

An Optimization Approach to Petri Net Monitor Design*

Francesco Basile (†), Pasquale Chiacchio (†), Alessandro Giua (‡)

(†) Dip. di Ing. dell'Informazione e Ing. Elettrica, Università di Salerno, Italy

(‡) Dip. di Ing. Elettrica ed Elettronica, Università di Cagliari, Italy

Abstract

This paper addresses the problem of enforcing generalized mutual exclusion constraints on a Petri net plant. Firstly, we replace the classical partition of the event set into controllable and uncontrollable events from supervisory control theory, by associating a control and observation cost to each event. This leads naturally to formulate the supervisory control problem as an optimal control problem. Monitor places which enforce the constraint are devised as a solution of an integer linear programming problem whose objective function is expressed in terms of the introduced costs. Secondly, we consider timed models for which the monitor choice may lead to performance optimization. If the plant net belongs to the class of mono-T-semiflow nets, we present an integer linear fractional programming approach to synthesize the optimal monitor so as to minimize the cycle time lower bound of the closed loop net. For strongly connected marked graphs the cycle time of the closed loop net can be minimized.

Index Terms: Discrete Event Systems, Supervisory Control, Petri Nets, Monitor Places.

1 Introduction

In this paper we consider discrete event systems modeled by Petri nets (PN) and address the problem of enforcing forbidden state specifications represented by *generalized mutual exclusion constraints* (GMEC) [5, 6, 8]. A GMEC (\mathbf{l}, k) limits the weighted sum of tokens in a subset of places and defines a set of legal markings $\mathcal{M}(\mathbf{l}, k) = \{\mathbf{m} \in \mathbb{N}^m \mid \mathbf{l} \cdot \mathbf{m} \leq k\}$. It was shown [5, 8] that it is possible to impose a GMEC by adding to a net a controller that takes the form of a single place p_c called *monitor*. Following the classical paradigm of supervisory control theory [14], the transitions of a PN may be labeled as *uncontrollable* or *unobservable*. When the monitor has arcs going to uncontrollable (going to and coming from unobservable) transitions we say that the monitor, and its corresponding GMEC, is not *admissible*. Moody and Antsaklis [9] propose an elegant approach to solve a GMEC problem when uncontrollable and/or unobservable transitions are present. Firstly, given a GMEC (\mathbf{l}, k) to be enforced, they propose a parameterization that gives a family of *safe* constraints and monitors. A constraint (\mathbf{l}', k') and its corresponding monitor p'_c is called *safe* if $\mathcal{M}(\mathbf{l}', k') \subseteq \mathcal{M}(\mathbf{l}, k)$: this means that (\mathbf{l}', k') is at least as restrictive as the original constraint and thus p'_c prevents the net from

*Published as: F. Basile, P. Chiacchio, A. Giua, "An optimization approach to Petri net monitor design," *IEEE Trans. on Automatic Control*, Vol. 52, No. 2, February 2007.

reaching any forbidden marking. Secondly in [9] a procedure is given to determine an *admissible* GMEC belonging to this family. This procedure may typically yield several admissible solutions; if the merit function to choose among them is "maximally permissiveness" (as is usually done in supervisory control) these solutions are often incomparable between them [1]. In this paper we propose a similar approach to enforce a GMEC framing it as an optimization problem.

In the first part of the paper, we associate a control and observation cost to each transition. We believe, in fact, that in many cases saying that a transition is not controllable is an oversimplification. It is more correct to say that to make a transition controllable some effort is required (modifying the software of low-level controllers, introduction of new actuators, establishing a network connection between different devices, etc.) and this effort can be quantified. Analogously, the effort to make a transition observable (introduction of sensors, connection of sensors to controllers, software modifications, ecc.) can be evaluated. The introduction of control and observation cost in discrete event systems leads to a new important class of problems. The motivation for this work can be found in related works recently appeared in the literature [11, 13] in the context of automata based supervisors. Here this problem is considered in the context of PN based supervisors.

Thus we consider two functions that associate to each transition t its control and observation costs. If the cost functions only take value in the binary set $\{\epsilon, K\}$, where $\epsilon \ll 1$ and $K \gg 1$ we go back to the controllable/uncontrollable and observable/unobservable case. We show how it is possible to compute, among the safe monitors given by the parameterization of Moody and Antsaklis the one that has minimal cost. We consider two cases.

In the first case, the monitor cost associated to the control and observation of a transition t depends on the number of arcs going to and coming from t . We show that the corresponding optimization problem takes the form of an integer linear programming problem with a linear objective function.

In a second case the cost does not depend on the number of arcs but only on the fact that there exists at least one arc between the monitor place and the transition t . In this case the optimization problem has a non linear objective function; we show, however, that this optimization problem can be re-formulated as a linear one.

The second case has a clear and intuitive interpretation: the cost of detecting or enabling an event is essentially the installation cost of a sensor or an actuator [11, 13]. It may be assumed that there is no extra cost associated with the use of that sensor or actuator. Furthermore, a monitor is usually software-implemented and the tokens in a monitor places represent values of an integer variable in a computer program. Then, if there are more than one arc from a monitor to a transition, the cost of enabling such a transition does not depend on the number of arcs, but only on the cost of installing a connection link between the actuator and the controller.

The first case has a less intuitive motivation. We present it for sake of completeness and also because it helps the formal presentation of the second case. It may have sense if the control and observation actions are associated with physical actions like a material flow or a signal carrying energy. In such a situation, tokens in the control places have a physical meaning and thus if more than one token in a control place is required to enable some transitions, the enabling action of these transitions has a cost proportional to the required number of tokens.

In the second part of the paper we consider another optimization criterion for monitor design. We add a deterministic firing delay to each transition, and assume that the best, among all safe monitors, is the one that minimizes an objective function that depends on the cycle time of the net, assuming a periodic execution of the net exists.

To set up this new optimization criterion, we assume that the plant net belongs to the special class of mono T-semiflow nets: this is a restricted but non trivial class of nets, that includes strongly connected marked graphs and that can be used to model meaningful systems (e.g., kanban manufacturing systems). We show how, using the structural results of [3], it is possible to compute — by solving a integer linear fractional programming problem — the monitor that minimizes a lower bound on the cycle time (if the closed loop net is a marked graph the actual cycle time is minimized). The presented results can be applied to the stochastic case yielding the monitor that minimizes a lower bound on the mean cycle time.

2 Background

2.1 Place/transitions nets

A place/transition (P/T) net is a structure $N = \langle P, T, \mathbf{Pre}, \mathbf{Post} \rangle$ where: P is a set of m places represented by circles; T is a set of n transitions represented by bars; $P \cap T = \emptyset$, $P \cup T \neq \emptyset$; \mathbf{Pre} (\mathbf{Post}) is the $|P| \times |T|$ sized, natural valued, pre-(post-)incidence matrix. For instance, $\mathbf{Pre}(p, t) = w$ ($\mathbf{Post}(p, t) = w$) means that there is an arc from p to t (from t to p) with weight w . The incidence matrix \mathbf{C} of the net is defined as $\mathbf{C} = \mathbf{Post} - \mathbf{Pre}$. A net having all arc weights equal to one is called *ordinary*. A *marking* is a $m \times 1$ vector $\mathbf{m} : P \rightarrow \mathbb{N}$ that assigns to each place of a P/T net a non-negative integer number of tokens. A P/T system or net system $\langle N, \mathbf{m}_0 \rangle$ is a P/T net N with an initial marking \mathbf{m}_0 . A transition $t \in T$ is enabled at a marking \mathbf{m} iff $\mathbf{m} \geq \mathbf{Pre}(\cdot, t)$. If t is enabled, then it may fire yielding a new marking $\mathbf{m}' = \mathbf{m} + \mathbf{C}(\cdot, t)$. The notation $\mathbf{m}[t > \mathbf{m}']$ will mean that an enabled transition t may fire at \mathbf{m} yielding \mathbf{m}' . A *firing sequence* from \mathbf{m}_0 is a (possibly empty) sequence of transitions $\sigma = t_1, \dots, t_k$ such that $\mathbf{m}_0[t_1 > \mathbf{m}_1[t_2 > \mathbf{m}_2 \dots [t_k > \mathbf{m}_k$, and we denote it as $\mathbf{m}_0[\sigma > \mathbf{m}_k$. A marking \mathbf{m} is reachable in $\langle N, \mathbf{m}_0 \rangle$ iff there exists a firing sequence σ such that $\mathbf{m}_0[\sigma > \mathbf{m}$. Given a net system $\langle N, \mathbf{m}_0 \rangle$ the set of reachable markings is denoted $R(N, \mathbf{m}_0)$. The function $\sigma : T \rightarrow \mathbb{N}$, where $\sigma(t)$ represents the number of occurrences of t in σ , is called *firing count vector* of the fireable sequence σ . A net system $\langle N, \mathbf{m}_0 \rangle$ is said to be *bounded* if there exists a nonnegative integer K such that $\mathbf{m}(p) \leq K$ for all $\mathbf{m} \in R(N, \mathbf{m}_0)$ and for all places $p \in P$. Right (left) annuller integer vectors of \mathbf{C} are called *T-semiflows* (*P-semiflows*), i.e. $\mathbf{x} : T \rightarrow \mathbb{N}, \mathbf{x} \neq \mathbf{0}$ (i.e. $\mathbf{y} : P \rightarrow \mathbb{N}, \mathbf{y} \neq \mathbf{0}$) such that $\mathbf{C}\mathbf{x} = \mathbf{0}$ ($\mathbf{y}^T\mathbf{C} = \mathbf{0}$). The *support* of a T-semiflow \mathbf{x} is defined as $\|\mathbf{x}\| = \{t \in T \mid \mathbf{x}(t) > 0\}$. A T-semiflow is said to be *minimal* iff the greatest common divisor of its components is 1 and there exists no other semiflow \mathbf{x}' such that $\|\mathbf{x}'\| \subset \|\mathbf{x}\|$.

Definition 1 A place/transition net is strongly connected if there exists a directed path from any node (place or transition) to any other node (place or transition). ■

Definition 2 A marked graph (MG) is an ordinary P/T net such that each place has a single input arc and a single output arc. ■

The following property is classical.

Property 1 *If the net N is a MG then the following statements are equivalent:*

(a) *it is strongly connected; (b) it is structurally bounded, i.e., $\langle N, \mathbf{m}_0 \rangle$ is bounded for any initial marking M_0 and its unique minimal T-semiflow is the vector $\mathbf{1}$.* ■

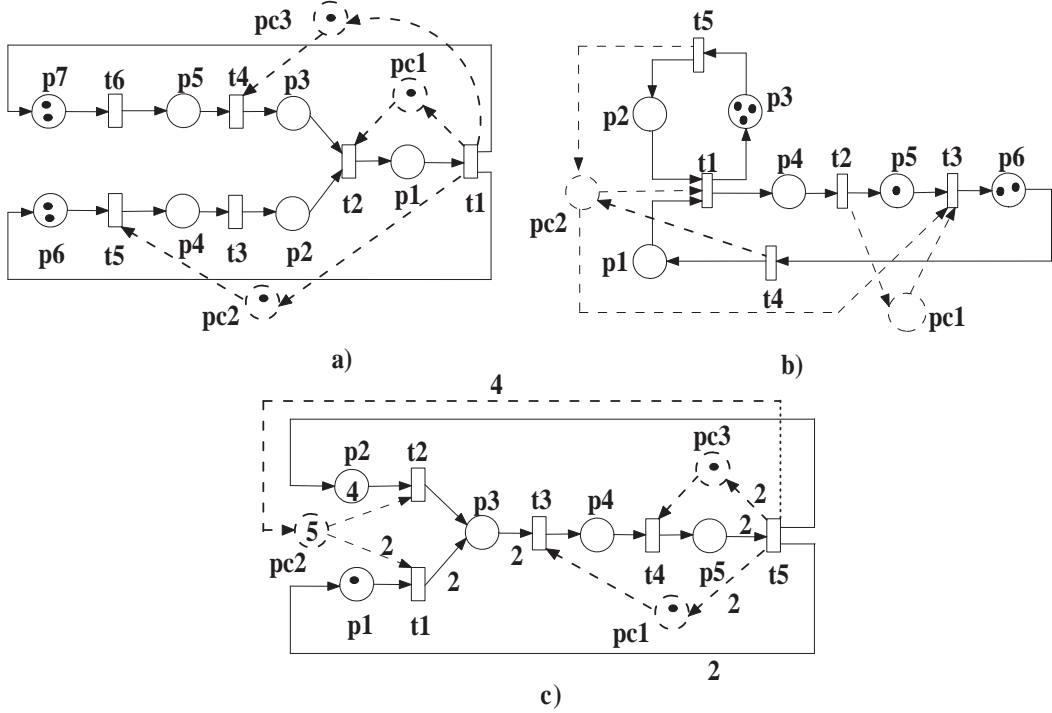


Figure 1: (a) Net system used in Examples 1 and 4; (b) net system used in Example 2; (c) net system used in Example 3. Monitor places and arcs are dashed.

2.2 Monitor approach

Assume we are given a set of legal markings $\mathcal{L} \subseteq \mathbb{N}^m$, and consider the basic control problem of designing a supervisor that restricts the reachability set of the plant in closed loop to $\mathcal{L} \cap R(N, \mathbf{m}_0)$. Of particular interest are those PN state-based control problems where the set of legal markings \mathcal{L} is expressed by a set of n_c linear inequality constraints called Generalized Mutual Exclusion Constraints. A single GMEC is a couple (\mathbf{l}, k) where $\mathbf{l} : P \rightarrow \mathbb{Z}$ is a $1 \times m$ weight vector and $k \in \mathbb{Z}$. Given the net system $\langle N, \mathbf{m}_0 \rangle$, a GMEC defines a set of markings that will be called *legal markings*: $\mathcal{M}(\mathbf{l}, k) = \{\mathbf{m} \in \mathbb{N}^m \mid \mathbf{l}\mathbf{m} \leq k\}$. The markings that are not legal are called *forbidden markings*. A controlling agent, called supervisor, must ensure the forbidden markings will not be reached. So the set of legal markings under control is $\mathcal{M}_c(\mathbf{l}, k) = \mathcal{M}(\mathbf{l}, k) \cap R(N, \mathbf{m}_0)$. It has been shown [8] that the Petri net controller that enforces (\mathbf{l}, k) is a place p_c called *monitor* with incidence matrix $\mathbf{c}_c \in \mathbb{Z}^{1 \times n}$ given by

$$\mathbf{c}_c = -\mathbf{l}\mathbf{C}_p \quad (1)$$

where \mathbf{C}_p is the incidence matrix of the plant. The initial marking of the monitor, denoted as $m_{c0} \in \mathbb{N}$, is given by

$$m_{c0} = k - \mathbf{l}\mathbf{m}_{p0} \quad (2)$$

where $\mathbf{m}_{p0} \in \mathbb{N}^{m \times 1}$ is the initial marking of the plant. The controller exists iff the initial marking is a legal marking, i.e. $k - \mathbf{l}\mathbf{m}_{p0} \geq \mathbf{0}$. By definition a monitor is loop-free¹, thus its post- and pre- incidence matrix \mathbf{c}_c^+ and \mathbf{c}_c^- such that $\mathbf{c}_c = \mathbf{c}_c^+ - \mathbf{c}_c^-$ can be uniquely defined as:

$$\mathbf{c}_c^+ = \max\{\mathbf{c}_c, \mathbf{0}\} \quad \text{and} \quad \mathbf{c}_c^- = \max\{-\mathbf{c}_c, \mathbf{0}\} \quad (3)$$

The monitor so constructed is maximally permissive, i.e. it prevents only transitions firings that yield forbidden markings. It has been shown that it is possible to transform a GMEC (\mathbf{l}, k) into a more restrictive GMEC (\mathbf{l}', k') as shown in the following proposition.

Proposition 1 (Moody and Antsaklis [9]) *Given a plant $\langle N, \mathbf{m}_{p0} \rangle$ with incidence matrix \mathbf{C}_p and a GMEC (\mathbf{l}, k) , let $\mathbf{r}_1 \in \mathbb{N}^{1 \times m}$ and $r_2 \in \mathbb{N}$ be such that $\mathbf{r}_1\mathbf{m}_{p0} + r_2\mathbf{l}\mathbf{m}_{p0} - r_2(k+1) \leq -1$. Consider the transformed GMEC (\mathbf{l}', k') with $\mathbf{l}' = \mathbf{r}_1 + r_2\mathbf{l}$, $k' = r_2(k+1) - 1$. Then, it holds $\mathcal{M}(\mathbf{l}', k') \subseteq \mathcal{M}(\mathbf{l}, k)$. The corresponding monitor has incidence matrix and initial marking given by $\mathbf{c}_c = -\mathbf{l}'\mathbf{C}_p$, $m_{c0} = k' - \mathbf{l}'\mathbf{m}_{p0}$, and the initial marking of the plant is legal with respect to (w.r.t.) to the transformed constraint. This parameterization in terms of \mathbf{r}_1 and r_2 is called Moody & Antsaklis' parameterization and the corresponding monitors are called safe w.r.t. (\mathbf{l}, k) .*

■

3 Monitor design with control and observation cost

Assume we are given a function² $z_c : T \rightarrow \mathbb{R}^+$ which associates a nonnegative control cost to each transition and a function $z_o : T \rightarrow \mathbb{R}^+$ which associates a nonnegative observation cost to each transition. Our problem consists in choosing, among the set of all monitors that are safe w.r.t. to a given GMEC (\mathbf{l}, k) , the one that minimizes an objective function representing the cost of the monitor based control net structure.

Proposition 2 *Consider a plant $\langle N, \mathbf{m}_{p0} \rangle$ with incidence matrix \mathbf{C}_p . Given a GMEC (\mathbf{l}, k) , the set of monitors that are safe w.r.t. it has, respectively, incidence matrix and initial marking $\mathbf{c}_c = \mathbf{c}_c^- - \mathbf{c}_c^+$, $m_{c0} = r_2(k+1) - 1 - (\mathbf{r}_1 + r_2\mathbf{l})\mathbf{m}_{p0}$ obtained by solving the set of equations*

$$\left\{ \begin{array}{l} (a) \quad \mathbf{r}_1\mathbf{C}_p + r_2\mathbf{l}\mathbf{C}_p = \mathbf{c}_c^- - \mathbf{c}_c^+ \\ (b) \quad \mathbf{r}_1\mathbf{m}_{p0} + r_2(\mathbf{l}\mathbf{m}_{p0} - (k+1)) \leq -1 \\ (c) \quad \mathbf{c}_c^- \geq \mathbf{0}_{1 \times n} \\ (d) \quad \mathbf{c}_c^+ \geq \mathbf{0}_{1 \times n} \\ (e) \quad \mathbf{r}_1 \geq \mathbf{0}_{1 \times m} \\ (f) \quad r_2 \geq 1 \end{array} \right. \quad (4)$$

with variables $\mathbf{r}_1 \in \mathbb{N}^{1 \times m}$, $r_2 \in \mathbb{N}$, $\mathbf{c}_c^- \in \mathbb{N}^{1 \times n}$, $\mathbf{c}_c^+ \in \mathbb{N}^{1 \times n}$.

Proof: Equations (4-a,c,d,e,f) impose that the incidence matrix of the controller is obtained from Moody & Antsaklis' parameterization: $\mathbf{l}'\mathbf{C}_p = \mathbf{c}_c = \mathbf{c}_c^- - \mathbf{c}_c^+$, with $\mathbf{l}' = \mathbf{r}_1 + r_2\mathbf{l}$. Equation (4-b) imposes that the initial marking condition is verified ($\mathbf{l}'\mathbf{m}_{p0} \leq k'$). □

Note that the solution of the previous system does not necessarily satisfies the condition $\mathbf{c}_c^-(p)\mathbf{c}_c^+(p) = 0$. However a (loop-free) monitor with incidence matrix $\mathbf{c}_c = \mathbf{c}_c^- - \mathbf{c}_c^+$ can always be obtained redefining its pre- and post- matrices as in eq. (3).

¹A transition t cannot be at same time input and output transition of a monitor.

²Here \mathbb{R}^+ denotes the set of nonnegative real numbers.

It is useful to represent the system of equations (4) in the form

$$\begin{cases} (a) & [\mathbf{C}_p^T \quad \mathbf{C}_p^T \mathbf{l}^T \quad -\mathbf{I}_n \quad \mathbf{I}_n] \mathbf{y} = \mathbf{0}_{n \times 1} \\ (b) & [\mathbf{m}_{p0}^T \quad (\mathbf{m}_{p0}^T \mathbf{l}^T - (k+1)) \quad \mathbf{0}_{1 \times 2n}] \mathbf{y} \leq -1 \\ (c) & \mathbf{I}_{m+1+2n} \mathbf{y} \geq [\mathbf{0}_{m \times 1} \quad 1 \quad \mathbf{0}_{2n \times 1}]^T \end{cases} \quad (5)$$

where $\mathbf{y} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{c}_c^- \quad \mathbf{c}_c^+]^T$, \mathbf{I}_n denotes the identity matrix of dimension n and $\mathbf{0}_{m \times n}$ denotes the zero matrix of dimension $m \times n$.

Definition 3 Let us denote by $\mathcal{F}(\mathbf{y})$ the set of the natural valued vectors that are solutions of the system of equations (5) with variables $\mathbf{y} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{c}_c^- \quad \mathbf{c}_c^+]^T$. ■

3.1 First case: a linear controller cost

In this first case, we assume that the cost associated to the control and observation of a transition t depends on the number of arcs going to and coming from t .

In this case if a monitor place p_c has an arc outgoing to a plant transition t with weight $\mathbf{c}_c^-(p_c, t)$, we define $\mathbf{c}_c^-(p_c, t)z_c(t)$ the cost of disabling a firing of the transition; so, if a monitor has an input arc from a plant transition with weight $\mathbf{c}_c^+(p_c, t)$, we define $\mathbf{c}_c^+(p_c, t)z_o(t)$ the cost of detecting a firing of this transition. Thus the optimal monitor can be found by solving the following integer linear programming (ILP) problem:

$$\begin{aligned} \min \Delta &= \mathbf{c}_c^- z_c + \mathbf{c}_c^+ z_o = \mathbf{z}^T \mathbf{y} \\ \text{s.t. } &\mathbf{y} \in \mathcal{F}(\mathbf{y}) \end{aligned} \quad (6)$$

where $\mathbf{z} = [\mathbf{0}_{m \times 1} \quad 0 \quad z_c \quad z_o]^T$.

Example 1 Let us consider the GMEC (\mathbf{l}, k) with $\mathbf{l} = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$, $k = 1$ and the net system in fig. 1-a. If we assume all transitions are observable and controllable, this GMEC can be enforced by the monitor p_{c1} , determined applying (1) and (2). Let us introduce the control and observation costs: $z_c = [1 \quad 6 \quad 5 \quad 3 \quad 2 \quad 3]^T$, $z_o = [1 \quad 3 \quad 4 \quad 4 \quad 3 \quad 2]^T$. If we adopt p_{c1} we have $\Delta = 7$. The optimal monitor can be computed solving ILP (6). It results to be p_{c2} with a control cost $\Delta^* = 3$. ■

We have noted that from a solution of ILP (4) a monitor (i.e., a loop-free controller) can always be obtained redefining its pre- and post- matrices as in eq. (3). However, this may change the objective function of ILP (6) that depends on the value of \mathbf{c}_c^- and \mathbf{c}_c^+ . The following property shows that an optimal solution of ILP (6) can always be implemented by a monitor.

Property 2 If ILP (6) admits a solution, then it also admits an optimal solution that verifies the condition $\mathbf{c}_c^{-*}(p)\mathbf{c}_c^{+*}(p) = 0$, $\forall p \in P$, i.e., there exists an optimal controller that is loop-free.

Proof: Suppose that $\exists p \in P$, $\mathbf{c}_c^{-*}(p)\mathbf{c}_c^{+*}(p) \neq 0$ and without loss of generality that $\mathbf{c}_c^{-*}(p) \geq \mathbf{c}_c^{+*}(p)$. Now let us build a new solution $\mathbf{c}_c^{-'}(p) = \mathbf{c}_c^{-*}(p) - \mathbf{c}_c^{+*}(p)$, $\mathbf{c}_c^{+'}(p) = 0$. It is immediate to verify that the (4-a,b,c,d,e,f) are verified and that $\Delta' = \Delta^* - z_c(p)\mathbf{c}_c^{-*}(p) - z_o(p)\mathbf{c}_c^{+*}(p) \leq \Delta^*$ hence the new solution is optimal. □

With a similar reasoning, it is easy to show that such a property also holds for all other optimizations problems presented in the rest of this note.

We remark that the monitor synthesis proposed in this paper is based on Moody and Antsaklis parameterization which has been devised on the basis of structural PN theory in order to avoid

the computation of reachability set. It may well happen that a transition is never plant enabled when disabled by a monitor and this may lead to a different notion of cost as shown in Example 2.

Example 2 Let us consider the GMEC (l, k) with $l = [-1 \ -1 \ 0 \ -1 \ -1 \ 0]$, $k = -1$ and the net system in fig. 1-b without dashed places and arcs. Let us introduce the control and observation costs: $z_c = [50 \ 1 \ 1 \ 1 \ 1]^T$, $z_o = [1 \ 10 \ 1 \ 1 \ 1]^T$. The monitor p_{c1} in fig. 1-b is obtained from a Moody parameterization with $r_1 = [1 \ 1 \ 0 \ 1 \ 0 \ 0]$ and $r_2 = 1$, and its control cost is equal to 11. The monitor p_{c2} in fig. 1-b is obtained from (1) and (2) and has a control cost $\Delta = 53$. However, since p_{c2} never disables t_1 whose control cost is equal to 50, one may argue that the control cost of t_1 should not be considered and thus the effective cost of p_{c2} is equal to 3 that is smaller than the cost of p_{c1} . ■

3.2 Second case: a nonlinear controller cost

In many cases the control and observation cost of a transition is just the cost of the device and its installation in order to perform these actions (sensor, network connection, etc.). In this case when adding a monitor p_c , the cost of controlling (resp., observing) a given transition t does not depend on the number of arcs from p_c to t (resp., from t to p_c). In this case the optimal monitor can be found by solving the following integer programming (IP) problem with non linear objective function³:

$$\begin{aligned} \min \Delta_{nl} &= \text{sign}(\mathbf{c}_c^-)z_c + \text{sign}(\mathbf{c}_c^+)z_o \\ \text{s.t. } \mathbf{y} &\in \mathcal{F}(\mathbf{y}). \end{aligned} \quad (7)$$

The *sign* function allows one to consider only the control or the observation cost without taking into account the weights of arcs from or to control places. The integer non linear programming problem (7) can be transformed into a integer linear programming problem (ILP) by introducing two auxiliary vectors $\mathbf{q}_c \in \{0, 1\}^{1 \times n}$ and $\mathbf{q}_o \in \{0, 1\}^{1 \times n}$ associated respectively to \mathbf{c}_c^- and \mathbf{c}_c^+ .

Definition 4 Let us denote by $\mathcal{F}'(\mathbf{y}')$ the set of vectors that are solutions of the system of equations obtained by adding to Equations (5a,b,c) the following ones

$$\begin{cases} (d) \ \mathbf{c}_c^- + W(\mathbf{1} - \mathbf{q}_c) \leq W \mathbf{1} \\ (e) \ \mathbf{c}_c^+ + W(\mathbf{1} - \mathbf{q}_o) \leq W \mathbf{1} \end{cases} \quad (8)$$

with variables $\mathbf{y}' = [r_1 \ r_2 \ \mathbf{c}_c^- \ \mathbf{c}_c^+ \ \mathbf{q}_c \ \mathbf{q}_o]^T$ and where W is a positive integer that should be chosen greater than the expected values of all elements of \mathbf{c}_c^- and \mathbf{c}_c^+ . ■

Property 3 Let us consider the ILP

$$\begin{aligned} \min \Delta &= \mathbf{q}_c z_c + \mathbf{q}_o z_o = \mathbf{z}'^T \mathbf{y}' \\ \text{s.t. } \mathbf{y}' &\in \mathcal{F}'(\mathbf{y}') \end{aligned} \quad (9)$$

with $\mathbf{z}' = [\mathbf{0}_{m \times 1} \ 0 \ \mathbf{0}_{n \times 1} \ \mathbf{0}_{n \times 1} \ z_c \ z_o]^T$, it gives the same optimal monitor of the IP (7).

Proof: Equations (8d-e) impose that an element of the auxiliary vector $\mathbf{q}_c(p)$ ($\mathbf{q}_o(p)$) has to be equal to one if $\mathbf{c}_c^-(p) > 0$ ($\mathbf{c}_c^+(p) > 0$), while $\mathbf{q}_c(p)$ ($\mathbf{q}_o(p)$) may be equal to zero or one if

³Given a \mathbf{A} $m \times n$ matrix of positive integers, we assume that $\text{sign}(\mathbf{A}(i, j)) = 0$ if $\mathbf{A}(i, j)$ is 0, else $\text{sign}(\mathbf{A}(i, j)) = 1$.

$\mathbf{c}_c^-(p) = 0$ ($\mathbf{c}_c^+(p) = 0$). Since the cost of the objective function has to be minimized $\mathbf{q}_c(p)$ ($\mathbf{q}_o(p)$) is selected equal to zero when $\mathbf{c}_c^-(p) = 0$ ($\mathbf{c}_c^+(p) = 0$). It follows the proof. \square

Example 3 Let us consider the GMEC (\mathbf{l}, k) with $\mathbf{l} = [0 \ 0 \ 0 \ 0 \ 1]$, $k = 2$ and the net system in fig. 1-c. Let $\mathbf{z}_c = [1 \ 2 \ 4 \ 5 \ 1]^T$, $\mathbf{z}_o = [3 \ 2 \ 2 \ 2 \ 1]^T$ be the control and observation costs of the transitions. Applying the ILP (4) to this system, the optimal monitor results p_{c1} with a control cost $\Delta^* = 6$. While, if we solve (7), we obtain that the optimal monitor in this case is the one labeled p_{c2} in the figure, and $\Delta_{nl}^* = 4$. Note that p_{c3} , simply obtained from (1) and (2), has a cost $\Delta_{nl} = 6$. \blacksquare

4 Optimal monitor design for timed Petri nets

Adding time to transitions a further criterion to select the suboptimal monitor could be the optimization of the cycle time lower bound of the closed loop net.

4.1 Deterministic Timed nets

In *deterministic timed PN* [10] we suppose that there is a delay of at least d_i units of time associated with the firing of transition t_i , $i = 1, \dots, n$; the delay may be greater than d_i units of time depending on the firing policy. This means that when t_i is enabled, a number of $\mathbf{Pre}(p_j, t_i)$ tokens will be reserved in the place p_j for at least d_i units of time before their removal by firing t_i . We are interested in finding how fast each transition can initiate firing in a periodically operated timed Petri net, where a period Γ is defined as the time to complete a *stationary* firing sequence (i.e., a sequence that leads back to the initial marking) after firing each transition at least once. Γ is called *cycle time* (CT) of the net system. It is well know that a firing sequence is stationary if and only if its firing count vector is a T-semiflow. Thus, it only makes sense to speak of CT for a *consistent* net - a net that admits a T-semiflow containing all the transitions, i.e. $\exists \mathbf{x} \in \mathbb{N}^n$ such that $\mathbf{x} > \mathbf{0}$ and $\mathbf{C}\mathbf{x} = \mathbf{0}$. We denote $\sigma[\tau]$ the firing sequence at time τ and we define the limit *firing count vector per time unit* $\bar{\sigma} = \lim_{\tau \rightarrow \infty} \sigma[\tau]/\tau$. We say that the firing process of a net system is weakly ergodic, if such limit exists. In this paper we consider any firing policy provided that the firing process is weakly ergodic and in this case the average time between two consecutive firings of a selected transition t_i , (CT of t_i) is defined as $\Gamma_i = 1/\bar{\sigma}(t_i)$.

Proposition 3 ([3]) *Given a deterministic timed strongly connected MG we have that*

1. $\Gamma_i = \Gamma, \forall t_i \in T$;
2. *the lower bound of the CT, denoted Γ_{min} , can be computed by the following fractional programming problem*

$$\Gamma_{min} = \max_{\mathbf{y}} \frac{\mathbf{y}^T \cdot \mathbf{Pre} \cdot \mathbf{d}}{\mathbf{y}^T \cdot \mathbf{m}_0} \quad (10)$$

where \mathbf{y} is a P-semiflow, that can be reduced to the following linear programming problem:

$$\Gamma_{min} = \max_{\mathbf{y}} \mathbf{y}^T \cdot \mathbf{Pre} \cdot \mathbf{d}$$

$$s.t. \begin{cases} \mathbf{y}^T \cdot \mathbf{C} = \mathbf{0} \\ \mathbf{y}^T \cdot \mathbf{m}_0 = 1 \\ \mathbf{y} \geq \mathbf{0} \end{cases} \quad (11)$$

where $\mathbf{d}(t_i)$ is the time delay of transition t_i ;

3. $\Gamma = \Gamma_{min}$ if each transition fires as soon as it is enabled – earliest firing policy (e.f.p). ■

We recall that in the case of MGs each minimal P-semiflow corresponds to an elementary circuit. In the system (10) \mathbf{y} is a P-semiflow, thus $\mathbf{Pre}^T \mathbf{y}$ is the characteristic vector (but for a scalar factor) of the transitions along the circuit and, finally, $\mathbf{y}^T \cdot \mathbf{Pre} \cdot \mathbf{d}$ is the sum of the time delay of all transitions along the circuit. Thus, an interpretation of the system (10) is that the CT can be computed looking at the slowest subsystem generated by the P-semiflows [4], considered in isolation w.r.t. delay nodes, where the CT of each subsystem can be computed making the summation of the time delays of all the transitions involved in it, and dividing by the tokens present in it (i.e. the division by $\mathbf{y}^T \cdot \mathbf{m}_0$). This important result allows one to compute the CT of deterministic strongly connected MGs.

It is also possible to generalize this result: let us first introduce a class of nets with a unique consistent firing vector.

Definition 5 *A structurally bounded net N is called a mono-T-semiflow net if it admits a unique minimal T-semiflow, and this semiflow contains all transitions.* ■

Obviously, a mono-T-semiflow net is a generalization of a strongly connected MG. For the class of mono-T-semiflow nets, we speak of CT of a certain transition since in order to complete a net system cycle each transition has to fire a different number of times. Note that, if we optimize the CT of a transition t_i , the CT of other transitions is optimized since it is scaled by a constant factor.

Proposition 4 ([3]) *Given a mono-T-semiflow net let \mathbf{x} be its unique minimal T-semiflow. Consider a solution of the LPP (11) changing the objective function to*

$$\Gamma_{min} = \max_{\mathbf{y}} \mathbf{y}^T \cdot \mathbf{Pre} \cdot \mathbf{D},$$

where for all $j = 1, \dots, n$: $\mathbf{D}(t_j) = \mathbf{d}(t_j)\mathbf{x}(t_j)$. It holds that $\Gamma_i \geq \Gamma_{min}/\mathbf{x}(t_i)$. ■

Note that for mono T-semiflow nets the lower bound may not be attainable under any firing policy. Moreover, if a mono T-semiflow net is not persistent ⁴, the lower bound may be finite even if the net system is not live, thus it may not be a good approximation of the transitions CT for some mono-T-semiflow nets.

4.2 Optimal monitor design for deterministic timed nets

Let us first recall two classical results of monitor controlled net.

Proposition 5 *Consider a PN where a monitor corresponding to the GMEC (\mathbf{l}, k) has been added:*

⁴A net system is said to be persistent if it never occurs that, if two transitions are both enabled, the firing of a transition disables the other one.

- a) \mathbf{x} is a T -semiflow of the plant net iff it is a T -semiflow of the closed loop net;
- b) the closed loop net has all the P -semiflows of the plant plus the vectors that are multiple of the P -semiflow $[\mathbf{l} \ 1]$. ■

Proposition 5a) is classical and Proposition 5b) has been proved in [8]. By Proposition 5a) it follows that, if the plant net is mono T -semiflow, the closed loop one is also mono- T -semiflow. In this section we only consider the case of open loop nets that are *mono- T -semiflow*.

Proposition 6 *Let consider the problem of imposing a GMEC (\mathbf{l}, k) on a timed mono- T -semiflow system with CT of a transition $\Gamma_i \geq \Gamma_p/\mathbf{x}(t_i)$ where \mathbf{x} is the unique minimal T -semiflow. A necessary condition to find a safe monitor that does not increase the CT of a transition is that the following system of equations has a solution:*

$$\left\{ \begin{array}{l} (a) \mathbf{y} \in \mathcal{F}(\mathbf{y}) \\ (b) \begin{bmatrix} \mathbf{D}^T \mathbf{Pre}^T & \mathbf{D}^T \mathbf{Pre}^T \mathbf{l}^T \\ -\Gamma_p(k+1) & \mathbf{D}^T \mathbf{O}_{1 \times n} \end{bmatrix} \mathbf{y} \leq -\Gamma_p \end{array} \right. \quad (12)$$

If the closed loop net is a strongly connected MG⁵ under e.f.p., the condition is sufficient.

Proof: Constraints (a) impose that the incidence matrix of the controller is obtained from Moody & Antsaklis' parameterization. Constraint (b) is equivalent to

$$\frac{[\mathbf{r}_1 + r_2 \mathbf{l} \ 1] \begin{bmatrix} \mathbf{Pre} \\ \mathbf{c}_c^- \end{bmatrix} \mathbf{D}}{[\mathbf{r}_1 + r_2 \mathbf{l} \ 1] \begin{bmatrix} \mathbf{m}_{p0} \\ m_{c0} \end{bmatrix}} = \frac{[\mathbf{r}_1 + r_2 \mathbf{l} \ 1] \begin{bmatrix} \mathbf{Pre} \\ \mathbf{c}_c^- \end{bmatrix} \mathbf{D}}{r_2(k+1) - 1} \leq \Gamma_p$$

where $[\mathbf{r}_1 + r_2 \mathbf{l} \ 1]$ is the new P -semiflow added by the monitor, $\begin{bmatrix} \mathbf{Pre} \\ \mathbf{c}_c^- \end{bmatrix}$ is the pre-incidence matrix of the closed loop net and $r_2(k+1) - 1$ represents the weighted sum of tokens contained in the new P -semiflow. Thus, this constraint imposes that the P -semiflow subnet introduced by the monitor has a CT lower bound less or equal than the plant net CT; this is a necessary condition to make the closed loop net not slower than the plant net because of Proposition 4.

If the closed loop net is a strongly connected MG under e.f.p., Equation (12-b) guarantees that the actual CT is not increased because of Proposition 3. □

If the set of constraints (12) has a solution, one may also use an objective function as in ILP (6) or (7) to find, among all monitors whose CT lower bound does not exceed the plant CT, one that is optimal w.r.t. the control and observation cost. On the other hand, if the system of equations (12) does not have a solution, it is necessary to make the closed loop net slower in order to impose the GMEC (\mathbf{l}, k) . To find a monitor with minimal CT lower bound one could solve the system of equations (12) by replacing Γ_p with a value $\Gamma'_p > \Gamma_p$ and keep increasing the value of Γ'_p until a solution is found. If a monitor based solution w.r.t. Moody and Antsaklis parameterization exists, i.e. system (5) admits a solution, its associated P -semiflow subnet has a finite CT lower bound and, thus, it is possible to conclude that a finite value Γ'_p also exists such that system (12) admits a solution.

A more direct approach can be taken, by using a *fractional* objective function. This complicates the optimization problem but there exist tools and techniques to solve such a problem [7].

⁵The addition of a monitor to a strongly connected MG, that is a special mono- T -semiflow net subclass, leads to a closed loop net that is a strongly connected MG if the monitor has only one input and output arc.

Proposition 7 *Let us consider the problem to impose a GMEC (\mathbf{l}, k) on the timed mono-T-semiflow system $\langle N, \mathbf{m}_0 \rangle$. A safe monitor with minimal CT lower bound can be obtained by solving the following integer linear fractional programming (ILFP) problem*

$$\begin{aligned} \min \Delta_t &= \frac{[\mathbf{D}^T \mathbf{Pre}^T \quad \mathbf{D}^T \mathbf{Pre}^T \mathbf{l}^T \quad \mathbf{D}^T \quad \mathbf{O}_{1 \times n}] \mathbf{y}}{[\mathbf{O}_{1 \times m} \quad k+1 \quad \mathbf{O}_{1 \times n} \quad \mathbf{O}_{1 \times n}] \mathbf{y} - 1} \\ \text{s.t. } \mathbf{y} &\in \mathcal{F}(\mathbf{y}). \end{aligned} \quad (13)$$

with $\mathbf{D}(t) = \mathbf{d}(t)\mathbf{x}(t)$ and $\mathbf{d}(t)$ is the delay associated to transition t and \mathbf{x} is the unique minimal T-semiflow of the net. If the closed loop net is a strongly connected MG under e.f.p. the safe monitor obtained from the ILFP problem (13) has a minimal actual CT.

Proof: The constraint $\mathbf{y} \in \mathcal{F}(\mathbf{y})$ imposes that the incidence matrix of the controller is obtained from Moody & Antsaklis' parameterization. The objective function

$$\begin{aligned} \Delta_t &= \frac{[\mathbf{D}^T \mathbf{Pre}^T \quad \mathbf{D}^T \mathbf{Pre}^T \mathbf{l}^T \quad \mathbf{D}^T \quad \mathbf{O}_{1 \times n}] \mathbf{y}}{[\mathbf{O}_{1 \times m} \quad k+1 \quad \mathbf{O}_{1 \times n} \quad \mathbf{O}_{1 \times n}] \mathbf{y} - 1} = \\ &= \frac{[\mathbf{r}_1 + r_2 \mathbf{l} \quad 1] \begin{bmatrix} \mathbf{Pre} \\ \mathbf{c}_c^- \end{bmatrix} \mathbf{d}}{[\mathbf{r}_1 + r_2 \mathbf{l} \quad 1] \begin{bmatrix} \mathbf{m}_{p0} \\ m_{c0} \end{bmatrix}} \end{aligned}$$

represents the CT lower bound of the subnet introduced by a monitor. If the closed loop net is a strongly connected MG under e.f.p., the safe monitor has a minimal actual CT because of Proposition 3. \square

The closed loop net, once that the monitor has been added, may be not live, since the existence of a finite CT lower bound does not imply liveness in a mono-T-semiflow net. This is true only if the closed loop net is a persistent mono-T-semiflow (e.g. strongly connected MGs) as recalled in Subsection 4a. Some conditions presented in [2] can be used to check liveness of a net controlled by a monitor if the plant net is a MG, otherwise liveness of the closed loop net has to be checked with classical techniques [12].

We also remark that the integrality constraint on the decision variables does not allow one to rewrite ILFP problem (13) in a ILP form, using the same technique used to rewrite (10) as (11).

Example 4 *Let us consider again the MG net system in fig. 1-a, the GMEC (\mathbf{l}, k) with $\mathbf{l} = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$, $k = 1$ and the following control (observation) costs, CT unit cost and the time delays for the transitions: $\mathbf{z}_c = [1 \ 10 \ 7 \ 8 \ 2 \ 8]^T$, $\mathbf{z}_o = [1 \ 3 \ 4 \ 4 \ 3 \ 2]^T$, $z_\Gamma = 1$, $\mathbf{d} = [1 \ 1 \ 3 \ 1 \ 3 \ 1]^T$. By solving (11) it results $\Gamma_p = 4$. If we adopt p_{c1} , simply obtained from (1) and (2), we have $\Delta = 11$; notice that the CT of the closed loop net remains equal to 4. By optimizing Δ w.r.t. the set of constraints (12) the optimal monitor place, that does not increase the closed loop net CT, results p_{c3} with a cost $\Delta^* = 9$. We remark that, since by adding p_{c1} or p_{c3} the closed loop net is still a MG, we have referred to CT and not to CT lower bound in this example. \blacksquare*

4.3 Generalization of the approach

The approach presented in this section can be generalized to include stochastic transition timing and to extend these results to more general net subclasses.

In *stochastic PN* we suppose that a random process is associated to each transition. Such random variable, called *service time*, represents the delay associated with the firing of the transition. As it has been shown in [3], if we denote $d(t_i)$ the mean value of service time associated to the transition t_i the bound obtained from Proposition 4 can be interpreted as a bound of the mean CT. For stochastic strongly connected MGs it cannot be improved only on the basis of the mean and variance of transition service times, but moments of order greater than two of the service time random variables are needed. Hence, it is a good approximation of the CT also for stochastic strongly connected MGs. For mono-T-semiflow nets the lower bound may not be attainable under any probability distribution function of service times [3]. Thus, in presence of stochastic PN the results presented in Section IVb, that have been derived from Proposition 4, can be interpreted in terms of mean CT lower bound optimization.

It is possible to extend the results presented in this section to more general net system when a not unique minimal T-semiflow exists. In the programming problem (13) the unique minimal T-semiflow is replaced with a T-semiflow obtained by imposing that its elements satisfy the routing rates of the conflicting transitions [4]. The computation of such a T-semiflow requires, if the net is a live and bounded free choice net, the resolution of a linear system of equations that depends only on the net structure and routing rates and so it is not computation demanding. On the contrary, for general net system a major computational effort is required.

5 Conclusions

In this paper we have dealt with the control of Petri Nets and we have shown how it is possible to generalize the classical notion of uncontrollable/unobservable transition introducing the notion of control and observation costs. We have shown how the problem of enforcing a GMEC so as to minimize the control and observation cost can be framed as an integer linear programming problem. For plant modelled by timed Petri net it is also possible to use a similar approach to optimize the cycle time of the closed loop net (or at least a lower bound on it). The results are valid for plants modelled by mono T-semiflow nets. This approach can be generalized to include stochastic transition timing and to extend these results to more general net subclasses.

Acknowledgement

We are grateful to an anonymous reviewer for suggesting Example 2.

References

- [1] F. Basile, P. Chiacchio and A. Giua “Suboptimal supervisory control of petri nets in presence of uncontrollable transitions via monitor places”, *Automatica*, vol. 42, pp. 995-1004, 2006.
- [2] F. Basile, L. Recalde, P. Chiacchio, M. Silva, *Closed-loop live Petri net supervisors for generalized mutual exclusion constraints*, 5th International Workshop on Discrete Event

- Systems (WODES 00), Gent, B, Aug. 2000, in *Discrete Event Systems: Analysis and Control*, R. Boel, G. Stremesch (Eds.), Kluwer Academic Publishers, pp. 169-180, 2000.
- [3] J. Campos, G. Chiola and M. Silva, "Ergodicity and throughput bounds of petri nets with unique consistent firing count vector.", *IEEE Trans. On Software Engineering*, vol.17, no. 2, pp. 117-125, 1991.
- [4] J. Campos and M. Silva, "Structural techniques and performance bounds of stochastic petri net models.", In *Advances in Petri Nets 1992*, no. 609 of LNCS, pp. 352-391, 1992, Springer Verlag, Berlin.
- [5] A. Giua, F. Di Cesare, M. Silva, "Generalized Mutual Exclusion Constraints on Nets with Uncontrollable Transitions", *Proc. 1992 IEEE Int. Conf. on Systems, Man, and Cybernetics (Chicago)*, pp. 974-979, October 1992.
- [6] B. H. Krogh and L. E. Holloway, "Synthesis of Feedback Control Logic for Discrete Manufacturing Systems", *Automatica*, vol. 27, no. 4, pp. 614-651, January 1991.
- [7] I.M. Stancu-Minasian, "Fractional Programming: Theory, Methods and Applications", Kluwer Academic Publishers, 1997.
- [8] J.O. Moody, K. Yamalidou, M.D. Lemmon and P.J. Antsaklis, "Feedback Control of Petri Nets Based on Place Invariants", *Automatica*, vol. 32, no. 1, pp. 15-28, January 1996.
- [9] J.O. Moody and P.J. Antsaklis, "Petri net supervisors for DES with uncontrollable and unobservable transitions," *IEEE Trans. on Automatic Control*, vol. 45, no. 3, pp. 462-476, March 2000.
- [10] T. Murata, "Petri nets: properties, analysis and applications.", *Proc. IEEE*, vol.77, no. 4, pp. 541-580, 1989.
- [11] K.R. Rohloff, S. Khuller, G. Kortsarz, "Approximating the Minimal Sensor Selection for Supervisory Control", *Discrete Event Dynamical Systems*, vol. 16, pp. 143-170, 2006.
- [12] L. Recalde, E. Teruel, and M. Silva, "On linear algebraic techniques for liveness analysis of P/T systems," *Journal of Circuits, Systems, and Computers*, vol. 1, no. 8, pp. 223-265, 1998.
- [13] S.D. Young, V.K. Garg, "Optimal Sensor and Actuator Choices for Discrete Event Systems," *Proc. 31st Allerton Conference on Communication, Control, and Computing*, 1993.
- [14] P.J. Ramadge and W.M. Wonham, "Control of Discrete-Event Systems", *Proc. IEEE*, vol. 77, no. 1, pp. 81-98, Jan. 1989.