking M can be represented in vector form, with the i^{th} component associated with the i^{th} element of P. A *marked P/T net* $\langle \mathcal{N}, M_0\rangle$ is a P/T net $\mathcal{N}$ with an *initial marking* M_0.

A transition $t \in T$ is *enabled* in marking M iff $\forall p \in P$ $M(p) \geq Pre(p,t)$. A transition t_j enabled in M can *fire* yielding a new marking M' defined by $M'(p) = M(p) - Pre(p,t_j) + Post(p,t_j)$ (or in vector form, $M'[i] = M[i] + C[i,j]$). The notation $M[t\rangle M'$ denotes that transition t

is enabled in M and that M' is *reached* from M by firing t in it.

A finite sequence of transitions $\sigma = t_1 t_2 \dots t_n$ is a *finite firing sequence* of $\langle \mathcal{N}, M_0 \rangle$ iff there exist a sequence of markings such that $M_0[t_1\rangle M_1[t_2\rangle M_2 \dots [t_n\rangle M_n$. In this case, marking M_n is said to be *reachable* from M_0 by firing σ, and this is denoted by $M_0[\sigma\rangle M_n$. Similarly, an *infinite firing sequence* $\sigma = t_1 t_2 \dots$ is defined for $\langle \mathcal{N}, M_0 \rangle$ iff there exist an infinite sequence of markings such that $\forall i \in \mathbb{N}\ M_{i-1}[t_i\rangle M_i$.

The notation $M[\sigma\rangle$ denotes a firable sequence σ from marking M. The function $\vec{\sigma} : T \rightarrow \mathbb{N}$ is the *firing count vector* of the firable sequence σ, i.e. $\vec{\sigma}[t]$ represents the number of occurrences of $t \in T$ in σ. If $M_0[\sigma\rangle M$, then we can write in vector form $M = M_0 + C\vec{\sigma}$, which is referred to as the *linear state equation* of the net. A marking M' is said to be *potentially reachable* iff $\exists \vec{\sigma} \geq 0$ such that $M' = M_0 + C\vec{\sigma}$.

A.3 Basic properties

The *reachability set* $R(\mathcal{N}, M_0)$ is the set of all markings reachable from the initial marking. Denoting by $PR(\mathcal{N}, M_0)$ the set of all potentially reachable markings we have the following relation: $R(\mathcal{N}, M_0) \subseteq PR(\mathcal{N}, M_0)$. $L(\mathcal{N}, M_0)$ is the set of all firing sequences and their suffixes in $\langle \mathcal{N}, M_0 \rangle$: $L(\mathcal{N}, M_0) = \{\sigma | M[\sigma\rangle \text{ and } M \in R(\mathcal{N}, M_0)\}$.

A place $p \in P$ is said to be k–bounded iff $\forall M \in R(\mathcal{N}, M_0)\ M(p) \leq k$. A marked net $\langle \mathcal{N}, M_0 \rangle$ is said to be (marking) K–bounded iff each of its places is K–bounded. A net $\mathcal{N}$ is *structurally bounded* iff $\forall M_0$ the marked nets $\langle \mathcal{N}, M_0 \rangle$ are K–bounded for some $K \in \mathbb{N}$.

A transition $t \in T$ is *live* in $\langle \mathcal{N}, M_0 \rangle$ iff $\forall M \in R(\mathcal{N}, M_0)\ \exists M' \in R(\mathcal{N}, M)$ such that M' enables t. The marked net $\langle \mathcal{N}, M_0 \rangle$ is live iff all its transitions are live (i.e. liveness of the net guarantees the possibility of an infinite activity of all transitions). A net $\mathcal{N}$ is *structurally live* iff $\exists M_0$ such that the marked net $\langle \mathcal{N}, M_0 \rangle$ is live. The marked net $\langle \mathcal{N}, M_0 \rangle$ is *deadlock-free* iff $\forall M \in R(\mathcal{N}, M_0)$ $\exists t \in T$ such that M enables t.

A *repetitive component* is a function (vector) $X : T \rightarrow \mathbb{N}$ such that $X \neq 0$ and $C \cdot X \geq 0$. A *consistent repetitive component* (or *T-semiflow*) is a repetitive component X such that $C \cdot X = 0$. A *conservative component* (or *P-semiflow*) is a function (vector) $Y : P \rightarrow \mathbb{N}$ such that $Y \neq 0$ and $Y^T \cdot C = 0$. The *support* of (T- and P-) semiflows are defined by $||X|| = \{t \in T | X(t) > 0\}$ and $||Y|| = \{p \in P | Y(p) > 0\}$. A (T- or P-) semiflow I is *minimal support* iff there exist no other semiflow I' such that $||I'|| \subset ||I||$. A (T- or P-) semiflow is *canonical* iff the greatest common divisor of its components is 1. A (T- or P-) semiflow is *elementary* iff it is canonical and minimal support.

A net $\mathcal{N}$ is *repetitive* if there exist a repetitive component $X \geq \vec{1}$. A net $\mathcal{N}$ is *consistent* if there exist a T-semiflow $X \geq \vec{1}$. A net $\mathcal{N}$ is *conservative* if there exist a P-semiflow $X \geq \vec{1}$.

A.4 Additional properties

An *implicit* place is one which never restricts the firing of its output transitions. Let $\mathcal{N}$ be any net and $\mathcal{N}_p$ be the net resulting from adding a place p to $\mathcal{N}$. If M_0 is an initial marking of $\mathcal{N}$, $M_0 \cup m_0(p)$ denotes the initial marking of $\mathcal{N}_p$. The place p is implicit in the marked net $\langle \mathcal{N}_p, M_0 \cup m_0(p) \rangle$ iff $L(\mathcal{N}_p, M_0 \cup m_0(p)) = L(\mathcal{N}, M_0)$.

A *livelock* is a maximal subset of strongly connected states that have no connections outside the subset itself.

DEFINITION A.1 [BV84], [Bra83] *A marked Petri net has the directedness (or confluence) property iff for all pair of reachable markings, $M_0[\sigma_1\rangle M_1$ and $M_0[\sigma_2\rangle M_2$, there exist two sequences σ_1' and σ_2' such that $M_1[\sigma_1'\rangle M$ and $M_2[\sigma_2'\rangle M$.*

DEFINITION A.2 *$M \in R(\mathcal{N}, M_0)$ is a home state iff $\forall M_i \in R(\mathcal{N}, M_0) : M \in R(\mathcal{N}, M_i)$.*

DEFINITION A.3 *A marked net is reversible iff its initial marking is a home state.*

A.5 Marked Graphs

DEFINITION A.4 [CHEP71] *MGs are ordinary Petri nets (pre and post incidence functions taking values in $\{0, 1\}$) such that $|{}^\bullet p| = |p^\bullet| = 1, \forall p \in P$.*

PROPERTY A.1 *MGs are structurally persistent nets.*

THEOREM A.1 [Mur77] *Let $\langle \mathcal{N}, M_0 \rangle$ be a live (possibly unbounded) MG. The two following statements are equivalent:*

i) *$M \in R(\mathcal{N}, M_0)$, i.e. M is reachable from M_0.*

ii) *$B_f \cdot M = B_f \cdot M_0$, with B_f the fundamental circuit matrix of the graph, and $M \geq 0$.*

According to the above theorem $M \in R(\mathcal{N}, M_0) \Leftrightarrow M_0 \in R(\mathcal{N}, M)$. In other words:

PROPERTY A.2 *Live MGs are reversible.*

PROPERTY A.3 *Let $\mathcal{N}$ be a MG.*

i) *$\mathcal{N}$ is structurally bounded (i.e. $\langle \mathcal{N}, M_0 \rangle$ is bounded $\forall M_0$) iff it is strongly connected.*

ii) *Let $\langle \mathcal{N}, M_0 \rangle$ be live. Then $\langle \mathcal{N}, M_0 \rangle$ is bounded iff $\mathcal{N}$ is structurally bounded.*