Design of Observers for Timed Continuous Petri Net Systems *

Jorge Júlvez ◇ †, Emilio Jiménez ‡, Laura Recalde†, Manuel Silva †
†Dep. de Informática e Ingeniería de Sistemas Centro Politécnico Superior, Universidad de Zaragoza
Maria de Luna 3, E-50015 Zaragoza, Spain, {julvez, lrecalde, silva}@unizar.es
‡Dep. Ingeniería Eléctrica, Universidad de La Rioja, Luis de Ulloa 20, E-26004 Logroño, Spain,
emilio.jimenezm@die.unirioja.es

***Abstract*— The design of observers is a topic of major importance in systems theory. Under a given firing semantics for transitions, the formalism of timed continuous Petri net systems represents a specific kind of piecewise linear system: the linear system that drives the evolution of the net system depends exclusively on the marking. This implies that a switch occurs only when an internal event happens. This paper deals with observability on continuous Petri nets and focuses on the design of observers. The proposed observers are piecewise linear systems that assure the continuity of the estimate even when a switch occurs. The use of the system simulation may allow to estimate even the unobservable space of the net system during a given time period.**

1 Introduction

The state of a dynamic system is defined by means of state variables. Some of them can be directly measured (sensed), while, under some conditions, some of the others can be estimated. This estimate constitutes the observation. The observability problem, i.e., the characterization of which state variables are observable and its observation, has been studied in detail in the framework of linear systems [6]. For these systems, the observable space can be characterized algebraically. A system state estimate based on such algebraic characterization can be *theoretically* obtained from the computation of the derivatives of the output signal. The estimate loses its reliability when "high" frequency noise appears in the output signal. In order to overcome this problem, linear observers came up [4]. A linear observer is a linear system whose state converges asymptotically to the state of the system being observed.

Petri nets represent a powerful formalism for the modelling of discrete concurrent systems. The continuous *relaxation* of Petri nets has been introduced in order to tackle the state explosion problem inherent to large discrete systems. Under an infinite servers semantics, a timed continuous Petri net system can be seen as a piecewise linear system [9], [10], i.e., the evolution of the state of the system is ruled by a set of switching differential equation systems. The timed continuous Petri net has the particularity that, at a given instant, the differential equation system that rules its evolution depends uniquely on the state of the system (marking). Hence, the switch from one linear differential equation system to another one is activated by an *internal event*, i.e., by a certain change of the marking.

The main goal of the paper is the design of observers for continuous Petri nets (see [3] for a work on observability of discrete Petri nets). One Luenberger's observer will be considered per each differential equation system that may rule the evolution of the net system (in a similar way to [1]). Each observer will yield an estimate that will be classified as *suitable* or *non-suitable* with respect to the current system's output. We propose an algorithm that filters non-suitable estimates and simulates the net system from a given instant in order to compute an estimate for the system's marking.

The paper is structured as follows: In Section 2 continuous Petri nets are introduced. In Section 3 the observability problem for continuous Petri nets is stated in a similar way to the observability problem for linear systems. Section 4 establishes the guidelines to detect non-suitable estimates. Section 5 shows how a set of linear observers can be created for a net system and the different classes of non-suitable observers' estimates that can appear. Section 6 is devoted to the design of an observer that uses a filter for the estimates and the simulation of the system. Conclusions are drawn in Section 7.

2 Continuous Petri Net Systems

2.1 Untimed Continuous Petri Net Systems

The reader is assumed to be familiar with Petri nets (PNs) (see for example [5], [8], [2]). The Petri net systems that will be considered are *continuous*. Continuous systems are obtained as a relaxation of *discrete* ones. Unlike 'usual' discrete systems, the amount in which a transition can be fired in a continuous Petri net system is not restricted to be a natural number. Firing a transition a non-negative real amount of times may cause the marking of the system to become a vector of real numbers. A PN system is a pair $\langle \mathcal{N}, \mathbf{m_0} \rangle$, where $\mathcal{N}$ specifies the net structure, $\mathcal{N} = \langle P, T, \mathbf{Pre}, \mathbf{Post} \rangle$ and $\mathbf{m_0}$ is the initial marking. The sets of places and transitions are denoted by P and T respectively and $\mathbf{C} = \mathbf{Post} - \mathbf{Pre}$ is the token flow matrix.

In continuous Petri net systems a transition t is *enabled* at a marking $\mathbf{m}$ iff every input place of t is marked (every $p \in {}^\bullet t$, $\mathbf{m}[p] > 0$). As in discrete systems, the *enabling degree* at marking $\mathbf{m}$ of a transition measures the maximum amount in which the transition can be fired in a single occurrence, i.e., $\text{enab}(t, \mathbf{m}) = \min_{p \in {}^\bullet t}\{\mathbf{m}[p]/\mathbf{Pre}[p, t]\}$. The firing of t in an amount $\alpha \leq \text{enab}(t, \mathbf{m})$ produces a new marking $\mathbf{m}'$, and it

◇ Supported by a grant from D.G.A. ref B106/2001
† Partially supported by project CICYT and FEDER TIC2001-1819 and DPI2003-06376

is denoted as $\mathbf{m}\xrightarrow{\alpha t}\mathbf{m}'$. It holds $\mathbf{m}' = \mathbf{m} + \alpha \cdot \mathbf{C}[P,t]$, hence, as in discrete systems the state equation $\mathbf{m} = \mathbf{m_0} + \mathbf{C} \cdot \boldsymbol{\sigma}$ summarizes the way the marking evolves, where $\boldsymbol{\sigma}$ is the firing count vector.

2.2 Timed Continuous Petri Net Systems

For the timing interpretation, a first order (or deterministic) approximation of the discrete case will be used [7], assuming that the delays associated to the firing of the transitions can be approximated by their mean values. Each transition t has associated an internal firing speed $\boldsymbol{\lambda}[t] > 0$. The state equation has an explicit dependence on time $\mathbf{m}(\tau) = \mathbf{m_0} + \mathbf{C} \cdot \boldsymbol{\sigma}(\tau)$. Deriving with respect to time, $\dot{\mathbf{m}}(\tau) = \mathbf{C} \cdot \dot{\boldsymbol{\sigma}}(\tau)$ is obtained. Let us denote $\mathbf{f} = \dot{\boldsymbol{\sigma}}$, since it represents the *flow* through the transitions. In this paper it will be assumed that every transition has at least one input place. Infinite servers semantics will be considered. Under this semantics, the flow of a transition is given by the product of $\boldsymbol{\lambda}$ and its enabling degree, i.e., $\mathbf{f}[t] = \boldsymbol{\lambda}[t] \cdot \mathrm{enab}(t, \mathbf{m}) = \boldsymbol{\lambda}[t] \cdot \min_{p \in {}^\bullet t}\{\mathbf{m}[p]/\mathbf{Pre}[p,t]\}$, what lead us to a non-linear system. More precisely, a piecewise linear system is obtained. The evolution of the system at a given instant is expressed by $\dot{\mathbf{m}} = \mathbf{A}_i \cdot \mathbf{m}$, where $\mathbf{A}_i$ is a constant matrix. To compute $\mathbf{A}_i$ it is necessary to know the set of places that is actually enabling the transitions, i.e., the set of places that are giving the minimum in the expression for the enabling degree. Once this set is computed, it is easy to establish a linear relationship between the marking of the places in this set and the flow of the transitions ($\mathbf{f}[t] = \boldsymbol{\lambda}[t] \cdot \mathbf{m}[p]/\mathbf{Pre}[p,t]$ if $p \in {}^\bullet t$ and p is giving the minimum). From the flow of transitions the derivative of the marking is obtained ($\dot{\mathbf{m}}(\tau) = \mathbf{C} \cdot \mathbf{f}$).

For each marking $\mathbf{m}$, its PT-set (PT-set($\mathbf{m}$)) can be defined as the set of all the pairs, (p,t), such that the marking of p is restricting the flow of transition t at marking $\mathbf{m}$.

Definition 1: Given a net system, the *PT-set* at marking $\mathbf{m}$ is

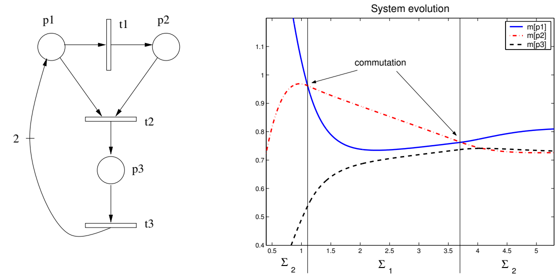
$$\text{PT-set}(\mathbf{m}) = \{(p,t) \mid \mathbf{f}[t] = \boldsymbol{\lambda}[t] \cdot \mathbf{m}[p]/\mathbf{Pre}[p,t]\} \qquad (1)$$

In this way, for every marking $\mathbf{m}$, there exists a PT-set k that has associated a square matrix $\mathbf{A}_k$ and a linear system $\Sigma_k : \dot{\mathbf{m}} = \mathbf{A}_k \cdot \mathbf{m}$ that rules the evolution of the system. An interesting issue is that the switch between the linear systems is activated by internal events, i.e., the change from one PT-set to another occurs when the place giving the minimum enabling degree of a transition changes.

Let us consider the system in Figure 1(a) with initial marking $\mathbf{m_0} = (3\ 0\ 0)$ and transition speeds $\boldsymbol{\lambda} = (0.9\ 1\ 1)$. If $\mathbf{m}[p_1] \le \mathbf{m}[p_2]$ (Σ_1), the flow of transition t_2 will be defined by the marking of p_1 and the PT-set will be $\{(p_1,t_1),(p_1,t_2),(p_3,t_3)\}$. Similarly, if $\mathbf{m}[p_1] \ge \mathbf{m}[p_2]$ (Σ_2) the flow of t_2 will be restricted by p_2 and the PT-set will be $\{(p_1,t_1),(p_2,t_2),(p_3,t_3)\}$. The matrices $\mathbf{A}_i$ are:

$$\mathbf{A}_1 = \begin{pmatrix} -1.9 & 0 & 2 \\ -0.1 & 0 & 0 \\ 1.0 & 0 & -1 \end{pmatrix} \mathbf{A}_2 = \begin{pmatrix} -0.9 & -1 & 2 \\ 0.9 & -1 & 0 \\ 0.0 & 1 & -1 \end{pmatrix}$$

At the time instant in which $\mathbf{m}[p_1] = \mathbf{m}[p_2]$, Σ_1 and Σ_2 behave in the same way and any of them can be taken. Figure 1(b) shows the evolution of the system along time. At the beginning the system evolves according to Σ_2. Then a switch occurs and the dynamics of the system is described by Σ_1. A second switch turns the system back to Σ_2, the system stabilizes and no more switches take place.



(a) (b)

Fig. 1. Marking evolution of a system with two PT-sets.

3 Observability: Problem Statement

Observability has been thoroughly studied in the framework of linear time invariant systems [4], [6]. A linear time invariant system without inputs is usually expressed by equations $\dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}, \mathbf{y} = \mathbf{S} \cdot \mathbf{x}$ where $\mathbf{x}$ is the state of the system and $\mathbf{y}$ is the output. The state space is denoted as $\mathbf{X}$. Knowing the matrices $\mathbf{A}$ and $\mathbf{S}$ and being able to watch the evolution of $\mathbf{y}$, a linear system is said to be observable iff it is possible to compute its initial state, $\mathbf{x}(\tau_0)$.

An observability criterion exists that allows to decide whether a linear system is observable or not. Given a linear system of dimension n expressed in discrete time, $\mathbf{x}(k+1) = \mathbf{F} \cdot \mathbf{x}(k)$, $\mathbf{y}(k) = \mathbf{S} \cdot \mathbf{x}(k)$ the output of the system in the first $n-1$ periods is given by:

$$\begin{pmatrix} \mathbf{y}(1) \\ \mathbf{y}(1) \\ \mathbf{y}(2) \\ .. \\ \mathbf{y}(n-1) \end{pmatrix} = \begin{pmatrix} \mathbf{S} \\ \mathbf{S} \cdot \mathbf{F} \\ \mathbf{S} \cdot \mathbf{F}^2 \\ .. \\ \mathbf{S} \cdot \mathbf{F}^{n-1} \end{pmatrix} \cdot \mathbf{x}_0 = \vartheta \cdot \mathbf{x}_0 \qquad (2)$$

The matrix ϑ is called observability matrix. The linear system is observable iff ϑ has full rank. For a non observable system it is possible to decompose the state space $\mathbf{X}$ into two subspaces: the observable subspace, $\mathbf{X}_o$, and the non observable subspace, $\mathbf{X}_{no}$ ($\mathbf{X}_{no}$ is the kernel of ϑ, i.e., $\vartheta \cdot \mathbf{X}_{no} = \mathbf{0}$).

In a timed continuous Petri net system every linear system $\Sigma_i : \dot{\mathbf{m}} = \mathbf{A}_i \cdot \mathbf{m}$ associated to a PT-set can be discretized in time. The equivalent discrete time system can be written as $\Sigma_i^d : \mathbf{m}(k+1) = \mathbf{F}_i \cdot \mathbf{m}(k)$, with $\mathbf{F}_i = e^{\mathbf{A}_i \cdot \delta}$ where δ is the time period. The output of the net system is given by $\mathbf{y} = \mathbf{S} \cdot \mathbf{m}$. Let us introduce the concept of observability for a continuous Petri net system and for a given PT-set.

Definition 2: A continuous Petri net system will be said to be observable iff given the structure of the net, $\mathcal{N}$, the internal speeds of the transitions, $\boldsymbol{\lambda}$, and the evolution of the output, $\mathbf{y}$, it is possible to compute the initial marking of the net, $\mathbf{m}(\tau_0)$.

Definition 3: Given a Petri net system, a PT-set system, $\dot{\mathbf{m}} = \mathbf{A}_k \cdot \mathbf{m}$, will be said to be observable iff its associated observability matrix ϑ_k has full rank.

Notice that every $\mathbf{F}_i$ is an exponential matrix and therefore it can be inverted. Hence continuous Petri net systems can be simulated backwards. This implies that if the marking of the system at a given instant is known then the marking of the system at any previous instant can be computed.

4 Computation of suitable estimates

An estimate per each possible PT-set of the net can be obtained by means of Equation 2 defined for the first $n-1$ periods. Theoretically, time discretization (τ) can be done as small as desired. It will be assumed that, for a small enough τ, no switch between PT-sets takes place in the first $n-1$ periods. The computed estimates may be used to filter those PT-sets that for sure are not ruling the evolution of the system. If only one PT-set remains, then the system evolves according to it.

The set of non-suitable estimates can be divided into three subsets: ***infeasible*** estimates, i.e., no solution of Equation 2, ***non-coherent*** estimates, i.e., the PT-set of the estimate is not the one for which it was computed, and ***suspicious*** estimates, i.e., the estimate belongs simultaneously to several PT-sets. Let us show through an example how infeasible and suspicious estimates can be used to filter PT-sets.

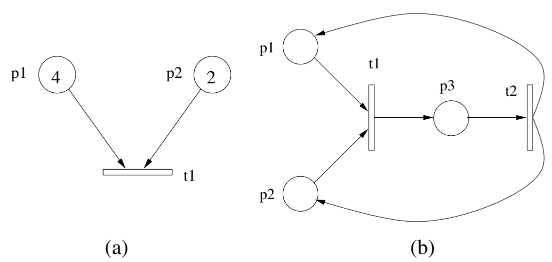


Fig. 2. (a) A synchronization with two input places. (b) A simple general Petri net system with two PT-sets.

Consider a system composed of a single synchronization with two input places p_1 and p_2, see Figure 2(a). The net has two possible PT-sets, either $W_1 = \{(p_1, t_1)\}$ or $W_2 = \{(p_2, t_1)\}$. If the time period is one time unit, the evolution of the system according to PT-set W_1 is ruled by the matrix $\mathbf{F}_1 = (e^{-1} \quad 0; e^{-1} - 1 \quad 1)$. The system matrix for PT-set W_2 is $\mathbf{F}_2 = (1 \quad e^{-1} - 1; 0 \quad e^{-1})$. Considering that the initial marking is $\mathbf{m_0} = (4; 2)$, the initial PT-set for the system is W_2, and after one time unit the marking will be $\mathbf{m}(\tau = 1) = (2 \cdot e^{-1} + 2; 2 \cdot e^{-1})$.

Assume that the output of the system is the marking of p_1 ($\mathbf{S} = (1\ 0)$) then after two steps we have $\mathbf{y} = (4; 2 \cdot e^{-1} + 2)$. Since the marking of p_2 is not in the output, an external agent of the system cannot know the PT-set in which the system is initially. Let us assume that the initial PT-set of the system is W_1. An estimate for the PT-set W_1 can be computed using Equation 2 with $\vartheta = (1\ 0; e^{-1}\ 0)$. It turns out that such a equation has no solution. This means that the estimate for W_1 is ***infeasible***, and therefore the PT-set W_1 is not the initial PT-set of the system.

Let us now consider the same system, Figure 2(a), with the marking of p_2 as the only output of the system. After two steps the output of the system is $\mathbf{y} = (2; 2 \cdot e^{-1})$. Again, we cannot know in advance what the real initial PT-set of the system is. Let us assume the initial PT-set is W_1. The observability matrix for PT-set W_1 is $\vartheta = (0\ 1; e^{-1} - 1\ 1)$. The estimate that Equation 2 yields for this PT-set is $\hat{\mathbf{m}}_0 = (2\ 2)$ that is different from real initial marking of the system. An initial marking $\mathbf{m_0} = (2\ 2)$ would mean that at the beginning the system is in both PT-sets, W_1 and W_2. The same estimate, $\hat{\mathbf{m}}_0 = (2\ 2)$, would have been obtained for any initial marking of p_1 greater than or equal to 2. Hence, this estimate is considered ***suspicious*** and should be filtered if a suitable estimate exists for other PT-set.

A basic feature that must be verified by an estimate is that it must be ***coherent*** with the PT-set for which it is computed. In other words, it does not make sense to consider an estimate that assigns a greater marking to p_1 than to p_2, if the PT-set for which it is computed happens when $\mathbf{m}[p_1] \leq \mathbf{m}[p_2]$. The net system in Figure 2(b) has two PT-sets: we will say that the system is in PT-set W_1 if $\mathbf{m}[p_1] \leq \mathbf{m}[p_2]$ and the system is in PT-set W_2 if $\mathbf{m}[p_1] \geq \mathbf{m}[p_2]$. In the case that $\mathbf{m}[p_1] = \mathbf{m}[p_2]$, the system is considered to be in both PT-sets simultaneously. Two estimates will be computed for this system, one per PT-set. The estimate corresponding to PT-set W_1, $\hat{\mathbf{m}}_0^{(1)}$, (W_2, $\hat{\mathbf{m}}_0^{(2)}$) has to be solution of Equation 2 with ϑ computed for the linear system associated to the PT-set and $\hat{\mathbf{m}}_0^{(1)}$ ($\hat{\mathbf{m}}_0^{(2)}$) has to fulfill $\hat{\mathbf{m}}_0^{(1)}[p_1] < \hat{\mathbf{m}}_0^{(1)}[p_2]$ ($\hat{\mathbf{m}}_0^{(2)}[p_1] > \hat{\mathbf{m}}_0^{(2)}[p_2]$). The use of strict inequalities allow us to filter also suspicious estimates like the one just shown in this Section.

For a general Petri net system with k PT-sets, a set of equation systems, $E_1 \ldots E_k$, can be defined. Each E_i contains Equation 2 with ϑ_i and the set of strict inequalities that defines the PT-set. The following proposition establishes when the initial PT-set can be uniquely determined before a switch happens.

Proposition 4: Assuming that a continuous Petri net is not initially in several PT-sets, the PT-set of $\mathbf{m_0}$ can be determined before a switch to another PT-set happens iff only one system $E_i,\ 1 \leq i \leq k$ has solution.

5 Observers and estimates

Previous section shows how an estimate can be computed by using Equation 2. The main drawback of that method is that it is very sensitive to the noise that may appear in the output $\mathbf{y}$. In order to overcome the problem of noise, observers are introduced. Basically, an observer is a dynamical system whose input is the output of the system to be observed. The state of an observer is the estimate for the system to be observed. It will be shown that a great parallelism exists between algebraically computed estimates and observers' estimates. A well designed observer should converge asymptotically to the real state of the observed system. For linear systems, Luenberger's observers [4], [6] are widely used. A Luenberger observer for a Petri net with a single PT-set can be expressed as: $\dot{\tilde{\mathbf{m}}} = \mathbf{A} \cdot \tilde{\mathbf{m}} + \mathbf{K} \cdot (\mathbf{y} - \mathbf{S} \cdot \tilde{\mathbf{m}})$ where $\tilde{\mathbf{m}}$ is the marking estimate, $\mathbf{A}$ and $\mathbf{S}$ (see Section 3) are the matrices defining the evolution of the system marking and its output in continuous time, $\mathbf{y}$ is the output of the system and $\mathbf{K}$ is a design matrix of parameters. The eigenvalues of the observer can be chosen arbitrarily, by means of $\mathbf{K}$, iff the system to be observed is observable. If the eigenvalues of the observer are appropriately chosen then the estimate will converge asymptotically to the marking of the system. In the case that the system is not observable, an observer to estimate the observable subspace, $\mathbf{X}_o$, can be designed.

The reliability of an estimate can be measured by means of a *residual* [1]. Let us define a norm $||\cdot||$ as $||\mathbf{x}|| = |\mathbf{x}_1| + \ldots + |\mathbf{x}_n|$. The residual at a given instant, $r(\tau)$, is the distance between the output of the system and the output that the observer's estimate, $\tilde{\mathbf{m}}(\tau)$, yields, i.e., $r = ||\mathbf{S}\cdot\tilde{\mathbf{m}}(\tau) - \mathbf{y}(\tau)||$.

5.1 Filtering estimates

One (Luenberger) linear observers will be designed per PT-set of the Petri net system. The designed observers will be launched simultaneously, and each one of them will yield an estimate. Some estimates may not be suitable for the PT-sets for which they are computed. Such estimates cannot represent the marking of the system and must be filtered. Three conditions will be presented that the estimates of the observers have to fulfill in order to be suitable: 1) the residual must tend to zero; 2) the estimates of places in synchronization have to be coherent with the PT-set for which they are computed; 3) the estimate must not be *suspicious*, i.e., it must not belong to several PT-sets at the same time.

Let us consider the system in Figure 2(b) with $\boldsymbol{\lambda} = (1\ 1)$ to show the behavior of the observers and their estimates under different conditions. The net has two PT-sets: $W_1 = \{(p_1, t_1), (p_3, t_2)\}$ and $W_2 = \{(p_2, t_1), (p_3, t_2)\}$. The system has a single T-semiflow, (1 1). Hence, in the steady state the flow of both transitions is the same. Since the net has two PT-sets, two linear observers can be designed.

1) Residuals: Let us consider the observer designed for PT-set W_1. Such observer assumes that $\mathbf{m}[p_1] \leq \mathbf{m}[p_2]$ and so the system matrix in continuous time is $\mathbf{A} = (-1\ 0\ 1; -1\ 0\ 1; 1\ 0\ -1)$. Let us assume that the output of the system is the marking of places p_1 and p_3, i.e., $\mathbf{S} = (1\ 0\ 0; 0\ 0\ 1)$. Under this conditions the observable subspace, $\mathbf{X}_o$, corresponds just to the marking of places p_1 and p_3. That is, p_2 is *timed-implicit*, i.e., its marking is not giving the minimum in the expression for the enabling degree, and cannot be observed (in this case it is also implicit and could have been removed). Therefore, the observer sees the evolution of a dynamical system ruled by matrix $\mathbf{A}' = (-1\ 1; 1\ -1)$ for places p_1 and p_3.

If the real PT-set of the system is W_1 the system will evolve to a steady state marking, $\mathbf{m}$, at which $\mathbf{m}[p_1] = \mathbf{m}[p_3] < \mathbf{m}[p_2]$. An observer with appropriate eigenvalues will asymptotically converge to an estimate marking $\tilde{\mathbf{m}}[p_1] = \tilde{\mathbf{m}}[p_3]$ and the residual will go to 0 as time increases. If the real PT-set of the system is p_2, p_3, then in the steady state $\mathbf{m}[p_2] = \mathbf{m}[p_3] < \mathbf{m}[p_1]$. It can be checked, that the observer will not reach a steady state estimate in which $\tilde{\mathbf{m}}[p_2] = \tilde{\mathbf{m}}[p_3]$ and therefore the residual will not tend to 0. In this case, the information given by the residual allows to decide that W_1 is not the PT-set of the net system. In general, every observer's estimate whose residual is not converging to 0 has to be filtered.

2) Coherent Estimates: It will be said that an estimate is coherent with the PT-set for which it was computed if it belongs to that PT-set. Let us consider again the observer designed for PT-set W_1 of the system in Figure 2(b) and let now the output matrix be $\mathbf{S} = (1\ 0\ 0; 0\ 1\ 0)$. In this case the observable subspace is complete, i.e., $\mathbf{X} = \mathbf{X}_o$ and the marking of every place can be estimated. For the observer, p_2 has no influence on the dynamics of the system since it is not in W_1. In the steady state it verifies $\tilde{\mathbf{m}}[p_2] = \mathbf{m}[p_2]$. The observer *thinks* (assuming that PT-set W_1 is the real PT-set) that in the steady state $\mathbf{m}[p_1]$ has to equal $\mathbf{m}[p_3]$ so that the flow of t_1 is equal to the flow of t_2. Therefore, the system estimate converges to $\tilde{\mathbf{m}}[p_1] = \tilde{\mathbf{m}}[p_3] = \mathbf{m}[p_1]$. In this way, the residual, $r = |\tilde{\mathbf{m}}[p_1] - \mathbf{m}[p_1]| + |\tilde{\mathbf{m}}[p_2] - \mathbf{m}[p_2]|$, is always equal to 0 in the steady state, independently of the real PT-set of the system.

The same phenomenon appears in the observer for PT-set W_2. The estimate converges to $\tilde{\mathbf{m}}[p_1] = \mathbf{m}[p_1]$ and $\tilde{\mathbf{m}}[p_2] = \tilde{\mathbf{m}}[p_3] = \mathbf{m}[p_2]$, independently of the real PT-set of the system. So, the residual always converges to 0.

Therefore, residuals are not helping to decide which PT-set of the system is the correct one. In principle, both observers are equally good since both residuals tend to 0. However, in order to choose the correct one it is enough to consider the marking of the places in the synchronization. Since, in this case both places are output of the system, it can be directly decided in which of the PT-sets the system is. In a general case, the estimate of an observer that is not coherent with its PT-set has to be filtered.

3) Suspicious Estimates: Let us consider the system in Figure 2(b) with output matrix $\mathbf{S} = (1\ 0\ 0; 0\ 0\ 1)$. The observable subspace of the observer for PT-set $W_2 = \{(p_2, t_1), (p_3, t_2)\}$ is complete. For this observer, the marking of place p_1 does not play any role in the evolution of the system. The estimate of place p_1 will always converge to the real marking of p_1, $\tilde{\mathbf{m}}[p_1] = \mathbf{m}[p_1]$. In the steady state, the observer will equal its estimates of p_2 and p_3 in order to fire transitions t_1 and t_2 in the same amount. Since $\mathbf{m}[p_3]$ is taken as output, the estimate will converge to $\tilde{\mathbf{m}}[p_2] = \tilde{\mathbf{m}}[p_3] = \mathbf{m}[p_3]$.

Let us assume that the real PT-set of the system is PT-set W_2. Then, in the steady state $\mathbf{m}[p_1] > \mathbf{m}[p_2] = \mathbf{m}[p_3]$ and according to the above reasonings the observer's estimate will correctly converge to the real marking of the system. If the real PT-set is PT-set W_1, the marking reached in the steady state fulfills $\mathbf{m}[p_2] > \mathbf{m}[p_1] = \mathbf{m}[p_3]$, and therefore the observer will converge to an estimate marking, $\tilde{\mathbf{m}}$, such that $\tilde{\mathbf{m}}[p_1] = \tilde{\mathbf{m}}[p_2] = \tilde{\mathbf{m}}[p_3]$. This estimate is considered *suspicious* because it assigns exactly the same markings to two places in synchronization for any initial marking $\mathbf{m_0}$ of the system such that $\mathbf{m_0}[p_2] \geq \mathbf{m_0}[p_1]$.

5.2 Observers' steady state

Some conditions to detect non-suitable observers' estimates have just been presented. A very tight relationship can be established between these conditions and those described in Section 4 for the estimates computed using Equation 2: for example, Equation 2 has no solution for a given PT-set iff the observer's estimate for that PT-set has a non null residual in the steady state. In the same way, suspicious or non-coherent estimates appear according to Equation 2 when there exists an observer whose estimate in the steady state is suspicious or non-coherent. Unlike the estimates computed in Section 4, the estimate yielded by an observer becomes more reliable as time increases.

When the system enters the steady state, its marking can be considered to remain constant and observers stabilize. At this point, those estimates that generate a non null residual or are not coherent must be filtered. Suspicious estimates must also be filtered if there exist at least another estimate that is not suspicious. If there is only one observer's estimate that has not been filtered, it is associated to the real PT-set of the system in the steady state. Assuming that in the steady state the system is not in more than one PT-set, the necessary and sufficient condition that has to be verified in order to filter every but one estimate, is equivalent to that of Proposition 4.

Proposition 5: Let us assume that the steady state marking of a system belongs only to one PT-set. In the steady state only one observer's estimate is not filtered, i.e., it generates a null residual, it is coherent with its PT-set and it is not suspicious iff only one equation system $E_i,\ 1 \leq i \leq n$ as defined in Section 4 associated to PT-set i has solution.

During the transient state, the estimates given by the observers may not be very reliable since there has not been enough time to get stabilized. At a given instant in the transient, the number of observer's estimates that are coherent with their PT-set may differ from the number of E_i systems that have solution. When this happens, a way to choose an estimate consists of choosing the one with minimum residual.

6 Design of a switching observer

This Section illustrates the concepts presented in the previous Sections and shows the design of an observer based on the filter of non-suitable estimates and the simulation of the system.

6.1 Filter based observer

Let us consider the continuous Petri net system in Figure 1(a). Let the output of the system be the marking of place p_1, that is, $\mathbf{S} = (1\ 0\ 0)$. The net has two PT-sets: let one of the PT-sets be $Z_1 = \{(p_1, t_1), (p_1, t_2), (p_3, t_3)\}$ and the other be $Z_2 = \{(p_1, t_1), (p_2, t_2), (p_3, t_3)\}$. The observable subspace of the PT-set Z_1 is the marking of places p_1 and p_3, while the observable subspace of PT-set Z_2 is the marking of all the places. Let the internal speeds of the transitions be $\boldsymbol{\lambda} = (0.9\ 1\ 1)$ and the initial marking be $\mathbf{m_0} = (3\ 0\ 0)$. The marking evolution of this system is depicted in Figure 1(b).

One observer per PT-set will be designed: observer 1 for PT-set Z_1 and observer 2 for PT-set Z_2. Let the initial state of observer 1 be $\mathbf{e}_{01} = (1\ 2)$ and its eigenvalues be $(-12+2\cdot\sqrt{3}\cdot i,\ -12-2\cdot\sqrt{3}\cdot i)$. Since observer 1 can only estimate p_1 and p_3, the first component of its state vector corresponds to the estimate for $\mathbf{m}[p_1]$, and its second component to the estimate for $\mathbf{m}[p_3]$. For observer 2, let the initial state be $\mathbf{e}_{02} = (1\ 0\ 2)$ and its eigenvalues be $(-15,\ -12+2\cdot\sqrt{3}\cdot i,\ -12-2\cdot\sqrt{3}\cdot i)$. The evolutions of the estimates of the observers are depicted in Figures 3(a) and 3(b).

The estimate of observer 1 gets quite close to the real marking of the system when it is in PT-set Z_1. At time $\tau = 3.7$ the system switches to PT-set Z_2 and the estimate for the marking of place p_3 moves away from the real marking.

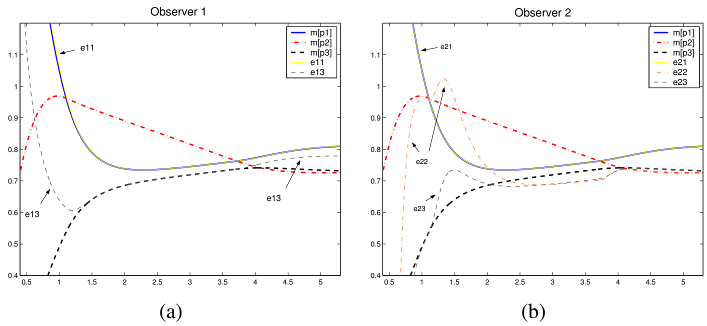


Fig. 3. (a) Evolution of observer 1, $(e11,\ e13)$ is the estimate for $(\mathbf{m}[p_1],\ \mathbf{m}[p_3])$. (b) Evolution of observer 2, $(e21,\ e22,\ e23)$ is the estimate for $(\mathbf{m}[p_1],\ \mathbf{m}[p_2],\ \mathbf{m}[p_3])$.

Similarly, the estimate of observer 2 gets very close to the marking of the system before it switches to PT-set Z_1 at time $\tau = 1.1$. As soon as the system switches, the observer loses the goodness of the estimate. When the system switches back to PT-set Z_2, the estimate approaches back quickly to the marking of the system.

After launching the observers for the PT-sets, a criterion must be adopted to decide which the best observer's estimate is. First, let us just filter the observer's estimate that has the greatest residual, see Figure 4(a). Before the first switch, observer 2 is chosen. After the switch, some time elapses till the residual of observer 2 becomes greater than the residual of observer 1. When this happens, the estimate of observer 2 is filtered. A similar phenomenon can be seen when the system switches from PT-set Z_1 to PT-set Z_2: after a little time period the estimate of observer 2 becomes smaller than the estimate of observer 1.

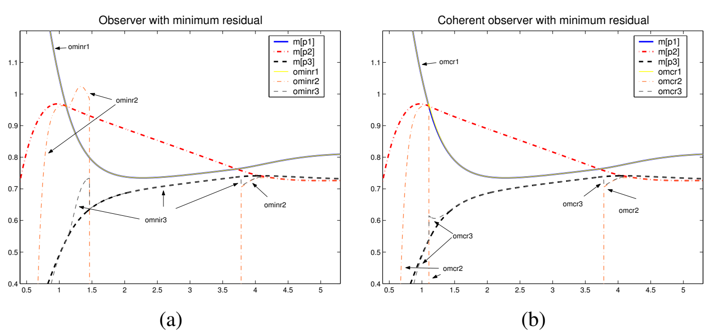


Fig. 4. (a) Minimum residual observer, $(ominr1,\ ominr2,\ ominr3)$ is the estimate for $(\mathbf{m}[p_1],\ \mathbf{m}[p_2],\ \mathbf{m}[p_3])$. (b) Minimum residual and coherent observer, $(omcr1,\ omcr2,\ omcr3)$ is the estimate for $(\mathbf{m}[p_1],\ \mathbf{m}[p_2],\ \mathbf{m}[p_3])$.

Notice that from the first system switch till the switch of observers, observer 2 has the minimum residual. However, it is not coherent with the PT-set for which it was designed, since $\tilde{\mathbf{m}}[p_1] < \tilde{\mathbf{m}}[p_2]$. Let us improve the estimate given by the observers by filtering those estimates that are not coherent with their PT-sets, see Figure 4(b). In this way, the first switch is immediately detected.

6.2 Improving the observer's estimate

The filter described in the previous Subsection allows to eliminate non-suitable estimates, i.e., infeasible, non-coherent and suspicious estimates. However, the resulting estimate can still be improved by taking into account some considerations. Let us have a look at Figure 4(b). When the first system switch happens, the estimate of observer 2 is very close to the marking of the system. By switching from observer 2 to observer 1, the estimate became discontinuous and, what it is more undesirable, the estimate for the marking of p_3 becomes worse. A similar effect happens when the second system switch occurs. Another undesirable phenomenon is that the estimate of the marking of p_2 just disappears (since $\mathbf{m}[p_2]$ in unobservable for observer 1) when the estimate of observer 2 is filtered.

One way to avoid discontinuities in the resulting estimate, is to use the estimate of the observer that is going to be filtered to update the estimate of the observer that is not going to be filtered. This estimate update must be done when a system switch is detected. In order not to lose the estimate of the marking of a place when it was almost perfectly estimated (recall the case of p_2 when the first switch happened) a simulation of the system can be launched. The initial marking of this simulation is the estimate of the system just before the observability of the place is lost (in the case of the example, the estimate of observer 2 when the first switch took place). Such simulation can be seen as an estimate for those places that are not observable by the observer being considered. The simulation can only be carried out when an estimate for all the places exists and the residual is quite small. Figure 5 shows the evolution of the estimate of the system taking into account the following: when the first switch is detected (observer 2 becomes non-coherent) the estimate of observer 1 is initialized with the estimate of observer 2. At that point a simulation is launched to estimate the marking of p_2. When the second switch is detected (the estimate of $\mathbf{m}[p_2]$ becomes smaller than the estimate of $\mathbf{m}[p_1]$) the estimate of observer 2 is initialized with the estimate of observer 1. Notice that once the estimate is close to the system marking, it does not move away from it, even if a switch happens. Based on these ideas, Algorithm 6 sketches how an estimate, est, can be computed as the system evolves.

Algorithm 6

Launch simultaneously one observer per PT-set
Repeat
  est_0 := suitable observer's estimate with minimum residual
  If $est_0 \neq \emptyset$ **then** % There exists a suitable estimate
    If est_0 does not estimate every place and there exists a simulation that is coherent with est_0 **then**
      est := est_0 plus the values of the simulation that are not in est_0
    Else
      est := est_0
    End_If
    If a system switch is detected **then**
      Update the estimates of the observers with est
      If est estimates every place with small residual **then**
        Create/substitute a simulation taking est as the initial marking
      End_If
    End_If
  Else % No estimate is suitable
    Take any observer's estimate
  End_If
End_Repeat

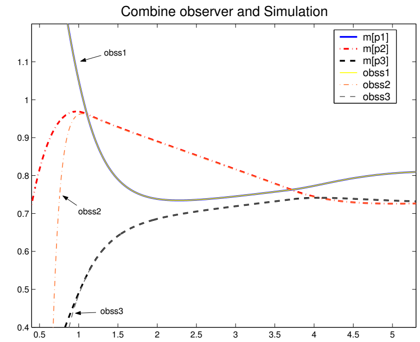


Fig. 5. Resulting observer's estimate that makes use of a simulation, ($obss1$, $obss2$, $obss3$) is the estimate for ($\mathbf{m}[p_1]$, $\mathbf{m}[p_2]$, $\mathbf{m}[p_3]$).

The resulting observer can be seen as a set of switching linear observers. One of the main advantages is that the residual does not increase sharply when the PT-set of the system changes. Another interesting feature is that the use of a simulation allows to estimate the marking of places that in some PT-sets are in principle not observable: in Figure 5 it can be seen that the marking of p_2 can be estimated, even when it is unobservable due to PT-set Z_1 being active.

7 Conclusions

In order to design an observer for a timed continuous Petri net one linear (Luenberger) observer per PT-set has been considered. As it happens when dealing with estimates computed algebraically, the estimate yielded by a given observer may not be suitable for the linear system for which it was designed: either the estimate generates a *non null residual* or it is *non-coherent* or *suspicious*. Based on the idea of choosing the suitable estimate with the smallest residual, a switching observer has been proposed. An interesting feature is that the observer launches a simulation of the system when the marking estimate is good enough. The use of such simulation allows to improve the estimate in two ways: the estimate does not change drastically when a system switch is detected; the unobservable space of the PT-set driving the evolution of the system may be estimated.

References

[1] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, and A. L. Sangiovanni-Vincentelli. Design of observers for hybrid systems. In Claire J. Tomlin and Mark R. Greenstreet, editors, *Hybrid Systems: Computation and Control*, volume 2289 of *Lecture Notes in Computer Science*, pages 76–89. Springer-Verlag, Berlin Heidelberg New York, 2002.

[2] R. David and H. Alla. *Petri Nets and Grafcet*. Prentice-Hall, 1992.

[3] A. Giua and A. Seatzu. Observability of place/transition nets. *IEEE Transactions on Automatic Control*, 47(9):1424–1437, September 2002.

[4] D.G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, 16(6):596–602, December 1971.

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

[6] K. Ogata. *Discrete-Time Control Systems, 2nd. ed.* Prentice Hall, 1995.

[7] L. Recalde and M. Silva. Petri Nets fluidification revisited: Semantics and steady state. *APII-JESA*, 35(4):435–449, 2001.

[8] M. Silva. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, pages 1–62. Chapman & Hall, 1993.

[9] M. Silva and L. Recalde. Petri nets and integrality relaxations: A view of continuous Petri nets. *IEEE Trans. on Systems, Man, and Cybernetics*, 32(4):314–327, 2002.

[10] M. Silva and L. Recalde. On fluidification of Petri net models: from discrete to hybrid and continuous models. In *IFAC Conference on Analysis and Design of Hybrid Systems, ADHS03*, pages 9–20, Saint-Malo, France, June 2003.