Operational Analysis of Timed Petri Nets and Application to the Computation of Performance Bounds

G. Chiola,* C. Anglano,

Dipartimento di Informatica
Università di Torino
Torino, Italy 10149

J. Campos,† J.M. Colom,† and M. Silva†

Dpto. de Ingeniería Eléctrica e Informática
Universidad de Zaragoza
Zaragoza, Spain 50015

Abstract

We use operational analysis techniques to partially characterize the behaviour of timed Petri nets under very weak assumptions on their timing semantics. New operational inequalities are derived that are typical of the presence of synchronization and that were therefore not considered in queueing network models. We show an interesting application of the operational laws to the statement and the efficient solution of problems related to the estimation of performance bounds insensitive to the timing probability distributions. The results obtained generalize and improve in a clear setting results that were derived in the last few years for several different subclasses of timed Petri nets. In particular the extension to Well-Formed Coloured nets appears straightforward and allows an efficient exploitation of models symmetries.

1 Introduction

Operational analysis is a conceptually very simple way of deriving mathematical equations relating observable quantities in queueing systems [11]. In [10] the reader can find some nice examples of how the application of operational analysis techniques can help in explaining and proving fundamental results in queueing network analysis. Here we apply operational analysis techniques to derive linear equations and inequalities relating interesting performance measures in timed Petri net models. The main conceptual difference between queueing and Petri net models is the presence of a synchronization primitive in the latter. Early works on extensions of operational analysis to Petri nets include [12], where however synchronization was neglected. New operational inequalities are derived here for synchronization elements that have no counterpart in operational laws for queueing networks.

Some classical results of queueing networks were already proven to hold in stochastic Petri net models. In this paper we derive, under much weaker conditions, a generalization of the classical *utilization law* for the case of multiply enabled transitions and several inequalities that relate throughput, average marking, and average transition firing time in case of synchronization transitions. All these results are derived for each possible observable sample path. Therefore, in order to compare to classical queueing laws stated in a stochastic framework, the additional hypothesis of unique limit behaviour for each sample path must be assumed.

In addition to the mathematical interest of these derivations, we propose also an application of these results to the computation of performance bounds based on linear programming techniques. Such performance bounds are fairly inexpensive to compute compared to the cost of discrete event simulation or exact Markovian analysis, and moreover provide results that are insensitive of the probability distribution of the transition firing times. The linear programming problems (LPP's) presented in this paper represent also a generalization of some recent results published in [3, 4, 2] since they can be applied to arbitrary Petri net structures and reduce to the previous ones when the Petri net structure satisfies some particular constraints.

The paper is organized as follows. Section 2 presents the operational analysis of timed Petri nets and the derivation of the main equations. Section 3 shows the application of the operational laws, also considering the case of Well-Formed Coloured nets, to the statement of LPP's for the computation of perfor-

*This work was performed while G. Chiola was visiting the University of Zaragoza, Spain, in the framework of the ESPRIT–BRA project No.7269 "QMIPS."

†This work has been supported by the Spanish PRONTIC 354/91 and the Aragonese CONAI-DGA P-IT 6/91.

mance bounds depending only on the average transition firing times, the structure, and the initial marking of the net. Section 4 provides an example of computation of such bounds in the case of a Coloured Well-Formed timed Petri net model. Finally, Section 5 contains some concluding remarks and ideas for future research on the topics.

2 Observable quantities and operational laws

In this section we start by defining measurable quantities that characterize the state and the behaviour in time of a Petri net model. Then we derive and prove in a very simple and direct way some fundamental relations that hold true "operationally" among them, i.e. that are verified in any sample path that one can measure in an experiment.

We assume the reader to be familiar with the Petri net formalism and notation. We refer to [13] or [15] for an introduction to Petri nets and most of their behavioural properties and analysis techniques. We also refer to [1] for a detailed discussion of different timing semantics and related operation mechanisms. We just resume here the notation conventions that are used in the following of this paper.

$\mathcal{N} = (P, T, W, M_0)$ is a net system, where P is the set of places, T is the set of transitions, $W : P \times T \cup T \times P \rightarrow \mathbb{N}$ is the incidence function, and M_0 is the initial marking (in general, a marking is $M : P \rightarrow \mathbb{N}$, and $\forall p_i \in P$, $M[p_i]$ is the number of tokens in p_i). The input (output) set of $x \in P \cup T$ is ${}^\bullet x = \{y \in P \cup T \mid W(y,x) \geq 1\}$ ($x^\bullet = \{y \in P \cup T \mid W(x,y) \geq 1\}$).

2.1 Basic operational quantities

Assume that a generic timed Petri net is available for measurement, and that the following quantities can be collected during an experiment, starting at time $\tau = 0$ and ending at time $\tau = \theta > 0$, at which all transitions have been fired at least once. The total number of transitions firings during the experiment is assumed finite.

Instantaneous marking: $\forall p_k \in P, \forall \tau : 0 \leq \tau \leq \theta$, $M[p_k](\tau)$ represents the number of tokens in place p_k at time τ.

Average marking during the experiment interval:

$$\forall p_k \in P, \quad \bar{M}[p_k](\theta) = \frac{1}{\theta} \int_0^\theta M[p_k](\tau) d\tau$$

Instantaneous enabling degree: $\forall t_i \in T, \forall \tau : 0 \leq \tau \leq \theta$, $e_i(\tau)$ represents the internal concurrency of transition t_i at time τ, i.e.

$$e_i(\tau) = \max\{k \in \mathbb{N} : \forall p \in {}^\bullet t_i,\ M[p](\tau) \geq k\, W(p, t_i)\}$$

The following relation holds by definitions:

$$\forall t_i \in T, \quad \forall \tau, \quad e_i(\tau) = \min_{p \in {}^\bullet t_i} \left\lfloor \frac{M[p](\tau)}{W(p, t_i)} \right\rfloor \qquad (1)$$

(where $\forall a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the largest integer not greater than a).

Average enabling degree: $\forall t_i \in T$, $\bar{e}_i(\theta) = \frac{1}{\theta} \int_0^\theta e_i(\tau) d\tau$ represents the average number of servers active in transition t_i during the experiment interval.

Since we use an "infinite-server" semantics for transition enabling, we need to consider the activities of the different servers in a given transition t_i independently. Without loss of generality we assume an ordering of the servers associated with transitions such that busy servers always come before idle servers, i.e., at any point in time τ the first $e_i(\tau)$ servers are active inside transition t_i, while the remaining ones are idle.

Under this assumption we can define the:

Number of firings completed by the j-th server in t_i from time 0 up to time θ, denoted $F_{i,j}(\theta)$.

Total number of firings of t_i during the experiment interval

$$F_i(\theta) = \sum_{j=1}^{\infty} F_{i,j}(\theta)$$

(by assumptions, $0 < F_i(\theta) < \infty$).

Throughput of t_i $x_i(\theta) = \frac{F_i(\theta)}{\theta}$ that represents the average number of firings completed per time unit.

2.2 Conflict-free nets

In case of nets without conflicts one can easily define the average service time of transitions as a function of the busy times of all servers. In particular we define:

Instantaneous enabling of j-th server in t_i

$$e_{i,j}(\tau) = \text{ if } e_i(\tau) \geq j \text{ then } 1 \text{ else } 0$$

characteristic function that evaluates to 1 if and only if the j-th server in transition t_i is busy at time τ.

Busy time for the j-th server of t_i $\theta_{i,j}(\theta) = \int_0^\theta e_{i,j}(\tau)d\tau$

Service time for the j-th server of t_i $S_{i,j}(\theta) = \frac{\theta_{i,j}(\theta)}{F_{i,j}(\theta)}$

Average service time for t_i $\bar{S}_i(\theta) = \frac{\sum_{j=1}^{\infty} \theta_{i,j}(\theta)}{\sum_{j=1}^{\infty} F_{i,j}(\theta)}$

The following equation holds for any measurement experiment:

Enabling operational law

$$\forall t_i \in T, \quad \bar{e}_i(\theta) = x_i(\theta)\bar{S}_i(\theta) \tag{2}$$

Proof: By definition $x_i(\theta) = \frac{\sum_{j=1}^{\infty} F_{i,j}(\theta)}{\theta}$ then multiplying and dividing by $\sum_{j=1}^{\infty} \theta_{i,j}(\theta)$ and recalling the definition of $\bar{S}_i(\theta)$ we obtain:

$$\forall t_i \in T, \quad x_i(\theta)\bar{S}_i(\theta) = \frac{1}{\theta}\sum_{j=1}^{\infty} \theta_{i,j}(\theta)$$

and then substituting the definition of $\theta_{i,j}(\theta)$ and exchanging the integral and the sum signs:

$$\forall t_i \in T, \quad x_i(\theta)\bar{S}_i(\theta) = \frac{1}{\theta}\int_0^\theta \sum_{j=1}^{\infty} e_{i,j}(\tau)d\tau$$

Now it is trivial to identify the integrand to be the instantaneous enabling degree $e_i(\tau)$, so that the result follows. Q.E.D.

The above enabling law is the well-known "utilization law" derived in the framework of multiple server queues. From the enabling law it follows that if the average firing time of a transition is known, then its throughput is proportional to its average enabling degree. Of course in case of immediate transitions $\bar{S}_i(\theta) = 0$, so immediate transitions are never enabled for non-null intervals of time.

We are now in a position to state our *synchronization inequalities* that relate the throughput and the average marking of the input places for any transition.

Upper bound inequality. $\forall t_i \in T$,

$$x_i(\theta)\ \bar{S}_i(\theta)\ \leq\ \min_{p_k \in {}^\bullet t_i} \left(\frac{\bar{M}[p_k](\theta)}{W(p_k, t_i)}\right) \tag{3}$$

The inequality becomes an equality whenever $\sum_{p\in {}^\bullet t_i} W(p, t_i) = 1$.

Proof: We start from Equation (1) that is valid in each instant of the experiment. Of course this implies that $\forall p_k \in {}^\bullet t_i$, $\forall \tau\ :\ 0 \leq \tau \leq \theta$, $e_i(\tau) \leq \frac{M[p_k](\tau)}{W(p_k,t_i)}$. Therefore $\forall p_k \in {}^\bullet t_i$, $\bar{e}_i(\theta) \leq \frac{\bar{M}[p_k](\theta)}{W(p_k,t_i)}$, and applying the enabling operational law the result follows. Q.E.D.

This inequality establishes an upper bound for the average enabling (hence for the transition throughput once the service time is defined) in the case of transitions with more than one input place that model a synchronization. In the following we derive other inequalities that establish lower bounds as well. We shall see that in the particular case of transitions with a single input place the two inequalities reduce to a single equality.

Lower bound inequality for single input arc ($W(p, t_i) \geq 1$). $\forall t_i \in T\ :\ {}^\bullet t_i = \{p\}$,

$$x_i(\theta)\ \bar{S}_i(\theta)\ \geq\ \frac{\bar{M}[p](\theta) - W(p, t_i) + 1}{W(p, t_i)} \tag{4}$$

Notice that in case $W(p, t_i) = 1$ this reduces to $x_i(\theta)\ \bar{S}_i(\theta) \geq \bar{M}[p](\theta)$, that combined with the upper bound inequality (3) reduces to the equation $x_i(\theta)\ \bar{S}_i(\theta) = \bar{M}[p](\theta)$.

Proof: First define some auxiliary punctual marking functions:

$$\forall p \in P,\ \ \forall\tau,\quad M^v[p](\tau)\ =\ \max(0, M[p](\tau) - v)$$
$$\forall p \in P,\ \ \forall\tau,\quad M_l^u[p](\tau)\ =\ M^l[p](\tau) - M^u[p](\tau)$$

Consider now the following properties of the auxiliary function $\forall k \in \mathbb{N} :\ k > 0$,

$$0\ \leq\ M_{kw-1}^{(k+1)w-1}[p](\tau)\ \leq\ w$$

Moreover, notice that the k-th server in transition t_i is enabled if and only if $M_{kw-1}^{(k+1)w-1}[p](\tau) \geq 1$ in case $w = W(p, t_i)$. Therefore we can conclude that:

$$\forall t_i \in T\ :\ {}^\bullet t_i = \{p\}, \quad \forall k \geq 1, \quad \forall \tau,$$
$$e_{i,k}(\tau)\ \geq\ \frac{1}{W(p, t_i)}\ M_{kW(p,t_i)-1}^{(k+1)W(p,t_i)-1}[p](\tau)$$

Hence we derive: $\forall t_i \in T\ :\ {}^\bullet t_i = \{p\}$,

$$\forall\tau,\ e_i(\tau)\ \geq\ \frac{1}{W(p,t_i)}\ \sum_{k=1}^{\infty} M_{kW(p,t_i)-1}^{(k+1)W(p,t_i)-1}[p](\tau) = \ =\ \frac{M[p](\tau)-W(p,t_i)+1}{W(p,t_i)}$$

Finally, taking the average over the experiment interval and applying the enabling law, the result follows. Q.E.D.

Observe that in the case that the right-hand expression in (4) is negative, a trivial inequality can be used: $x_i(\theta)\bar{S}_i(\theta) \geq 0$.

Improvement for bounded nets: $\forall t_i \in T\ :\ {}^\bullet t_i = \{p\}$, if $\forall \tau,\ M[p](\tau) \le B_p$ and $w_{ip} = W(p,t_i)$ and $\exists k \in \mathbb{N}\ : w_{ip}k \le B_p < (k+1)w_{ip}$

$$x_i(\theta)\ \bar{S}_i(\theta)\ \ge\ k\ \frac{\bar{M}[p](\theta) - w_{ip}k + 1}{B_p - w_{ip}k + 1} \qquad (5)$$

Proof: Firstly note that $\forall t_i \in T$, $\forall j \in \mathbb{N}$

$$\bar{e}_i(\theta)\ \ge\ j\ \bar{e}_{i,j}(\theta)\ =\ j\ \frac{\theta_{i,j}(\theta)}{\theta}$$

Secondly, note that the marking in the input place p can be expressed as the sum of two components:

$$\forall \tau, \quad M[p](\tau)\ =\ w_{ip} \sum_{j=1}^{\infty} e_{i,j}(\tau)\ + N[p](\tau)$$

where the component $N[p](\tau) \le w_{ip} - 1$ represents the portion of marking not used to enable the transition. Now taking the integral and dividing by θ one obtains:

$$\bar{M}[p](\theta)\ =\ w_{ip}\ \bar{e}_i(\theta) + \bar{N}[p](\theta)$$

This equation shows that the average enabling depends only on the mean values of the input place marking and of the unused portion of the marking.
The worst case from the point of view of enabling the transition k times occurs when the place is marked with $w_{ip}k - 1$ tokens most of the time and with B_p tokens for the rest of the time, since this case maximizes the unused portion of the average marking in the input place. From these considerations the result follows. Q.E.D.

Lower bound inequality for binary synchronization with ordinary arcs. $\forall t_i \in T\ :\ {}^\bullet t_i = \{p_1, p_2\}$ and $W(p_1,t_i) = W(p_2,t_i) = 1$, if $M[p_1](\tau) \le B_1$ and $M[p_2](\tau) \le B_2$ and $B_1 \le B_2$ then

$$x_i(\theta)\ \bar{S}_i(\theta)\ \ge\ \bar{M}[p_1](\theta) + \frac{B_1}{B_2}\bar{M}[p_2](\theta) - B_1 \qquad (6)$$

Proof: Similarly to the previous case we can write two equations relating the average marking, the average enabling, and the average portion of unused marking for each of the two input places:

$$\bar{e}_i(\theta)\ =\ \bar{M}[p_1](\theta) - \bar{N}[p_1](\theta)$$

$$\bar{e}_i(\theta)\ =\ \bar{M}[p_2](\theta) - \bar{N}[p_2](\theta)$$

Now we can compute upper bounds on the unused part of the marking as follows. The maximum fraction of time during which $N[p_1](\tau)$ may be greater than zero is equal to the minimum time during which $M[p_2](\tau) = 0$ (otherwise the transition would be enabled and the marking of p_1 would contribute to the enabling instead); since place p_2 is B_2 bounded, this fraction of time is less than or equal to $1 - \frac{\bar{M}[p_2](\theta)}{B_2}$; moreover during this maximum time, the maximum value of the marking in p_1 is less than or equal to B_1. Hence

$$\bar{N}[p_1](\theta)\ \le\ B_1\ \left(1 - \frac{\bar{M}[p_2](\theta)}{B_2}\right)$$

and from this the result follows trivially. Q.E.D.

A general lower bound for bounded nets: $\forall t_i \in T\ :\ {}^\bullet t_i = \{p_1, p_2, \ldots, p_n\}$, $\forall j \le n$, $M[p_j](\tau) \le B_j$ and $B_1 \le B_j$

$$x_i(\theta)\ \bar{S}_i(\theta)\ \ge\ \frac{\bar{M}[p_1](\theta) - W(p_1,t_i) + 1 - B_1\ \max(f_j)}{W(p_1,t_i)} \qquad (7)$$

where $\forall j\ :\ 2 \le j \le n$, $f_j = 1 - \frac{\bar{M}[p_j](\theta) - W(p_j,t_i)+1}{B_j - W(p_j,t_i)+1}$

Proof: Similar to the previous ones writing the upper bound for the quantity $\bar{N}[p_1](\theta)$. Q.E.D.

2.3 General nets with conflicts

In the general case in which transitions may be enabled in conflict the definitions of service time and average enabling degree must be modified in order to take the possibility of preemption into account. In the literature two types of timed Petri net semantics have been proposed: *race* and *preselection* conflict resolution policies [1]. According to the race policy all enabled transitions start working, and the first one that completes its firing time seizes the tokens from the input places, thus possibly preempting other transitions. Instead, the preselection policy requires that conflicts be solved at the enabling time instant, so that only selected transitions put their servers to work and fire for sure after the elapsing of their firing time. In any case the same kind of results can be derived.

Conditional instantaneous enabling of j-th server in t_i:

$e'_{i,j}(\tau) =$ if "$e_i(\tau) \ge j$ and the enabling is not preempted" then 1 else 0,

characteristic function that evaluates to 1 if and only if the j-th server in transition t_i is busy at time τ and its work will not be wasted due to the preemption from a conflicting transition. Of course $e'_{i,j}(\tau) \le e_{i,j}(\tau)$ by definition.

Useful busy time for the j-th server of t_i
$\theta'_{i,j}(\theta) = \int_0^\theta e'_{i,j}(\tau) d\tau$

Useful service time for the j-th server of t_i
$S'_{i,j}(\theta) = \frac{\theta'_{i,j}(\theta)}{F_{i,j}(\theta)}$

Useful average service time for transition t_i
$\bar{S'}_i(\theta) = \frac{\sum_{j=1}^{\infty} \theta'_{i,j}(\theta)}{\sum_{j=1}^{\infty} F_{i,j}(\theta)}$

The enabling operational law is extended as:

$$\forall t_i \in T, \quad \bar{e'}_i(\theta) = x_i(\theta)\ \bar{S'}_i(\theta) \tag{8}$$

and the proof is similar to the one shown above. From the comparison with Equation 2 it also follows that $\bar{S'}_i(\theta) \leq \bar{S}_i(\theta)$ independently of the probability distribution of the firing time processes.

Equation (1) however becomes an inequality in case of nets with conflicting transitions:

$$\forall t_i \in T, \ \ \forall \tau, \quad e'_i(\tau) \leq \min_{p_k \in {}^\bullet t_i} \left(\frac{M[p_k](\tau)}{W(p_k, t_i)} \right) \tag{9}$$

The upper bound inequality (3) still holds in this more general setting by just substituting $S'_{i,j}(\theta)$ for $S_{i,j}(\theta)$.

2.3.1 Race versus preselection policy

The quantities $\bar{e'}_i(\theta)$ and $\bar{S'}_i(\theta)$ are in general measurable from an off-line processing of an experiment record without any further assumption.

Using the preselection policy, the useful service time of a transition is exactly the transition firing time. This allows one to derive an improved version of the upper bound inequality: $\forall p_k \in P$,

$$\sum_{t_i \in p_k^\bullet} \left(W(p_k, t_i)\ x_i(\theta)\ \bar{S}_i(\theta) \right) \ \leq \ \bar{M}[p_k](\theta) \tag{10}$$

In the case of race policy, instead, the useful average service time $\bar{S'}_i(\theta)$ might be strictly less than the nominal transition firing times due to the effect of preemption from conflicting transitions. Inequality (10) holds true in a race policy model only if all transitions that are output for place p_k are *behaviourally persistent* (i.e. their enabling is mutually exclusive). In other words, only the following modified version of inequality (3) holds true for behaviourally conflicting timed transitions with race policy:

$$\forall t_i \in T, \quad x_i(\theta)\ \bar{S'}_i(\theta) \ \leq \ \min_{p_k \in {}^\bullet t_i} \left(\frac{\bar{M}[p_k](\theta)}{W(p_k, t_i)} \right) \tag{11}$$

For what concerns the synchronization lower bounds, inequalities (4–7) in general apply only to persistent or immediate transitions (in the latter case $\bar{S}_i = \bar{S'}_i = 0$). The case of conflicting transitions with preselection policy may be treated by net transformation as follows, while for the case of conflicting timed transitions with race policy no synchronization lower bound inequality applies.

Consider transition t_i timed, potentially in conflict with other timed transitions and with preselection conflict resolution policy. Split t_i in two transitions t'_i and t''_i and add a new place p'_i such that $\forall p \in P$ $W(p, t'_i) = W(p, t_i)$ and $\forall p \in P$ $W(t''_i, p,) = W(t_i, p)$ and $W(t'_i, p'_i) = W(p'_i, t''_i) = 1$ and t'_i is immediate and $\bar{S_{t''_i}} = \bar{S}_i$. In the transformed net t''_i is persistent with single input arc (by construction), so that Inequality (4) applies. Transition t'_i is instead immediate, so that a subset of inequalities (4–7) applies even in presence of conflict.

3 Performance bounds based on operational laws

The inequalities that we derived in the previous section can be used to compute upper and lower bounds for the throughput of transitions or for the average marking of places for general timed Petri nets using linear programming techniques. The idea is to compute vectors $\bar{M}$ and $\vec{x}$ that maximize or minimize the throughput of a transition or the average marking of a place among those verifying the previous operational laws and other linear constraints that can be easily derived from the net structure.

A first set of linear equality constraints can be derived from the fact that the vector $\bar{M}$ is an average weight of reachable markings: $\bar{M} = \sum_{M_r \in RS(\vec{M}_0)} \beta_r M_r$. Since for each reachable marking $M_r = M_0 + C \cdot \vec{\sigma}_r$, we obtain that also the average marking must satisfy the same linear equation:

$$\bar{M} = M_0 + C \cdot \vec{\sigma} \tag{12}$$

where $\vec{\sigma} = \sum_{M_r \in RS(\vec{M}_0)} \beta_r \vec{\sigma}_r$.

The following set of linear inequalities imposes that for each place the token flow out is less than or equal to the token flow in: $\forall p_k \in P$,

$$\sum_{t_i \in {}^\bullet p_k} x_i W(t_i, p_k) \geq \sum_{t_o \in p_k^\bullet} x_o W(p_k, t_o) \tag{13}$$

If place p_k is known to be bounded, then the above inequality becomes an equality which represents the classical *flow balance* equation: $C[p_k] \cdot \vec{x} = 0$.

On the other hand, for each pair of transitions t_i, t_j in (behavioural) free conflict (i.e., such that they are always simultaneously enabled or disabled) the following equation is verified:

$$\frac{x_i}{\alpha_i} = \frac{x_j}{\alpha_j} \tag{14}$$

where α_i, α_j are the routing rates that define the resolution of the conflict between t_i and t_j.

Additionally, most of the operational inequality laws that were derived in the previous section linearly relate the average marking of places with the throughput of their output transitions. Hence they can be considered as constraints for an LPP.

3.1 Extension to TWN's

For timed Well-Formed Coloured nets (TWN's) [8] it is possible to derive, directly from the inequalities developed in the previous sections, operational relations allowing an efficient computation of performance bounds. Given a TWN, the basic idea is to consider the corresponding unfolded net and to apply the relations developed in the previous sections. The relations for the TWN are then obtained combining the partial results for the unfolded one.

A fundamental property that TWN's must have in order to be able to combine the results for the unfolded one is the *symmetry*, meaning that in the unfolded nets obtained from the Well-Formed ones all colour instances of a given place and of a given transition must be equivalent. To be more precise, if a transition t has average service time $\bar{S}_t$, then all of its instances have the same average service time. Moreover if a place p is bounded, then we assume that the maximum number of tokens that each of its instances can contain is the same.

In the rest of this section we show, as an example, the derivation of lower bound inequality for single input arc for TWN's. More details on the derivations can be found in [7].

3.1.1 Notation

In this section we give some notations used in the derivations of relations for TWN's ([8]).

Generic function $f = \sum_{j=1}^{k} F_j$, where F_j is the j^{th} tuple and its arity l is given by the number of colour classes composing the colour domain of the place. This definition of function is slightly different from the classical one, since here we allow linear combinations only outside the tuples (i.e. each tuple is composed only by elementary functions). For example the function $F = < S - x, y >$ is written as $F' = < S, y > - < x, y >$.

Cardinality of function $\mid f \mid = \sum_{j=1}^{k} \mid F_j \mid$, where $\mid F_j \mid = \alpha_j \times \prod_{i=1}^{l} \mid (F_j)_i \mid$ is the cardinality of the j^{th} tuple. The coefficient α_j denotes the product of the coefficients of the elementary functions composing the tuple and $(F_j)_i$ is the i^{th} function of the j^{th} tuple. For example if $F_j = < 3x, 2y >$, then $\alpha_j = 6$.
Family of arcs Each tuple F_j of a function f identifies a set of arcs (with weight α_j), whose cardinality is $A(F_j) = \prod_{i=1}^{l} \mid (F_j)_h \mid$. The global number of arcs corresponding to function f is $A(f) = \sum_{j=1}^{k} A(F_j)$, where each $A(F_j)$ has the sign of the corresponding tuple F_j. When $A(f) = 1$, then we denote as α_f the weights associated to the unique family of arcs corresponding to f.
Input and Output relations If t is an input transition of place p (with function f), then $IN(p,t) = \frac{|C(t)|}{|C(p)|} A(f)$ is the number of input instances of t for each instance of p. Similarly if t is an output transition of place p, then $OUT(p,t) = \frac{|C(t)|}{|C(p)|} A(f)$ is the number of output instances of t for each instance of p.

3.1.2 Lower bound inequality for single input arc

To apply this inequality to an unfolded net, the conditions for its applicability must be met for all transition instances. This means that each instance t_i of a coloured transition t must have only one input place. This condition is met if the function f labelling the arc contains only *projection* and *successor* elementary functions (that is $A(f) = 1$).
Inequality for single input arc

$$\forall t \in T: \; {}^{\bullet}t = \{p\}, W^-(p,t) = f, \;\; A(f) = 1$$

$$\alpha_f x_t \bar{S}_t \geq OUT(p,t)\bar{M}[p] - \mid C(t) \mid (\alpha_f - 1)$$

Proof: Assume to have a portion of a TWN containing transition t and its input place p and that $\mid C(t) \mid = n$ and $\mid C(p) \mid = m$. Considering the n instances of t we can write the following set of inequalities

$$\forall i \in \{1, \ldots, n\} \;\; \alpha_f x_{t_i} \bar{S}_{t_i} \geq \bar{M}[p_{t_i}] - \alpha_f + 1$$

where p_{t_i} is the unique input place of transition instance t_i. Summing the left-hand sides and the right-hand sides of the above inequalities we obtain:

$$\alpha_f x_t \bar{S}_t \geq (\sum_{i=1}^{n} \bar{M}[p_{t_i}] - \mid C(t) \mid (\alpha_f - 1)) \tag{15}$$

Since each instance of p appears exactly $OUT(p,t)$ times in the summation of the above expression we can rewrite inequality (15) as

$$\alpha_f x_t \bar{S}_t \geq (OUT(p,t) \sum_{i=1}^{m} M[p_i] - |C(t)| (\alpha_f - 1)) \quad (16)$$

and the result follows. Q.E.D.
In a similar way it is possible to derive, for TWN's, the equivalent of relations devised for timed Petri nets.

3.2 LPP formulation

Performance bounds for TWN's can be computed solving the LPP of table 1 (whose constraints are the relations derived in previous sections) where f is a linear function of $\bar{M}$ and $\vec{x}$. The linear programming problem for bounds computation for non coloured timed Petri nets can be obtained from that of table 1 setting $OUT(p,t) = |C(p)| = |C(t)| = 1$, $\forall p \in P, t \in T$ and observing that condition $A(f) = 1$ always holds true. The average marking equation is written here in explicit form, but it could be written also in matricial form. Moreover relation (c_7) has been derived for TWN's under the hypothesis of strong symmetries. In particular we assumed that, for each input place of transition t in inequality (c_7), the weights of the arcs belonging to the families corresponding to the function labelling the arc are the same. Obviously the uncoloured version of (c_7) has no such restriction.

As we remarked in the case of timed Petri nets, also for TWN's constraint (c_2) becomes an equality for bounded places (c_2'). The equality sign also holds true in (c_4) if $\alpha_f = 1$ (i.e. the unique family of arcs corresponding to function f have weight 1) since in this case it may be combined with the opposite inequality (c_5). For the case of places with several output conflicting transitions, inequality (10) derived in previous section (or its coloured counterpart) can be added if preselection policy is assumed for the resolution of the conflict. The constraint labelled with (c_5) can be improved if the input place to t_i is bounded, by introducing the additional constraint (c_5').

The LPP of table 1 provides a general method to compute upper and lower bounds for arbitrary linear functions of average marking of places and throughput of transitions. For instance, if $f(\bar{M}, \vec{x}) = x_i$, then the problem can be used to compute an upper or a lower bound (depending on the selection of "max" or "min" optimization for the objective function) for the throughput of transition t_i. In an analogous way, upper or lower bounds for the average marking of a given place p_j can be derived by solving the LPP of table 1 for the objective function $f(\bar{M}, \vec{x}) = \bar{M}[p_j]$. The bounds are insensitive to the timing probability distributions since they are based only on the knowledge of the average service times.

Notice also that most equalities and inequalities contain coefficients that depend only on the net structure and on the (known) average transition firing times (and probabilities in case of free choice immediate conflicts). The only coefficients that may be unknown at the time of the formulation of the model are the actual bounds for places B_i. If the modeller has no a-priori more precise knowledge of these bounds, notice that an upper bound for them that can be used in the LPP of table 1 may be computed from a simplified LPP that contains only constraint c_1 (structural marking bound).

An improvement of the proposed bounds can be obtained if additional constraints that improve the linear characterization of the average marking in terms of the equation $\bar{M} = M_0 + C \cdot \vec{\sigma}$ are considered. For instance, if a *trap* P_T (i.e., $P_T \subseteq P, P_T^\bullet \subseteq {}^\bullet P_T$) is not a P-semiflow, the net is live, and we are interested only in the steady state performance, then we can add the constraint: $\sum_{p_k \in P_T} \bar{M}[p_k] \geq 1$.

Similarly, if a *siphon* P_S ($P_S \subseteq P$, ${}^\bullet P_S \subseteq P_S^\bullet$) is not a P-semiflow and the net is live, then we can add the constraint: $\sum_{p_k \in P_S} \bar{M}[p_k] \geq 1$.

The systematic method for the improvement of linear characterization of reachable markings based on the addition of *implicit places*, presented in [9], can be also applied as in [5].

We remark that linear programming problems can be solved in *polynomial time* [14], therefore the above presented method for the computation of (upper and lower) bounds for the throughput and for the average marking of general timed nets has a polynomial complexity on the number of nodes of the net. Moreover, the *simplex* method for the resolution of LPP's proceeds in linear time in most cases even if it has a theoretically exponential complexity.

Similar results, based on linear programming techniques, were presented in previous works [3, 4, 2] for the computation of throughput upper bounds for particular net subclasses, such as marked graphs or free choice nets. The new approach derived in this section generalizes those recent results in two ways: first, it can be applied to arbitrary Petri net structures; second, it allows one to compute upper and lower bounds for throughput and average marking in a simple and unified way. The proposed method produces the same results as previous ones [3, 4, 2] when the same net subclasses are considered.

Table 1: Linear programming problem.

maximize [or minimize] $f(\bar{M},\vec{x})$	subject to
$\bar{M}[p] = M_0[p] + \sum_{t_i \in {}^\bullet p} \lvert f_i \rvert \, \sigma_{t_i} - \sum_{k_j \in p^\bullet} \lvert g_j \rvert \, \sigma_{k_j};$	(c_1) $\forall p \in P : W^+(p,t_i) = f_i,\ W^-(p,k_i) = g_j$
$\sum_{t_i \in {}^\bullet p} \lvert f_i \rvert \, x_{t_i} \geq \sum_{k_j \in p^\bullet} \lvert g_j \rvert \, x_{k_j};$	(c_2) $\forall p_k \in P : W^+(p,t_i) = f_i,\ W^-(p,k_i) = g_j$
$\sum_{t_i \in {}^\bullet p} \lvert f_i \rvert \, x_{t_i} = \sum_{k_j \in p^\bullet} \lvert g_j \rvert \, x_{k_j};$	(c_2') $\forall p_k \in P$ bounded
$\frac{x_i}{\alpha_i} = \frac{x_j}{\alpha_j},$	(c_3) $\forall t_i, t_j \in T :$ behaviourally free choice
$\lvert f \rvert \cdot x_t \bar{S}_t \leq OUT(p,t)\bar{M}[p]$	(c_4) $\forall t \in T, \forall p \in {}^\bullet t : W^-(p,t) = f$
$\alpha_f x_t \bar{S}_t \geq OUT(p,t)\bar{M}[p] - \lvert C(t) \rvert \, (\alpha_f - 1)$	(c_5) $\forall t \in T$ persistent or immediate $: {}^\bullet t = \{p\},$ $W^-(p,t) = f,\ \ A(f) = 1$
$x_t \bar{S}_t \geq k \frac{OUT(p,t)\bar{M}[p] + \lvert C(t) \rvert \, (1 - k\alpha_f)}{OUT(p,t) + \lvert C(t) \rvert \, (1 - k\alpha_f)}$	(c_5') $\forall t \in T$ persistent or immediate $: {}^\bullet t = \{p\},$ $W^-(p,t) = f,\ \ A(f) = 1$ $\wedge k \in \mathbb{N} : k\alpha_f \leq B_p \leq (k+1)\alpha_f$
$x_t \bar{S}_t \geq OUT(p,t)(\bar{M}[p] + \frac{B_p}{B_q}\bar{M}[q] - B_p)$	(c_6) $\forall t \in T$ persistent or immediate: ${}^\bullet t = \{p,q\},$ $W(p,t) = f, W(q,t) = g, A(f) = A(g) = 1, \lvert f \rvert = \lvert g \rvert = 1$
$\alpha_f x_t \bar{S}_t \geq OUT(p,t)\bar{M}[p] + \lvert C(t) \rvert \, (1 - \alpha_f) +$ $-OUT(p,t)B_p \cdot \left(\lvert C(t) \rvert - \frac{OUT(q,t)\bar{M}[q] + \lvert C(t) \rvert (1-\alpha_g)}{OUT(q,t)B_q + \lvert C(t) \rvert (1-\alpha_g)} \right)$	(c_6') $\forall t \in T$ persistent or immediate $: {}^\bullet t = \{p,q\},$ $B_p \leq B_q,\ W(p,t) = f, W(q,t) = g, A(f) = A(g) = 1$
$\alpha_1 x_t \bar{S}_t \geq OUT(p1,t)\bar{M}[p1] - \lvert C(t) \rvert \, (-\alpha_1 + 1) +$ $-OUT(p1,t)B_{p1} \max_{1 \leq j \leq n} fj$ $fj = 1 - \frac{OUT(pj,t)\bar{M}[pj] + \lvert C(pj) \rvert (-\alpha_j + 1)}{B_{pj} / \lvert C(pj) \rvert - \alpha_j + 1}$	(c_7) $\forall t \in T$ persistent or immediate: ${}^\bullet t = \{p1, \ldots, pn\}, B_{p1} \leq B_{pj}, j \in \{2, \ldots, n\},$ $W(pi,t) = f_i,\ A(f_1) = 1$
$\bar{M}, \vec{x}, \vec{\sigma} \geq 0$	(c_8)

4 An example of application

Let us present an example of application for the computation of bounds in the case of the TWN of figure 1. The architecture comprises a set of processing modules interconnected by a common bus called the "external bus". A processor can access its own memory module directly from its private bus through one port, or it can access non-local shared-memory modules by means of the external bus. In case of contention for the access to one shared-memory module, preemptive priority is given to external access through the external bus with respect to the accesses from the local processor. The experiments on the shared-memory model have been carried out assuming to have 4 processors and that the average service time of all the transitions are equal to 0.5. According to the arguments presented in the previous sections, bounds can be computed solving LPP's with constraints included in table 2, where the first letters of each transition name have been used for reasons of space. The solution for the LPP leads to upper and lower bounds, for the throughput of transitions, given by $\frac{8}{11} \leq x_{e\_e\_a} \leq 2$, while the "exact" solution with exponential distribution is $x_{e\_e\_a} = 1.71999$. An improvement in the lower bound can be obtained observing that when a token arrives in place Choice transition choose_m is enabled at least for one transition instance. This implies that the average marking of place Choice is equal to 0 (transition choose_m is immediate), so $\bar{M}[Choice] = 0$ and $B_{Choice} = 0$ (only tangible markings are considered) can be added to the set of constraints. Moreover place Memory is implicit w.r.t. the enabling of transition b_ext_acc, so we can consider this transition as having only two input places, so constraint (c_6) can be applied instead of constraint (c_7). Finally $B_{Queue} = 3$ can be added

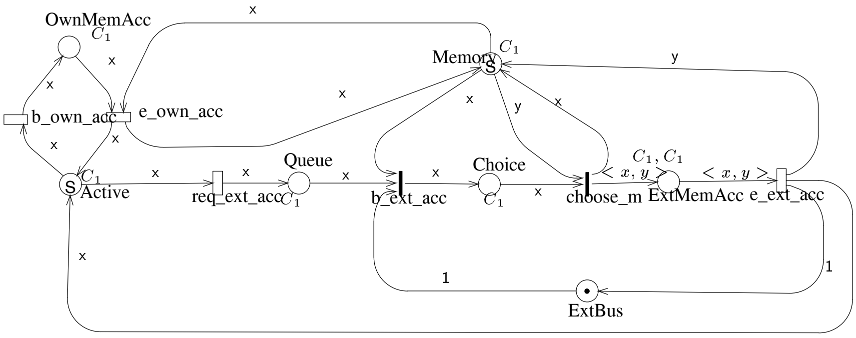

Figure 1: TWN model of a shared-memory multiprocessor.

Table 2: Constraints for the model in figure 1.

(c_1)	$\bar{M}[Active] = 4 + \sigma_{e\_e\_a} + \sigma_{e\_o\_a} +$
	$-\sigma_{r\_e\_a} - \sigma_{b\_o\_a};$
	$\bar{M}[Memory] = 4 + \sigma_{e\_e\_a} - \sigma_{b\_e\_a};$
	$\bar{M}[OwnMemAcc] = \sigma_{b\_o\_a} - \sigma_{e\_o\_a};$
	$\bar{M}[Queue] = \sigma_{r\_e\_a} - \sigma_{b\_e\_a};$
	$\bar{M}[Choice] = \sigma_{b\_e\_a} - \sigma_{c\_m};$
	$\bar{M}[ExtMemAcc] = \sigma_{c\_m} - \sigma_{e\_e\_a};$
	$\bar{M}[ExtBus] = 1 + \sigma_{e\_e\_a} - \sigma_{b\_e\_a};$
(c_2')	$x_{e\_e\_a} + x_{e\_o\_a} = x_{r\_e\_a} + x_{b\_o\_a};$
	$x_{b\_e\_a} = x_{c\_m} = x_{e\_e\_a} = x_{r\_e\_a};$
(c_3)	$x_{b\_o\_a} = x_{r\_e\_a};$
$(c_4\&c_5)$	$x_{b\_o\_a}\bar{S}_{b\_o\_a} = \frac{\bar{M}[Active]}{2};$
	$x_{r\_e\_a}\bar{S}_{r\_e\_a} = \frac{\bar{M}[Active]}{2};$
	$x_{e\_e\_a}\bar{S}_{e\_e\_a} = \bar{M}[ExtMemAcc];$
(c_4)	$x_{e\_o\_a}\bar{S}_{e\_o\_a} \leq \bar{M}[OwnMemAcc];$
	$x_{e\_o\_a}\bar{S}_{e\_o\_a} \leq \bar{M}[Memory];$
(c_6)	$x_{e\_o\_a}\bar{S}_{e\_o\_a} \geq \bar{M}[OwnMemAcc] +$
	$+\frac{B_{OwnMemAcc}}{B_{Memory}}\bar{M}[Memory] - B_{Memory};$
(c_7)	$4(\bar{M}[ExtBus] - B_{ExtBus}(1 - \frac{\bar{M}[Memory]}{B_{Memory}})) \leq 0$
	$4(\bar{M}[ExtBus] - B_{ExtBus}(1 - \frac{\bar{M}[Queue]}{B_{Queue}})) \leq 0;$

since the output transition of place Queue is immediate, and from the behaviour of the model it is clear that at most 3 processors can be waiting in the queue. The relations (c_7) in the above LPP can thus be replaced with the new constraint:

$$4(\bar{M}[ExtBus] + \frac{B_{ExtBus}}{B_{Queue}}\bar{M}[Queue] - B_{ExtBus}) \leq 0$$

where $B_{Queue} = 3$. Solving this reduced linear programming problem the values obtained for the upper and lower bounds are:

$$1 \leq x_{e\_e\_a} \leq 2$$

5 Conclusions

Operational analysis of timed Petri net models has been introduced. In particular, we have defined adequate observable quantities that allow the derivation of fundamental relations among them. These relations hold true "operationally," i.e., in each sample path that one can measure in an experiment. Among these relations the enabling operational law constitutes a restatement of the classical utilization law (derived in the framework of multiple server queues) for each timed transition of a general Petri net model with infinite server semantics. Bounding inequalities in both directions between throughput of a transition and average marking of its input places have also been derived. These results are typical on network models containing synchronization, and represent a novel result of operational analysis. Under the hypothesis of strong symmetries, analogous relations have been derived for Timed Well-Formed nets.

A direct and interesting application of the obtained operational laws is the computation of performance bounds insensitive to the timing probability distributions. Indeed the bounding technique proposed in this paper guarantees that the exact value of a given performance index falls in the computed interval, whatever its probability distributions is. In this sense this bound technique is substantially more robust with respect to practical application than any performance evaluation technique based on Markovian analysis or simulation (where in any case some hypothesis on the timing distribution must be introduced in order to produce sample execution traces).

Proper linear programming problems including the derived operational laws as constraints allow one to estimate upper and lower bounds for arbitrary linear functions of the throughput and the average marking (in particular, the throughput of a single transition or the average marking of a particular place). This approach constitutes a clear improvement and generalization of previous results valid only for particular net subclasses. An important characteristics of this new method is that it is "open" to the introduction of additional contraints besides the ones already described in this paper provided that they are expressed in linear algebraic form. The straightforward addition of some constraints deriving from a specific knowledge about some peculiar behavioural characteristics of a WN model may improve the quality of the bounds based on results developed for the analysis of the qualitative behaviour of untimed Petri net models.

The proposed method for bounds computation is cheap, since the solution of the LPPs is practically extremely fast in terms of CPU time compared to Markovian numerical analysis (not to mention simulation).

The size of the LPP depends only on the net structure (number of places, transitions and arcs); in particular it is also independent of the cardinality of the basic colour classes, thus adding a dimension on the parameterization of the results. If the computation of bounds for a 4 processor system takes less than 1 second of CPU time, it will take the same order of magnitude to compute bounds for a 1,000 processor system.

References

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

[2] J. Campos, G. Chiola, J.M. Colom, and M. Silva, "Properties and performace bounds for timed marked graphs," *IEEE Trans. on Circ. and Syst. I: Fundamental Th. and App.*, 39(5):386–401, May 1992.

[3] J. Campos, G. Chiola, and M. Silva, "Ergodicity and throughput bounds for Petri nets with unique consistent firing count vector," *IEEE Trans. on Soft. Eng.*, 17(2):117–125, Feb. 1991.

[4] J. Campos, G. Chiola, and M. Silva, "Properties and performance bounds for closed free choice synchronized monoclass queueing networks," *IEEE Trans. on Aut. Cont.*, 36(12):1368–1382, Dec. 1991.

[5] J. Campos, J.M. Colom, and M. Silva, "Improving throughput upper bounds for net based models," In *Proc. of the IMACS-IFAC Symp. Modelling and Control of Tech. Syst.*, pp. 573–582, Lille, May 1991.

[6] G. Chiola, "A graphical Petri net tool for performance analysis," In *Proc. of the* 3^{rd} *Intern. Workshop on Modeling Techniques and Performance Evaluation*, Paris, March 1987.

[7] G. Chiola and C. Anglano, "Linear programming performance bounds for symmetric coloured nets," *Tech. Rep.*, Dip. di Informatica, Univ. di Torino, Feb. 1993.

[8] G. Chiola, C. Dutheillet, G. Franceschinis and S. Haddad, "Stochastic Well-Formed Coloured nets for symmetric modelling applications," *IEEE Trans. on Comp.*, 1993. Accepted for publication.

[9] J.M. Colom and M. Silva, "Improving the linearly based characterization of P/T nets," In G. Rozenberg, ed. *Advances in Petri Nets 1990*, Vol. 483 of *LNCS*, pp. 113–145. Springer-Verlag, Berlin, 1991.

[10] Y. Dallery and X.R. Cao, "Operational analysis of stochastic closed queueing networks," *Performance Evaluation*, 14:43–61, 1992.

[11] P.J. Denning and J.P. Buzen, "The operational analysis of queueing network models," *ACM Computing Surveys*, 10:225–262, 1978.

[12] E. Gelenbe, "Stationary deterministic flows in discrete systems: I," *Theoretical Computer Science*, 3(2):107–127, April 1983.

[13] T. Murata, "Petri nets: Properties, analysis, and applications," *Proceedings of the IEEE*, 77(4):541–580, April 1989.

[14] G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, eds. *Optimization*, Vol. 1 of *Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam, 1989.

[15] M. Silva, "Introducing Petri Nets," Chapter 1 of *Practice of Petri Nets in Manufacturing* (F. Dicesare et al.). Chapman & Hall, 1993.