First-Order Hybrid Petri Nets: A Model for Optimization and Control

Fabio Balduzzi, *Member, IEEE*, Alessandro Giua, and Giuseppe Menga

***Abstract*—We consider in this paper first-order hybrid Petri Nets, a model that consists of continuous places holding fluid, discrete places containing a nonnegative integer number of tokens, and transitions, either discrete or continuous. We set up a linear algebraic formalism to study the first-order continuous behavior of this model and show how its control can be framed as a conflict resolution policy that aims at optimizing a given objective function. The use of linear algebra leads to sensitivity analysis that allows one to study of how changes in the structure of the model influence the optimal behavior. As an example of application, we show how the proposed formalism can be applied to flexible manufacturing systems with arbitrary layout and different classes of products.**

***Index Terms*—Flexible manufacturing systems, hybrid Petri nets, optimization, performance evaluation, sensitivity analysis.**

I. Introduction

A. Motivation

THE CONTROL of hybrid systems, i.e., systems with both continuous-time and discrete-event dynamics, is a domain of increasing importance and several hybrid models have been presented in the literature. Petri nets (PN's) [17] have originally been introduced to describe and analyze discrete event systems. Recently, much effort has been devoted to apply these models to hybrid systems (see Section I-C for a discussion of relevant literature).

The usual tradeoff between modeling power and analytical tractability also applies to hybrid PN models. On one hand, there exist Markovian models with a powerful set of analytical tools [22]. On the other, there exist models that can describe larger classes of systems but that can be validated only by simulation [10], or possibly if the incidence matrix can be defined for them, by invariant analysis [2], [3], [14], [16].

In this paper, we present a model called *first-order hybrid PN's* (FOHPN's) [6]–[8] that is general enough to model classes of systems of practical interest and yet whose first-order continuous behavior can be studied by linear algebraic tools. An FOHPN consists of continuous places holding fluid, discrete places containing a nonnegative integer number of tokens and transitions, either discrete or continuous. We assume that an autonomous timing structure (deterministic or stochastic) is associated to the discrete transitions. On the contrary, we assume that the *instantaneous firing speeds* (IFS) of the continuous transitions are control variables which may be chosen by the system operator within a given set $\mathcal{S} \subset (\mathbb{R}_0^+)^{n_c}$, where n_c is the number of continuous transitions and $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$.

In the paper, we address two main issues: *on-line control* and *structural optimization*. The *on-line control problem*, i.e., the computation of an optimal IFS vector $\boldsymbol{v} \in \mathcal{S}$, corresponds to a *conflict resolution*.

Conflict resolution is an important issue in the study of discrete nets. Conflicts arise when a limited number of tokens enables more than one transition, but it is only sufficient to fire a subset of them. Several schemes have been devised to tackle this problem for discrete timed nets, including *token reservations* [2], *resampling rules*, and *priorities* [1]. To any of these schemes corresponds a timing structure which imposes an *a priori* scheduling policy, i.e., the firing of discrete transitions is not controlled on-line. In FOHPN's, we adopt a similar approach and we assume that *discrete transitions* fire according to their timing structure. Note that this is not a limitation, as these transitions are generally used to represent the occurrence of events whose delay is not controllable (e.g., the failure or the repair of a machine).

The on-line control problem we address in this paper is that of computing an "optimal" mode of operation of the net by solving conflicts among *continuous transitions* at continuous places where continuous flows must be routed in the net. The choice of an optimal IFS vector is framed as a linear programming problem (LPP). This leads quite naturally to the *structural optimization problem*. In fact, the use of linear programming and sensitivity analysis tools allows one to study of how changes in the structure of the model influence the optimal behavior.

In the final section, we discuss as an example of application a manufacturing system. Manufacturing systems are discrete event systems whose number of reachable states is typically very large, hence approximating fluid models have often been used in this context [5], [11], [21]. The FOHPN model is rich enough to model manufacturing systems consisting of unreliable machines and buffers of finite capacity in the most general multiclass multimachine setting. Several criteria to assess performances of a manufacturing system, such as throughput rate, buffer levels, machines utilization, etc., can be framed as conflict resolution policies for the FOHPN model. In addition, a designer or analyst of a manufacturing system may want to perform "*what if*" analysis. For example, what happens to throughput if a machine produces faster? These problems

Manuscript received September 22, 1998; revised August 23, 1999. This paper was recommended for publication by Associate Editor M. Zhou and Editor P. Luh upon evaluation of the reviewers' comments. This paper was presented in part at the IEEE International Conference on Systems, Man and Cybernetics, San Francisco, CA, Ocotober 1998, and in part at the 4th Workshop on Discrete Event Dynamics Systems, Calgiari, Italy, August 1998.

F. Balduzzi and G. Menga are with the Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy (e-mail: balduzzi@polito.it; menga@polito.it).

A. Giua is with the Dipartimento di Ingegneria Elettrica ed Elettronica, Università di Cagliari, Cagliari, Italy (e-mail: giua@diee.unica.it).

Publisher Item Identifier S 1042-296X(00)06959-7.

will be addressed applying sensitivity analysis to the FOHPN model.

B. Proposed Model

In the first part of the paper, we describe a general hybrid PN model based on the framework proposed by Alla and David [2]. As in all hybrid models, we distinguish two behavioral levels.

At the lower level, the continuous evolution of the net is described by first-order fluid models, i.e., models in which the continuous flows have constant rates and the fluid content of each continuous place varies linearly with time. Each model is relative to a given *macrostate* characterized by the following: 1) the discrete marking of the net and 2) the set of empty continuous places.

The macro-state defines the set $\mathcal{S}$ of admissible IFS vectors of continuous transitions. Since we consider the first-order behavior, we assume that the IFS vector remains constant within a macro-state.

At the higher level, a discrete-event model describes the macro-behavior of the net that, upon the occurrence of *macro-events*, evolves through a sequence of macro-states. As it will be clear in the formal description of the model, we consider the following two types of macro-events: 1) the firing of a discrete transition, according to a given timing structure, and 2) the emptying of a continuous place.

A useful formalism for the analysis of discrete timed nets, assumes that the timing structure associated to the transitions is stochastic with exponentially distributed delays (see, e.g., [1]). This has the advantages of leading to a Markovian model that can be easily analyzed. In our model, however, the occurrence of the second type of macro-events does not preserve the Markovian property, hence, we allow arbitrary stochastic or deterministic timing structures. Thus, the resulting macro-behavior can be described as a generalized semi-Markovian process (GSMP).

The key features that represent the novelty of our approach are the following.

- We allow the IFS of a continuous transition t_i to be chosen by a control agent in a given range $[V_i', V_i]$, where V_i' is the *minimum firing speed* (mfs) and V_i is the *maximum firing speed* (MFS).
- We explicitly define the set of all admissible IFS vectors $\boldsymbol{v}$ within a given macro-state. This set is characterized by the feasible solutions of a linear constraint set.
- We consider the problem of choosing an optimal IFS vector according to a given objective function. Our optimization scheme can only be *myopic* [5], in the sense that it generates a piecewise optimal solution, i.e., a solution that is optimal only in a macro-state. Although we provide the basic tools for the analysis of the overall behavior of a net, this paper is only concerned with the analysis and optimization within a macro-state.
- We adopt the optimal basis approach, i.e., the simplex method, to solve LPP. We show how one can naturally apply to the FOHPN framework those sensitivity analysis techniques that pertain to LPP [15], [19] to study how changes in the structure of the model influence its optimal behavior.

C. Relevant Literature

The hybrid PN model we propose follows the formalism described by David and Alla [2], [3]. In effect, our model can be seen as an extension of timed hybrid nets with constant maximal speeds with the following additional features: token reservation is not used as a conflict resolution policy; stochastic transitions are also included in our model; minimal firing speeds are considered as well. The novel contribution of our work is that of showing how the first-order behavior of such a net can be efficiently analyzed with a linear algebraic formalism.

Linear algebraic techniques have also been used by Amrah *et al.* [4] when modeling manufacturing systems with continuous PN's. These authors deal with open and closed transfer lines modeled by *controlled variable speed continuous PN's*, a type of continuous PN [3] with controllable maximal firing speeds. Then by using a constrained optimization approach, they obtained optimal values for the machine production rates that bring the average levels of buffers to a desired values. However, it must be observed that transfer lines are not a general model, in the sense they do not require scheduling and routing. Our work, instead may deal with manufacturing systems in the more general configurations and settings.

Another approach that extends the stochastic discrete PN framework of [1] toward fluid approximations is *fluid stochastic PN's* presented by Trivedi and Kulkarni [22]. These authors proposed a model with places holding continuous tokens and arcs representing fluid flows. The flow rates are uniquely specified by the complete marking of the net. This is a main difference with our approach, in which the IFS vector of continuous transitions can be chosen in a given set that defines all admissible ones.

All these hybrid models discussed so far are fluid models, i.e., the marking of the continuous places is a nonnegative real number. However, more general hybrid PN's that also admit negative-real markings have been proposed. *Differential PN's* have been presented by Demongodin and Koussoulas [14] and can be used to model hybrid systems whose continuous evolution can be described by a finite number of linear first-order differential state equations. *High-level hybrid PN's* [16] extend the hybrid framework to colored nets: the use of real numbers as colors allows one to model general primitives such as jumps in the state space and switches in the continuous dynamics, which cannot be generally described with other formalisms.

In other approaches [10], the PN formalism is only used to represent the discrete state of a hybrid system, while the continuous state is represented by a vector (not by a marking) whose arbitrary continuous evolution is modeled by differential algebraic equations.

The rest of the paper is structured as follows. In Section II, we introduce FOHPN's. Section III is concerned with the computation of the instantaneous firing speed of continuous transitions and with different conflict resolution schemes. Section IV deals with the sensitivity analysis of this model. In Section V, we apply the previously developed results to a manufacturing system.

II. Background

We recall the PN formalism used in this paper. For a more comprehensive introduction to place/transition PN's, see [17]. The common notation and semantics for timed nets can be found in [1]. Hybrid PN's are defined in [2].

A. Structure and Marking

An FOHPN is a structure $N = (P, T, \mathrm{Pre}, \mathrm{Post}, \mathcal{D}, \mathcal{C})$.

The set of *places* $P = P_d \cup P_c$ is partitioned into a set of *discrete* places P_d (represented as circles) and a set of *continuous* places P_c (represented as double circles).

The set of *transitions* $T = T_d \cup T_c$ is partitioned into a set of discrete transitions T_d and a set of continuous transitions T_c (represented as double boxes). The set $T_d = T_I \cup T_D \cup T_E$ is further partitioned into a set of *immediate* transitions T_I (represented as bars), a set of *deterministic timed* transitions T_D (represented as black boxes), and a set of *exponentially distributed timed* transitions T_E (represented as white boxes). The cardinality of T, T_d, and T_c is denoted n, n_d, and n_c.

The *pre-* and *post-incidence functions* that specify the arcs are[1]

$$\mathrm{Pre}\colon \begin{cases} P_d \times T \to \mathbb{N} \\ P_c \times T \to \mathbb{R}_0^+ \end{cases}$$

and

$$\mathrm{Post}\colon \begin{cases} P_d \times T \to \mathbb{N} \\ P_c \times T \to \mathbb{R}_0^+. \end{cases}$$

We require (*well-formed nets*) that for all $t \in T_c$ and for all $p \in P_d$, $\mathrm{Pre}(p, t) = \mathrm{Post}(p, t)$.

The function $\mathcal{D}\colon T_d \backslash T_I \to \mathbb{R}^+$ specifies the timing associated to timed discrete transitions. We associate to a deterministic timed transition $t_i \in T_D$ its (constant) firing delay $\delta_i = \mathcal{D}(t_i)$. We associate to an exponentially distributed timed transition $t_i \in T_E$ its average firing rate $\lambda_i = \mathcal{D}(t_i)$, i.e., the average firing delay is $1/\lambda_i$, where λ_i is the parameter of the corresponding exponential distribution.

The function $\mathcal{C}\colon T_c \to \mathbb{R}_0^+ \times \mathbb{R}_\infty^+$ specifies the firing speeds associated to continuous transitions.[2] For any continuous transition $t_i \in T_c$, we let $\mathcal{C}(t_i) = (V_i', V_i)$, with $V_i' \le V_i$. Here V_i' represents the mfs and V_i represents the MFS. In the following, unless explicitly specified, the mfs of a continuous transition will be $V_i' = 0$.

We denote the preset (postset) of transition t as ${}^\bullet t$ ($t^\bullet$) and its restriction to continuous or discrete places as ${}^{(d)}t = {}^\bullet t \cap P_d$ or ${}^{(c)}t = {}^\bullet t \cap P_c$. Similar notation may be used for presets and postsets of places. The *incidence matrix* of the net is defined as $\boldsymbol{C}(p, t) = \mathrm{Post}(p, t) - \mathrm{Pre}(p, t)$. The restriction of $\boldsymbol{C}$ to P_X and T_Y ($X, Y \in \{c, d\}$) is denoted $\boldsymbol{C}_{XY}$. Note that by the well-formed hypothesis, $\boldsymbol{C}_{dc} = 0$.

A *marking*

$$\boldsymbol{m}\colon \begin{cases} P_d \to \mathbb{N} \\ P_c \to \mathbb{R}_0^+ \end{cases} \tag{1}$$

is a function that assigns to each discrete place a nonnegative number of tokens, represented by black dots and assigns to each continuous place a fluid volume; m_p denotes the marking of place p. The value of a marking at time τ is denoted $\boldsymbol{m}(\tau)$. The restriction of $\boldsymbol{m}$ to P_d and P_c are denoted with $\boldsymbol{m}^d$ and $\boldsymbol{m}^c$, respectively.

An *FOHPN system* $\langle N, \boldsymbol{m}(\tau_0)\rangle$ is an FOHPN N with an initial marking $\boldsymbol{m}(\tau_0)$.

[1]Here $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$.

[2]Here $\mathbb{R}_\infty^+ = \mathbb{R}^+ \cup \{\infty\}$.

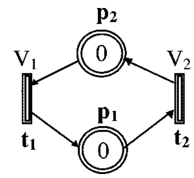

Fig. 1. FOHPN.

B. Enabling and Firing

The enabling of a discrete transition depends on the marking of all its input places, both discrete and continuous.

Definition 1: Let $\langle N, \boldsymbol{m}\rangle$ be an FOHPN system. A discrete transition t is *enabled* at $\boldsymbol{m}$ if for all $p \in {}^\bullet t$, $m_p \ge \mathrm{Pre}(p, t)$. ■

An enabled discrete transition t fires (after its associated delay) yielding the marking $\boldsymbol{m}' = \boldsymbol{m} + \boldsymbol{C}(\cdot, t)$. The firing of discrete transitions may follow any of the common enabling and firing rules discussed in [1]. These rules—that define the structure of the GSMP associated to the net—are well known and are not further discussed in this paper.

A continuous transition is enabled only by the marking of its input discrete places. The marking of its input continuous places, however, is used to distinguish between strongly and weakly enabling.

Definition 2: Let $\langle N, \boldsymbol{m}\rangle$ be an FOHPN system. A continuous transition t is *enabled* at $\boldsymbol{m}$ if for all $p \in {}^{(d)}t$, $m_p \ge \mathrm{Pre}(p, t)$.

We say that an enabled transition $t \in T_c$ is:

- *strongly enabled* at $\boldsymbol{m}$ if for all places $p \in {}^{(c)}t$, $m_p > 0$;
- *weakly enabled* at $\boldsymbol{m}$ if for some $\overline{p} \in {}^{(c)}t$, $m_{\overline{p}} = 0$. ■

Remark 3: We note that the definition of enabling we have given is slightly different from the one proposed by David and Alla in [3], where it was also required that a weakly enabled transition be "fed," i.e., that there exists an upstream transition strongly enabled feeding it. The two notions lead to different semantics for a cycle such as the one in Fig. 1. According to the definition of [3], the two transitions are not enabled and the cycle is blocked, while according to our definition they are both weakly enabled and the cycle is not blocked.

To overcome this limitation, David and Alla introduced in [13] a new concept—that of ϵ-marking. If an arbitrary small marking is initially assigned to any of the two places of the cycle in Fig. 1, then both transitions can be considered weakly enabled. Thus, in this generalized framework, it is possible to assign to empty cycles two semantics: blocked cycles (those that are empty) and nonblocked cycles (those ϵ-marked).

We believe that blocked cycles are not a useful modeling feature for systems of practical interest, thus we have chosen to keep just the second semantics. □

The enabling state of a continuous transition t_i defines its admissible *instantaneous firing speed* (IFS) v_i.

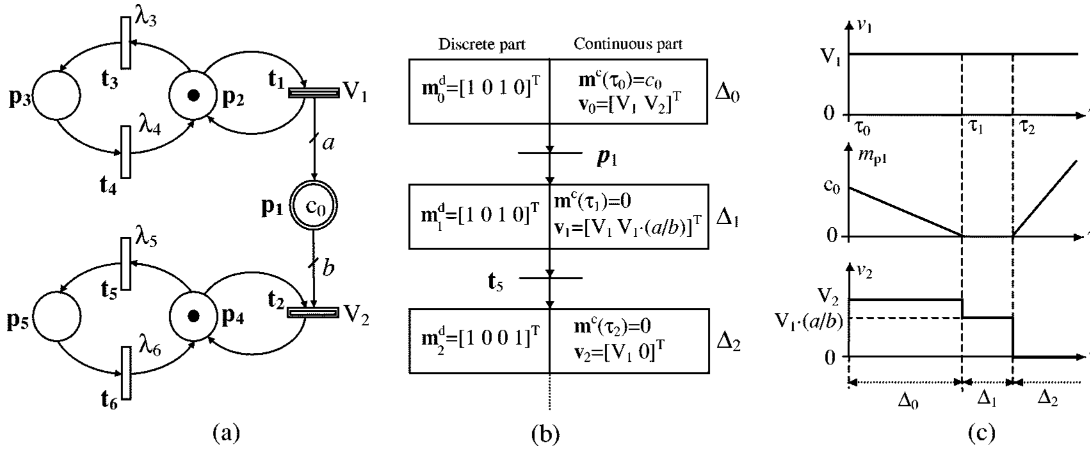

Fig. 2. (a) An FOHPN, (b) its phase-diagram, and (c) its evolution in time.

Definition 4: Let $\langle N, \boldsymbol{m}\rangle$ be an FOHPN system and $t_i \in T_c$ be a continuous transition with IFS v_i.

- If t_i is not enabled, then $v_i = 0$.
- If t_i is strongly enabled, then it may fire with any firing speed $v_i \in [V_i', V_i]$.
- If t_i is weakly enabled, then it may fire with any firing speed $v_i \in [V_i', \overline{V}_i]$, where $\overline{V}_i \le V_i$ depends on the amount of fluid entering the empty input continuous place(s) of t_i. In fact, the transition cannot remove more fluid from any empty input continuous place $\overline{p}$ than the quantity entered in $\overline{p}$ by other transitions. ■

The computation of the IFS of enabled transitions is not a trivial task. We will set up in Section III a linear-algebraic formalism to do this. Here, we simply discuss the net evolution assuming that the IFS are given.

The IFS at time τ of a transition $t_i \in T_c$ is denoted $v_i(\tau)$. The evolution in time of the marking of a place $p \in P_c$ is described by

$$\frac{dm_p(\tau)}{d\tau} = \sum_{t_i \in T_c} \boldsymbol{C}(p, t_i) \cdot v_i(\tau). \tag{2}$$

Indeed. (2) holds assuming that at time τ no discrete transition is fired and that all speeds $v_i(\tau)$ are continuous in τ.

C. Net Dynamics

A *macro-event* occurs when: 1) either a discrete transition fires, thus changing the discrete marking or enabling/disabling a continuous transition or 2) or a continuous place becomes empty, thus changing the enabling state of a continuous transition from strong to weak.

Let τ_k and τ_{k+1} be the occurrence times of consecutive macro-events; the interval of time $[\tau_k, \tau_{k+1})$ is called *macro-period* and its length is denoted $\Delta_k = \tau_{k+1} - \tau_k$.

We will assume that the IFS of continuous transitions are constant during a macro-period. Thus, the discrete marking and the IFS vector during a macro-period define a *macro-state* that corresponds to the *invariant behavior states* of [2].

We now describe the dynamics of an FOHPN. Let τ_0 be the initial time, $\tau_k (k > 0)$ be the instants in which macro-events occur, and $\boldsymbol{v}(\tau_k)$ be the IFS vector during the macro-period of length Δ_k. Let $\boldsymbol{\sigma}(\tau_k)$ be the *firing count vector* at time τ_k, i.e., a vector of dimension n_d that specifies the discrete transitions, if any, firing at time τ_k. Thus, the micro-behavior of an FOHPN is described during the kth macro-period by

$$\begin{cases} \boldsymbol{m}^c(\tau) &= \boldsymbol{m}^c(\tau_k) + \boldsymbol{C}_{cc} \cdot \boldsymbol{v}(\tau_k) \cdot (\tau - \tau_k) \\ \boldsymbol{m}^d(\tau) &= \boldsymbol{m}^d(\tau_k) \end{cases} \tag{3}$$

where $\tau \in [\tau_k, \tau_{k+1})$, while the evolution of the net at the occurrence of the macro-events is described by

$$\begin{cases} \boldsymbol{m}^c(\tau_k) &= \boldsymbol{m}^c(\tau_k^-) + \boldsymbol{C}_{cd} \cdot \boldsymbol{\sigma}(\tau_k) \\ \boldsymbol{m}^d(\tau_k) &= \boldsymbol{m}^d(\tau_k^-) + \boldsymbol{C}_{dd} \cdot \boldsymbol{\sigma}(\tau_k). \end{cases} \tag{4}$$

The macro-behavior of an FOHPN can be described by a *phase-diagram* in which every macro-state is represented by a box labeled on the right by the length Δ_k of the corresponding macro-period. Each box is partitioned into two parts. On the left (discrete part) the discrete marking $\boldsymbol{m}^d$ of the net is represented. On the right (continuous part) the continuous marking $\boldsymbol{m}^c$ of the net at time τ_k is represented with the IFS vector $\boldsymbol{v}$. Macro-states are connected through bars, representing the macro-events that caused the state transitions. Each bar is labeled on the left (respectively, right) by the discrete transition (respectively, continuous place) that caused the occurrence of the macro-event. An example is discussed in the following section.

D. Example

Consider the net in Fig. 2(a) and let the initial time be τ_0. Place p_1 is a continuous place with initial marking $m_{p_1}(\tau_0) = c_0 > 0$. Places p_2, p_3, p_4, p_5 are discrete places. Transitions t_1 and t_2 are continuous transitions with MFS V_1 and V_2. We assume $V_1\, a < V_2\, b$ (here a and b are the arc weights given by Pre and Post). Discrete transitions t_3, t_4, t_5, t_6 are exponentially distributed timed transitions whose average firing rates are λ_3, λ_4, λ_5, and λ_6, respectively.

Macro-Period MP0: In the initial state, p_1 is not empty and p_2, p_4 are marked. Thus, transitions t_1 and t_2 are strongly enabled and may fire at their maximum speeds, i.e., we choose $v_1 = V_1$ and $v_2 = V_2$. The continuous marking of the net during this macro-period is given, as in (2), by

$$\boldsymbol{m}^c(\tau) = m_{p_1}(\tau) = c_0 - (V_2\, b - V_1\, a)(\tau - \tau_0)$$

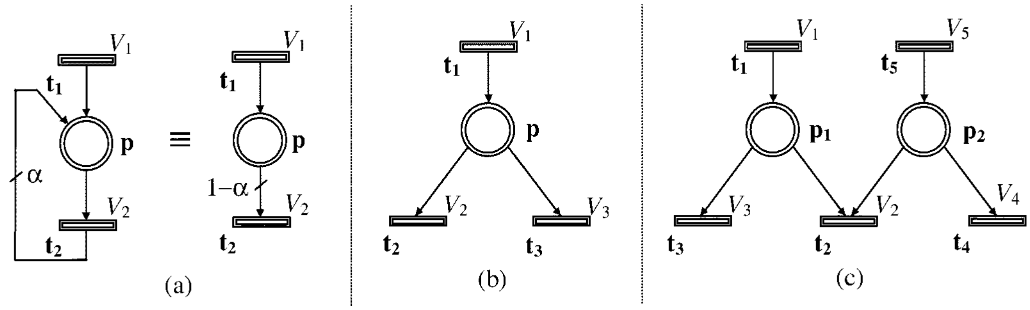

Fig. 3. FOHPN model of (a) a conflict-free re-entrant production line, (b) a free-choice conflict, and (c) a non free-choice conflict.

while the discrete marking is constant and given by

$$\boldsymbol{m}_0^d = [m_{p_2}\, m_{p_3}\, m_{p_4}\, m_{p_5}]^T = [1\ \ 0\ \ 1\ \ 0]^T.$$

Macro-Period MP1: At time $\tau_1 = (c_0/(V_2\, b - V_1\, a)) + \tau_0$, place p_1 becomes empty, thus causing a macro-event. In the new macro-state, t_1 remains strongly enabled, while t_2 is weakly enabled and may fire at most at speed $\overline{V}_2 = v_1(a/b) < V_2$. Assume $v_1 = V_1$ and $v_2 = V_1(a/b)$. Then, the continuous marking of the net during this macro-period is constant: $m_{p_1}(\tau) = 0$. The discrete marking maintains the value it had in the previous macro-period since no discrete transition has fired.

Macro-Period MP2: We assume that the enabled transition t_5 fires at time $\tau_2 > \tau_1$. This macro-event changes the discrete marking of the net to $\boldsymbol{m}_2^d = [1\quad 0\quad 0\quad 1]^T$. Now transition t_2 is disabled, i.e., $v_2 = 0$, while t_1 remains strongly enabled. Assume $v_1 = V_1$. Then, the continuous marking during this macro-period is given, as in (2), by

$$\boldsymbol{m}^c(\tau) = m_{p_1}(\tau) = V_1 a(\tau - \tau_2).$$

This behavior is represented in Fig. 2(b) and (c), which shows the phase-diagram of this net and the evolution in time of m_{p_1}, v_1, v_2.

III. Firing Speed and Dynamics of an FOHPN

The computation of an admissible IFS vector of continuous and hybrid nets is not trivial. In [3], an iterative algorithm was given to determine a *single* admissible vector; the algorithm aims at maximizing firing speeds while respecting priority rules. We propose a different approach in which we use linear inequalities to characterize the set $\mathcal{S} \subset (\mathbb{R}_0^+)^{n_c}$ of *all* admissible firing speed vectors. Each vector $\boldsymbol{v} \in \mathcal{S}$ represents a particular mode of operation of the system described by the net, and among all possible modes of operation, the system operator may choose the best according to a given objective. There are several advantages in our approach.

- We can explicitly characterize the set of all admissible IFS vectors in a given macro-state and not just compute a particular vector.
- We consider more general scheduling rules than priorities. For example, in an FMS, we may want to maximize machines utilization, maximize the throughput of the system, balance the load, etc. Each of these problems corresponds to a particular objective function. Note that each set $\mathcal{S}$ corresponds to a particular system macro-state. Thus, our optimization scheme can only be *myopic* [5], in the sense that it generates a piecewise optimal solution, i.e., a solution that is optimal only in a macro-period.
- We compute a particular (optimal) IFS vector solving a linear programming problem, rather than by means of an iterative algorithm, whose convergence properties may not be good.
- Linear programming leads to sensitivity analysis, which plays an essential role in performance evaluation and optimization. In fact, we may be able to compute analytically the objective function improvement due to a parameter variation.

A. Admissible IFS Vectors

In this section, we characterize the set of admissible IFS vectors.

Definition 5: Let $\langle N, \boldsymbol{m}\rangle$ be an FOHPN system with n_c continuous transitions and incidence matrix $\boldsymbol{C}$. Let $T_{\mathcal{E}}(\boldsymbol{m}) \subset T_c$ ($T_{\mathcal{N}}(\boldsymbol{m}) \subset T_c$) be the subset of continuous transitions enabled (not enabled) at $\boldsymbol{m}$, and $P_{\mathcal{E}} = \{p \in P_c | m_p = 0\}$ be the subset of empty continuous places. Any *admissible IFS vector* $\boldsymbol{v} = [v_1 \cdots v_{n_c}]^T$ at $\boldsymbol{m}$ is a feasible solution of the following linear set:

$$\begin{cases} \text{(a)} & V_j - v_j \geq 0 & \forall\, t_j \in T_{\mathcal{E}}(\boldsymbol{m}) \\ \text{(b)} & v_j - V_j' \geq 0 & \forall\, t_j \in T_{\mathcal{E}}(\boldsymbol{m}) \\ \text{(c)} & v_j = 0 & \forall\, t_j \in T_{\mathcal{N}}(\boldsymbol{m}) \\ \text{(d)} & \displaystyle\sum_{t_j \in T_{\mathcal{E}}} \boldsymbol{C}(p, t_j) \cdot v_j \geq 0 & \forall\, p \in P_{\mathcal{E}}(\boldsymbol{m}). \end{cases} \tag{5}$$

The set of all feasible solutions is denoted $\mathcal{S}(N, \boldsymbol{m})$. ■

Thus, the total number of constraints that define $\mathcal{S}(N, \boldsymbol{m})$ is: $2\,\text{card}\{T_{\mathcal{E}}(\boldsymbol{m})\} + \text{card}\{T_{\mathcal{N}}(\boldsymbol{m})\} + \text{card}\{P_{\mathcal{E}}(\boldsymbol{m})\}$ (here, $\text{card}\{A\}$ denotes the cardinality of the set A). Constraints of the form (5.a)–(5.c) follow from the firing rules of continuous transitions. Constraints of the form (5.d) follow from (2), because if a continuous place is empty then its fluid content cannot decrease.

Note that if $V_j' = 0$, then the constraint of the form (5.b) associated to t_j reduces to a nonnegativity constraint on v_j.

Example 6: Let $\langle N, \boldsymbol{m}\rangle$ be the net in Fig. 3(a), with $\alpha \in (0, 1)$, where place p is initially empty. Such a net is representative of a re-entrant production line. Hence, according to the previous definition

$$\begin{cases} V_1 - v_1 & \geq 0 \\ V_2 - v_2 & \geq 0 \\ v_1, v_2 & \geq 0 \\ v_1 - (1-\alpha)v_2 & \geq 0 \end{cases} \tag{6}$$

is the linear constraint set that defines $\mathcal{S}(N, \boldsymbol{m})$. ■

We now discuss under which conditions the set $\mathcal{S}(N, \boldsymbol{m})$ admits feasible solutions. When no feasible solution exists, no admissible modes of operation is allowed by the net. First, we make the following observation.

Remark 7: Any constraint of the form (5d) related to an empty continuous place p can be written as

$$\sum_{j \in J} \alpha_j v_j \geq \sum_{k \in K} \alpha_k v_k \tag{7}$$

with $J \cap K = \emptyset$ and $\alpha_j, \alpha_k \in \mathbb{R}^+$. The set J (respectively, K) contains the indices of the continuous transitions whose firing increases (respectively, decreases) the marking of p. □

Definition 8: Let $\langle N, \boldsymbol{m} \rangle$ be an FOHPN system. A transition $t_j \in T_c$ is called *mfs-free* at $\boldsymbol{m}$ if at least one of the following conditions holds:

1) the mfs of t_j is $V_j' = 0$;
2) t_j has no empty input continuous places, i.e., $\forall p \in {}^{(c)}t$: $m_p > 0$. ■

We can now provide a sufficient condition for the existence of admissible IFS vectors.

Proposition 9: Let $\mathcal{S}(N, \boldsymbol{m})$ be the linear set defined by (5). Such a set is nonempty if all $t_j \in T_{\mathcal{E}}(\boldsymbol{m})$ are mfs-free at $\boldsymbol{m}$.

Proof: Let $\boldsymbol{v}^{\min}$ be such that $v_j^{\min} = 0$ if $t_j \in T_{\mathcal{N}}(\boldsymbol{m})$, else $v_j^{\min} = V_j'$. We prove that $v^{\min} \in \mathcal{S}(N, \boldsymbol{m})$.

Clearly this vector satisfies all constraints of the form (5a)–(5c). Moreover, from Remark 7, it follows that any constraint of the form (5d) can be written as

$$\sum_{j \in J} \alpha_j v_j^{\min} \geq \sum_{k \in K} \alpha_k v_k^{\min}$$

and the right-hand side of this equation evaluates to 0 by the assumption that all enabled transitions are mfs-free. Thus, this constraint is satisfied. □

As a counterexample, we show that no feasible solution may exist if one (or more) continuous transition is not mfs-free.

Example 10: Consider the net in Fig. 3(a) with $\alpha = 0$. Let $m_p = 0$ and $V_2' > V_1 \geq V_1' = 0$. Thus, transition t_2 is not mfs-free. The set $\mathcal{S}(N, \boldsymbol{m})$ is defined by the following set of inequalities:

$$\begin{cases} V_1 - v_1 & \geq 0 \\ V_2 - v_2 & \geq 0 \\ v_1 & \geq 0 \\ v_2 - V_2' & \geq 0 \\ v_1 - v_2 & \geq 0 \end{cases} \tag{8}$$

that clearly admits no feasible solution. ■

B. Conflict-Free Firing Speed Computation

By the formalism previously introduced, we define the concept of *conflict* in a net. We will only treat conflicts at continuous places, since the computation of an admissible IFS vector is only affected by this type of conflicts.

Example 11: Consider the net shown in Fig. 3(b). When place p is not empty, both t_2 and t_3 can fire at their MFS. When place p is empty, however, the output flow $v_2 + v_3$ is bounded by the input flow v_1, thus in $\mathcal{S}(N, \boldsymbol{m})$ there will be a constraint of the form (5d) related to place p that writes $v_1 \geq v_2 + v_3$. This constraint expresses the fact that we have a limited amount of resource (the input flow) that must be shared between different processes (the output transitions). ■

There is no conflict in a net, instead, if each empty place $p \in P_c$ has at most one enabled output transition $t \in T_c$. This motivates the next definition.

Definition 12: Let $\langle N, \boldsymbol{m} \rangle$ be an FOHPN system and $\mathcal{S}(N, \boldsymbol{m})$ be the linear set defined by (5). We say that N is *continuous conflict-free* (CCF) at $\boldsymbol{m}$ if for all constraints of the form (5d) rewritten as (7) holds: $\text{card}\{K\} \leq 1$. ■

In the rest of this section, we discuss the relationship between conflict resolution (i.e., the computation of IFS vectors) and performance optimization.

If we set our goal to maximize the firing speed of the continuous transitions, it is possible to show that in a continuous conflict-free FOHPN each component of the IFS vector may be maximized independently.

Theorem 13: Let $\langle N, \boldsymbol{m} \rangle$ be an FOHPN system. If N is CCF at $\boldsymbol{m}$, the optimal solution $\boldsymbol{v}^*$ of the following LPP

$$\begin{aligned} \max \quad & \mathbf{1}^T \cdot \boldsymbol{v} \qquad \text{s.t.} \\ & \boldsymbol{v} \in \mathcal{S}(N, \boldsymbol{m}) \end{aligned}$$

is such that $\forall \boldsymbol{v} \in \mathcal{S}(N, \boldsymbol{m})$, $\boldsymbol{v} \leq \boldsymbol{v}^*$ (componentwise).

Proof: Let $\oplus$ be the (componentwise) max operator, i.e., $\boldsymbol{w} \oplus \boldsymbol{y} \equiv (w_i \oplus y_i)_i \equiv (\max\{w_i, y_i\}_i$. It is sufficient to prove that if the net is CCF, then $\boldsymbol{w}, \boldsymbol{y} \in \mathcal{S}(N, \boldsymbol{m}) \implies \boldsymbol{w} \oplus \boldsymbol{y} \in \mathcal{S}(N, \boldsymbol{m})$.

Clearly, if $\boldsymbol{w}$ and $\boldsymbol{y}$ satisfy (5), then $\boldsymbol{w} \oplus \boldsymbol{y}$ will satisfy all constraints of the form (5a)–(5c). Under the hypothesis of conflict-freeness, we can write any constraint of the form (5d) associated to a place p as follows:

1) $\sum_{j \in J} \alpha_j v_j \geq 0$ if no enabled transition outputs from place p;
2) $\sum_{j \in J} \alpha_j v_j \geq \alpha_{\text{out}} v_{\text{out}}$ if t_{out} is the only enabled transition outputting from place p;

with $\alpha_j, \alpha_{\text{out}}, v_j, v_{\text{out}} \in \mathbb{R}_0^+$.

In the first case, we have that

$$\begin{aligned} \sum_{j \in J} \alpha_j (w_j \oplus y_j) &\geq \left(\sum_{j \in J} \alpha_j w_j \right) \oplus \left(\sum_{j \in J} \alpha_j y_j \right) \\ &\geq 0 \oplus 0 = 0 \end{aligned}$$

while in the second case, we have

$$\begin{aligned} \sum_{j \in J} \alpha_j (w_j \oplus y_j) &\geq \left(\sum_{j \in J} \alpha_j w_j \right) \oplus \left(\sum_{j \in J} \alpha_j y_j \right) \\ &\geq (\alpha_{\text{out}} w_{\text{out}}) \oplus (\alpha_{\text{out}} y_{\text{out}}) \\ &= \alpha_{\text{out}} (w_{\text{out}} \oplus y_{\text{out}}) \end{aligned}$$

i.e., the vector $\boldsymbol{w} \oplus \boldsymbol{y}$ satisfies all constraints of the form (5d) as well. □

In the case of CCF nets, the optimal solution $\boldsymbol{v}^*$ in the previous theorem coincides with the solution computed with the priority algorithm in [3]. It may be interesting, however, to compare the two algorithms via an example.

Example 14: Let us consider again the net in Fig. 3(a) whose set of admissible IFS vectors is given by (6). If we compute the vector $\boldsymbol{v}^*$ solution of (6) that maximizes $J = v_1 + v_2$, we clearly obtain $v_1^* = V_1$ and $v_2^* = \min\{(1/(1-\alpha))V_1, V_2\}$. This example is so simple that we can write the solution in closed form; in more complex cases, the solution can still be easily found solving the associated LPP. If we apply the procedure proposed in [3], we obtain at the first iteration step $v_1 = V_1$, while to compute the IFS of transition t_2 we need to solve the following iterative problem:

$$\begin{cases} v_2^0 & = 0 \\ v_2^{r+1} & = \min(V_1 + \alpha \cdot v_2^r, V_2) \end{cases}$$

and for $V_1 \leq (1-\alpha)V_2$, the algorithm requires an infinite number of steps to converge to the correct value $v_2 = (1/(1-\alpha))V_1$. ■

C. Global Conflict Resolution

When the net is not conflict-free, not all firing speeds may be maximized independently. However, we can always find out a conflict resolution policy by solving an LPP aimed at a *global optimization* of the system resources. We may consider different performance indices as the objective function in the LP formulation of the problem. We consider some examples.

Maximize Flows: In an FOHPN, we may consider as optimal the solution $\boldsymbol{v}^*$ of (5) that maximizes the performance index $J = \mathbf{1}^T \cdot \boldsymbol{v}$, which is of course intended to maximize the sum over all flow rates. In the manufacturing domain, this may correspond to maximizing machines utilization.

Maximize Outflows: In an FOHPN, we may want to maximize the performance index $J = \boldsymbol{a}^T \cdot \boldsymbol{v}$ where

$$a_j = \begin{cases} 1, & \text{if } t_j \text{ is an exogenous transition,} \\ 0, & \text{if } t_j \text{ is an endogenous transition.} \end{cases}$$

In the manufacturing domain, this may correspond to maximizing throughput.

Dynamic Flow Balancing: This problem consists in reducing the difference between maximum and minimum utilization of continuous transitions. The utilization of a transition $t_j \in T_c$ can be given as the ratio v_j/V_j. Then, we may want to minimize the performance index $J = \max_{j\in K}\{v_j/V_j\} - \min_{j \in K}\{v_j/V_j\}$ for a suitable index set K. In the manufacturing domain, this may correspond to balancing the machines load.

Minimize Stored Fluid: In an FOHPN, we may want to minimize the derivative of the marking of a place $p \in P_c$. This can be done by minimizing the performance index $J = \boldsymbol{a}^T \cdot \boldsymbol{v}$ where

$$a_j = \begin{cases} \boldsymbol{C}(p, t_j), & \text{if } t_j \in p^{(c)} \cup {}^{(c)}p, \\ 0, & \text{otherwise.} \end{cases}$$

In the manufacturing domain, this may correspond to minimizing the work-in-process (WIP).

A different optimization procedure is based on *global priorities* (GP). In this case, we have a multiobjective performance in which the goals have different priorities. We first look for all solutions that optimize the first goal, then among them for those that optimize the second goal, and so forth. We discuss a simple case in which each goal consists in maximizing the IFS of a single transition, though this result can be easily generalized.

Definition 15: Let $\langle N, \boldsymbol{m}\rangle$ be an FOHPN system and $\mathcal{S}(N, \boldsymbol{m})$ be the linear set defined by (5). Assume that the continuous transitions of the net are ordered in a priority sequence $t_1 \succ t_2 \succ \cdots \succ t_{n_c}$. The GP-optimal solution for $\mathcal{S}(N, \boldsymbol{m})$ is defined by

$$\begin{aligned} v_1^* &= \max\{v_1 | \boldsymbol{v} \in \mathcal{S}(N, \boldsymbol{m})\} \\ v_2^* &= \max\{v_2 | \boldsymbol{v} \in \mathcal{S}(N, \boldsymbol{m}), v_1 = v_1^*\} \\ v_3^* &= \max\{v_2 | \boldsymbol{v} \in \mathcal{S}(N, \boldsymbol{m}), v_1 = v_1^*, v_2 = v_2^*\} \\ &\cdots \end{aligned}$$

where $\boldsymbol{v}^* = [v_1^* \cdots v_{n_c}^*]^T$. ■

The GP-optimal solution can be found by solving n_c LPP. First, we compute v_1^*; then, we add to (5) the constraint $v_1 = v_1^*$ and maximize $J = v_2$, etc. Note, however, that there exist other techniques based on lexicographic ordering [9] that may well be meaningfully used to compute the GP-optimal solution by solving a single LPP with a suitably modified objective function.

Example 16: Consider the net in Fig. 3(c) with $V_1 = V_5 = 10$, $V_2 = V_3 = V_4 = 7$, and places p_1, p_2 initially empty. We apply the method discussed above to obtain $\boldsymbol{v}^* = [10 \ \ 7 \ \ 3 \ \ 3 \ \ 10]^T$. Note that by applying the algorithm proposed in [3], we obtain $\boldsymbol{v} = [10 \ \ 3 \ \ 7 \ \ 7 \ \ 10]^T$, which is an admissible IFS vector even though it does not have the same properties of the GP-optimal solution. ■

According to the next theorem, a GP-optimal solution is a basic solution of any LPP subject to $\mathcal{S}(N, \boldsymbol{m})$, hence, it is amenable to sensitivity analysis as it will be discussed in the following section.

Theorem 17: The GP-optimal solution $\boldsymbol{v}^*$ is unique and it is a vertex, i.e., a basic solution, of the feasible region $\mathcal{S}(N, \boldsymbol{m})$.

Proof: Let $\mathcal{S}_0(N, \boldsymbol{m}) = \mathcal{S}(N, \boldsymbol{m})$ and $\mathcal{S}_i(N, \boldsymbol{m}) = \mathcal{S}_{i-1}(N, \boldsymbol{m}) \cap \{\boldsymbol{v} | v_i = v_i^*\}$ for $i = 1, \cdots, n_c - 1$. We will prove that all vertices of $\mathcal{S}_i(N, \boldsymbol{m})$ are also vertices of $\mathcal{S}_{i-1}(N, \boldsymbol{m})$. In fact, the hyperplane $\{\boldsymbol{v}|v_i = v_i^*\}$ does not cut the convex set $\mathcal{S}_{i-1}(N, \boldsymbol{m})$ in any internal point because by construction $\{\boldsymbol{v}|v_i > v_i^*\} \cap \mathcal{S}_{i-1}(N, \boldsymbol{m}) = \emptyset$. The solution $\boldsymbol{v}^*$ is necessarily unique because for all $\boldsymbol{v}'$ if $v_i' = v_i^*$ for $i = 1, \cdots, p < n_c$ and $v_{p+1}' < v_{p+1}^*$ then $\boldsymbol{v}'$ cannot be a GP-optimal solution. □

D. Local Conflict Resolution

The use of a performance index to be maximized (or minimized) over the space of all admissible IFS vectors corresponds to a global optimization procedure. It is often the case, however, that local rules are used to determine the operating mode of a system described by a hybrid net. These rules correspond to decisions that can be taken in a decentralized way.

We consider the case of nets where all conflicts are *free-choice*, i.e., if a continuous place p has more than one output continuous transition (e.g., $p^{(c)} = \{t_1, t_2, \cdots t_k\}$ with $k > 1$), then it is the only continuous input place for all those transitions (i.e., ${}^{(c)}t_j = \{p\},\ j = 1, \cdots, k$). The conflict in Fig. 3(b) is free-choice, while the two conflicts in Fig. 3(c) are not. When the conflicts are not free-choice, the local optimization rules described below may not be well founded.

Fixed Ratio: One particular simple rule that may be used to locally solve free-choice conflicts is that of assigning a *fixed ratio* of fluid volume to all enabled continuous transitions that take fluid out of an empty continuous place. As an example, in Fig. 3(b), we may assign a ratio $v_2 = s\,v_3,\ s \in \mathbb{R}$. This new constraint can be added to the set $\mathcal{S}$ or, even better, by substitution we can reduce by one the number of variables in (5).

Local Priorities: We can also consider the case of *local priority rules* by a suitable modification of the linear set (5). Assume that in Fig. 3(b) a legal solution is such that t_2 has priority over t_3, i.e., all fluid entering place p should be consumed by t_2 and only if $v_2 = V_2$ the remaining fluid should be consumed by t_3. This can be done adding the following constraints:

$$\begin{cases} M\,x \geq V_2 - v_2 \\ v_3 \leq M\,(1-x) \end{cases}$$

where $x \in \{0, 1\}$, $M \in \mathbb{R}$ with $M \gg 0$. Thus, if $v_2 < V_2$ it follows $v_3 = 0$. The problem with this technique is that a simple LPP is transformed into a more complex mixed integer-linear problem.

IV. Sensitivity Analysis for FOHPN

The LPP stated in the previous section may be solved taking into account only the constraints related to enabled transitions since we know that the IFS of transitions that are not enabled are 0. Let $I_t = \{\alpha_1, \cdots, \alpha_k\}$ be the set of indices of the enabled continuous transitions and $I_p = \{\alpha_{2k+1}, \cdots, \alpha_\ell\}$ be the set of indices of the empty continuous places. Thus, we can write

$$\max \sum_{j \in I_t} c_j\, v_j \qquad \text{s.t.}$$
$$\begin{cases} v_{\alpha_1} + s_1 & = V_{\alpha_1} \\ \cdots \\ v_{\alpha_k} + s_k & = V_{\alpha_k} \\ v_{\alpha_1} - s_{k+1} & = V'_{\alpha_1} \\ \cdots \\ v_{\alpha_k} - s_{2k} & = V'_{\alpha_k} \\ \sum\limits_{j \in I_t} \boldsymbol{C}(p_{\alpha_{2k+1}}, t_j) v_j - s_{2k+1} & = 0 \\ \cdots \\ \sum\limits_{j \in I_t} \boldsymbol{C}(p_{\alpha_\ell}, t_j) v_j - s_\ell & = 0 \end{cases}$$
$$s_j \geq 0. \qquad (9)$$

Defining vector $\boldsymbol{x} = [v_{\alpha_1} \cdots v_{\alpha_k}\ s_1 \cdots s_\ell]^T$, we obtain the following standard form:

$$\max_{\boldsymbol{x}} \{\boldsymbol{c}^T \boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq 0\}. \qquad (10)$$

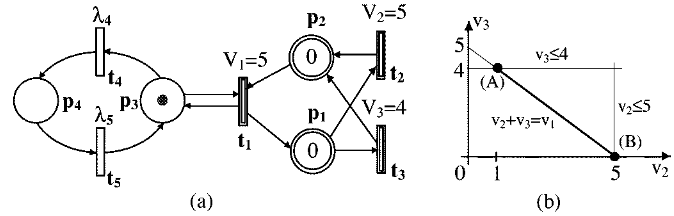

Fig. 4. (a) FOHPN model of a manufacturing service. (b) The feasible region for the IFS vectors.

Note that when $V'_{\alpha_j} = 0$, the slack s_{k+j} and the corresponding constraint can be removed from the LPP given by (9) by adding a nonnegativity constraint on v_{α_j}.

Here $\boldsymbol{x}$ is a vector with $\ell + k$ variables, $\boldsymbol{A}$ is the $\ell \times (\ell + k)$ matrix constraints, and we assume that $\boldsymbol{A}$ has full rank, $\boldsymbol{c}$ is the $(\ell + k)$-vector of the objective coefficients, while $\boldsymbol{b}$ represents the ℓ-vector of the right-hand side constants.

In this work, the *simplex method* will be used to solve LPP. This is an iterative method in which at each step and in an efficient manner, a new basis is computed. Each basis represents a vertex of the feasible region. We denote an optimal solution $\boldsymbol{x}^o$, the corresponding optimal basis $\mathcal{B}$ (a set of ℓ variables), and $\boldsymbol{A}_\mathcal{B}$ the optimal basis matrix obtained by taking only those columns of $\boldsymbol{A}$ whose corresponding variables are in $\mathcal{B}$. An optimal solution $\boldsymbol{x}^o$ can always be written as

$$\boldsymbol{x}^o = \begin{bmatrix} \boldsymbol{x}^o_\mathcal{B} \\ \boldsymbol{x}^o_\mathcal{N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_\mathcal{B}^{-1} \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix}.$$

The variables in $\mathcal{B}$ are the basic variables, while the others, whose set is denoted $\mathcal{N}$, are called nonbasic. Note that the optimal solution may be degenerate, i.e., we have many basis associated with it. It may also be the case that more than one basic optimal solution exists.

Example 18: In Fig. 4(a), it is represented the FOHPN model of a manufacturing service, where transition t_1 models an unreliable machine and transitions t_2 and t_3 represent the outflows from buffer p_1. A buffer capacity 0 is imposed by the co-buffer place p_2. The maximum production rate of the machine is bounded by the MFS V_1, while the maximum outflows rates cannot exceed V_2 and V_3, respectively. The discrete part of the net models the failure/repair stochastic process of the machine by means of exponential transitions t_4 and t_5 with average firing rates λ_4 and λ_5, respectively. The machine is operating while place p_3 is marked (i.e., transition t_1 is enabled) and it is down when place p_4 is marked.

The constraint set associated to this net from the given marking is

$$\begin{cases} v_1 & \leq 5 \\ v_2 & \leq 5 \\ v_3 & \leq 4 \\ -v_1 + v_2 + v_3 & \leq 0 \\ v_1 - v_2 - v_3 & \leq 0 \\ v_1,\ v_2,\ v_3 & \geq 0. \end{cases} \qquad (11)$$

We take as objective function to be maximized $J = v_2 + v_3$, representing the overall output flow, and we obtain the following LPP in standard form:

$$\max v_2 + v_3 \quad \text{s.t.} \quad \begin{cases} v_1 + s_1 & = 5 \\ v_2 + s_2 & = 5 \\ v_3 + s_3 & = 4 \\ v_1 - v_2 - v_3 & = 0 \end{cases}$$

Note that we have not written the nonnegativity constraints and we have packed together the last two inequalities of (11). There are infinitely many optimal solutions of the form $v_1 = 5$, $v_2 = y$, $v_3 = 5 - y$ with $y \in [1, 5]$, represented by the thick line in Fig. 4(b) in the plane $v_1 = 5$. Two of these are basic solutions: $\boldsymbol{v}_{(A)} = [5 \quad 1 \quad 4]^T$ and $\boldsymbol{v}_{(B)} = [5 \quad 5 \quad 0]^T$. Point (A) is a nondegenerate solution with basic variables v_1, v_2, v_3, s_2 and basis $\mathcal{B}_A = \{v_1, v_2, v_3, s_2\}$. Point (B) is a degenerate solution with two optimal basis: $\mathcal{B}_{B1} = \{v_1, v_2, v_3, s_3\}$ and $\mathcal{B}_{B2} = \{v_1, v_2, s_2, s_3\}$. Furthermore, we observe that in (B) there is also another basis $\mathcal{B}_{B3} = \{v_1, v_2, s_1, s_3\}$, which is not optimal. ■

Sensitivity analysis refers to the study of how optimal solutions change according to changes of the given linear program in terms of the coefficients of the matrix, the right-hand side and the objective function. Suppose that the LPP (10) has an optimal solution. If there is any change in the values of b_j, c_j, or a_{ij}, the optimal solution is likely to change in general.

In the next sections, we will develop sensitivity analysis with respect to the design parameters by assuming changes in the right-hand side vector and in the matrix coefficients. Perturbations in the cost coefficients will not be considered in this work.

A. Perturbed Model

The perturbed linear programming problem considered in this paper is defined as follows:

$$\max_{\boldsymbol{x}} \{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{A}(\boldsymbol{q})\boldsymbol{x} = \boldsymbol{b}(\boldsymbol{q}),\ \boldsymbol{x} \geq 0\} \tag{12}$$

where $\boldsymbol{q} = [q_0 \cdots q_p]^T$ is a vector of uncertain parameters. The nominal value is denoted $\overline{\boldsymbol{q}}$. For a given value of $\boldsymbol{q}$, the optimal solution of (12) is

$$\boldsymbol{x}^o(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{x}_{\mathcal{B}}(\boldsymbol{q}) \\ \boldsymbol{x}_{\mathcal{N}}(\boldsymbol{q}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{B}}^{-1}(\boldsymbol{q})\boldsymbol{b}(\boldsymbol{q}) \\ \boldsymbol{0} \end{bmatrix}. \tag{13}$$

We compute with the simplex method an optimal solution in $\overline{\boldsymbol{q}}$ and the corresponding optimal basis $\mathcal{B}$. The sensitivity of the basic variables $\boldsymbol{x}_{\mathcal{B}}(\overline{\boldsymbol{q}})$ with respect to q_i can be computed, at least within a certain domain where the optimal basis does not change, by taking the partial derivatives

$$\frac{\partial \boldsymbol{x}_{\mathcal{B}}(\overline{\boldsymbol{q}})}{\partial q_i} = \boldsymbol{A}_{\mathcal{B}}^{-1}(\overline{\boldsymbol{q}}) \left(\frac{\partial \boldsymbol{b}(\overline{\boldsymbol{q}})}{\partial q_i} - \frac{\partial \boldsymbol{A}_{\mathcal{B}}(\overline{\boldsymbol{q}})}{\partial q_i} \boldsymbol{x}_{\mathcal{B}}(\overline{\boldsymbol{q}}) \right) \tag{14}$$

while the nonbasic variables $\boldsymbol{x}_{\mathcal{N}}(\overline{\boldsymbol{q}})$ do not change. Equation (14) shows the effect on the optimal solution caused by a small change of q_i. It is only required first-order differentiability of $\boldsymbol{A}_{\mathcal{B}}(\overline{\boldsymbol{q}})$ and $\boldsymbol{b}(\overline{\boldsymbol{q}})$ with respect to q_i. Furthermore, if the optimal solution is not degenerate, then the obtained sensitivity is unique. For simplicity in this presentation, we make the following assumptions.

1) Only one parameter q_i varies at a time, that is, $\boldsymbol{q} = \overline{\boldsymbol{q}} + \lambda \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the ith canonical basis vector. Under this assumption, the sensitivity given by (14) can be regarded as function of λ in the allowable range.
2) Matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ are linear functions of the parameter λ. Thus, we can write
$$\boldsymbol{A}_{\mathcal{B}}(\lambda) = \boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{A}_{\mathcal{B}}^*$$
$$\boldsymbol{b}(\lambda) = \boldsymbol{b} + \lambda \boldsymbol{b}^*$$
where $\boldsymbol{A}_{\mathcal{B}} = \boldsymbol{A}_{\mathcal{B}}(\overline{\boldsymbol{q}})$, $\boldsymbol{b} = \boldsymbol{b}(\overline{\boldsymbol{q}})$.
3) The variation of each parameter q_i influences only one column, say the jth, of matrix $\boldsymbol{A}_{\mathcal{B}}(\lambda)$. Then
$$\boldsymbol{A}_{\mathcal{B}}(\lambda) = \boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{A}_{\mathcal{B}}^* = \boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T.$$

In what follows, we consider separately linear perturbations of the right-hand side vector and of the matrix coefficients.

B. Perturbation of the Right-Hand Side Vector

We assume that the right-hand side constant vector $\boldsymbol{b}$ varies linearly with the parameter $\lambda \in \mathbb{R}$, that is, $\boldsymbol{b}(\lambda) = \boldsymbol{b} + \lambda \boldsymbol{b}^*$. In the FOHPN framework, this perturbation corresponds to changes in the entries of the vector $\boldsymbol{V} = [V_{\alpha_1} \cdots V_{\alpha_k}]^T$, which denotes the MFS vector, and of the vector $\boldsymbol{V'} = [V'_{\alpha_1} \cdots V'_{\alpha_k}]^T$, which denotes the mfs vector. As an example, in a manufacturing system we may want to add servers to a machine in order to increase the overall productivity of the system.

If only V_{α_i} is perturbed, then $\boldsymbol{b}^* = \boldsymbol{e}_i$ for $i = 1, \cdots, k$. We may also consider the case where V_{α_i} and V_{α_j} vary simultaneously with the parameter λ. As an example, if we consider that some servers are shifted from transition t_j to t_i or vice versa, then we have $V_{\alpha_i} = \overline{V}_{\alpha_i} + \lambda$ and $V_{\alpha_j} = \overline{V}_{\alpha_j} - \lambda$, hence $\boldsymbol{b}^* = \boldsymbol{e}_i - \boldsymbol{e}_j$. Similar considerations apply when the mfs of a continuous transition is perturbed.

Let $\boldsymbol{x}^o$ be an optimal basic solution of (10) and $\mathcal{B}$ an associated optimal basis. The perturbed optimal solution $\boldsymbol{x}^o(\lambda)$ has basic components

$$\boldsymbol{x}_{\mathcal{B}}^o(\lambda) = \boldsymbol{A}_{\mathcal{B}}^{-1}\boldsymbol{b}(\lambda) = \boldsymbol{A}_{\mathcal{B}}^{-1}(\boldsymbol{b} + \boldsymbol{b}^*\lambda) = \boldsymbol{x}_{\mathcal{B}}^o + \lambda \boldsymbol{x}_{\mathcal{B}}^* \tag{15}$$

where $\boldsymbol{x}_{\mathcal{B}}^o = \boldsymbol{A}_{\mathcal{B}}^{-1}\boldsymbol{b} = [\beta_1 \cdots \beta_\ell]^T$ and $\boldsymbol{x}_{\mathcal{B}}^* = \boldsymbol{A}_{\mathcal{B}}^{-1}\boldsymbol{b}^* = [\beta_1^* \cdots \beta_\ell^*]^T$. The optimal value of the objective function is

$$J(\lambda) = \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}_{\mathcal{B}}^o(\lambda) = \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}_{\mathcal{B}}^o + \lambda \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}_{\mathcal{B}}^* = J + \lambda J^*. \tag{16}$$

When $\boldsymbol{b}^* = \boldsymbol{e}_i$, the derivative of the objective function with respect to the parameter λ, i.e., $(dJ(\lambda)/d\lambda) = J^*$, is also called *dual price* of the ith resource. It represents the amount by which the optimum will increase if the availability of the resource associated to the ith constraint (i.e., the right-hand side of the constraint) is increased by one unit.

Equations (15) and (16) hold only when λ belongs to a certain interval $\Lambda_{\mathcal{B}} = [\underline{\lambda}_{\mathcal{B}}, \overline{\lambda}_{\mathcal{B}}]$, also called the *allowable range*, where the optimal basis $\mathcal{B}$ remains unchanged. This requires nonnegativity of the basic variables, $\boldsymbol{x}^o_{\mathcal{B}}(\lambda) \geq \boldsymbol{0}$, and the bounds for the parameter λ can be computed as follows:

$$\underline{\lambda}_{\mathcal{B}} = \begin{cases} -\infty, & \text{if} \quad I^+ = \emptyset \\ \max_{i \in I^+} \left\{ -\dfrac{\beta_i}{\beta_i^*} \right\} \end{cases} \tag{17}$$

and

$$\overline{\lambda}_{\mathcal{B}} = \begin{cases} +\infty, & \text{if} \quad I^- = \emptyset \\ \min_{i \in I^-} \left\{ -\dfrac{\beta_i}{\beta_i^*} \right\} \end{cases} \tag{18}$$

where $I^+ = \{i \geq 1 | \beta_i^* > 0\}$ and $I^- = \{i \geq 1 | \beta_i^* < 0\}$. Since $\boldsymbol{A}_{\mathcal{B}}^{-1}$ is invertible, then $\boldsymbol{A}_{\mathcal{B}}^{-1}\boldsymbol{b}^* \neq \boldsymbol{0}$, i.e., either $\underline{\lambda}_{\mathcal{B}}$ or $\overline{\lambda}_{\mathcal{B}}$ must be finite.

Much attention has been devoted in the literature [18], [15] to the case in which the optimal solution $\boldsymbol{x}^o$ of the nominal LPP is unique. In this case, $\boldsymbol{x}^o$ is not a degenerate solution and the unique optimal basis remains constant within the allowable range, therefore the value of the objective function is linear in λ. As λ reaches the boundary of the allowable range, a degenerate solution is found, a new basis can be computed with an allowable range that will not overlap the previous one except at the end points. As the basis changes, the derivative of the objective function with respect to the parameter λ, i.e., $(dJ(\lambda)/d\lambda) = J^*$, may also change, thus it may not be defined only at a finite number of points whereas we can instead provide right and left values. In the manufacturing domain, this non-differentiability behavior has been already observed in tandem lines by Fu and Suri [20] when the average production rates of two machines are equal. With our approach, the result is immediately generalized to more general cases.

However, the situation can be more complex when more than one optimal solution exists, as we show in the following example. Multiple optimal solutions represent the degrees of freedom in the optimization procedure.

Example 19: Let us consider again the net in Example 18. There are two optimal basic solutions, (A) and (B), and three optimal bases. We apply the previous methodology to each basis to obtain the following allowable ranges: $\Lambda_{\mathcal{B}_A} = [-1, 4]$, $\Lambda_{\mathcal{B}_{B1}} = [0, 4]$, and $\Lambda_{\mathcal{B}_{B2}} = [-5, 0]$. As expected, the intervals $\Lambda_{\mathcal{B}_{B1}}$ and $\Lambda_{\mathcal{B}_{B2}}$, corresponding to the same optimal basic solution (B), do not overlap. However, we note that the interval $\Lambda_{\mathcal{B}_A}$ corresponding to the optimal basic solution (A) overlaps both of them. This observation allows us to state that the interval in which the derivative of the objective function remains constant is $\Lambda = [-5, 4]$, hence, it is larger than the allowable range associated to each basis. ■

Motivated by the previous example, we can state the next proposition that applies to the case in which there are two optimal basic solutions of a given LPP and that can be naturally extended to the case of more than two solutions.

Proposition 20: Let $\boldsymbol{x}_0^{nd}$ and $\boldsymbol{x}_0^d$ be the optimal basic solutions of the LPP (10), and let the perturbed solutions take the form given by (15). Let $\boldsymbol{x}_0^{nd}$ be a nondegenerate optimal solution with allowable range $\Lambda_{\mathcal{B}_1} = [\underline{\lambda}_{\mathcal{B}_1}, \overline{\lambda}_{\mathcal{B}_1}]$ associated to the unique optimal basis $\mathcal{B}_1$, and $\boldsymbol{x}_0^d$ be a degenerate optimal solution with allowable ranges $\Lambda_{\mathcal{B}_2} = [\underline{\lambda}_{\mathcal{B}_2}, 0]$ and $\Lambda_{\mathcal{B}_3} = [0, \overline{\lambda}_{\mathcal{B}_3}]$ associated to the optimal basis $\mathcal{B}_2$ and $\mathcal{B}_3$, respectively. Then, the derivative $dJ(\lambda)/d\lambda$ of the objective function (16) with respect to the parameter λ is continuous and constant over all the interval $\Lambda = \Lambda_{\mathcal{B}_1} \cup \Lambda_{\mathcal{B}_2} \cup \Lambda_{\mathcal{B}_3}$.

Proof: Over each interval $\Lambda_{\mathcal{B}_i}$, for $i = 1, 2, 3$, the derivative $dJ(\lambda)/d\lambda$ is continuous and constant, say J_1^*, J_2^*, and J_3^*. Since $\Lambda_{\mathcal{B}_1} \cap \Lambda_{\mathcal{B}_2}$ is an interval of nonzero length, then $J_1^* = J_2^*$. A similar reasoning shows that $J_1^* = J_3^*$ and this complete the proof. □

C. Perturbation of the Matrix Coefficients

We assume that the basis matrix $\boldsymbol{A}_{\mathcal{B}}$ varies linearly with the parameter $\lambda \in \mathbb{R}$, according to $\boldsymbol{A}_{\mathcal{B}}(\lambda) = \boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{A}_{\mathcal{B}}^* = \boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T$, i.e., we assume that only the jth column of $\boldsymbol{A}_{\mathcal{B}}$ may vary.

In the FOHPN framework, this perturbation of the jth column corresponds to changes in the weights of the arcs between continuous places and transition t_j, as it can be seen from (9). Multiple variations of the coefficients along a column correspond to a redistribution of the inflow or outflow of a single continuous transition. In a manufacturing system, this situation is quite common and it arises when we deal with changes of the percentage of parts that need to be reworked or with changes of the routing coefficients. The results we present here also hold when a single row of $\boldsymbol{A}_{\mathcal{B}}$ varies linearly with the parameter λ. Nevertheless, this case is less relevant in the context of FOHPN.

Let $\boldsymbol{x}^o$ be an optimal basic solution of (10) and $\mathcal{B}$ an associated optimal basis. We recall the matrix equality

$$\boldsymbol{A}_{\mathcal{B}}^{-1}(\lambda) = \left(\boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T\right)^{-1} = \boldsymbol{A}_{\mathcal{B}}^{-1} - \frac{\boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{a}^* \boldsymbol{e}_j^T \boldsymbol{A}_{\mathcal{B}}^{-1}}{1 + \boldsymbol{e}_j^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{a}^* \lambda} \lambda.$$

The perturbed optimal solution $\boldsymbol{x}^o(\lambda)$ has basic components

$$\boldsymbol{x}^o_{\mathcal{B}}(\lambda) = \boldsymbol{A}_{\mathcal{B}}^{-1}(\lambda)\boldsymbol{b} = \boldsymbol{x}^o_{\mathcal{B}} - \frac{\lambda}{1 + v\lambda} \boldsymbol{x}^*_{\mathcal{B}} \tag{19}$$

where $\boldsymbol{x}^o_{\mathcal{B}} = \boldsymbol{A}_{\mathcal{B}}^{-1}\boldsymbol{b}$, $\boldsymbol{x}^*_{\mathcal{B}} = \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{a}^* \boldsymbol{e}_j^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{b}$, and $v = \boldsymbol{e}_j^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{a}^*$. The relative cost coefficient vector of the optimal solution $\boldsymbol{x}^o(\lambda)$ is

$$\boldsymbol{r}(\lambda) = \left(\boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{A}_{\mathcal{B}}^{-1}(\lambda) \boldsymbol{A}\right)^T - \boldsymbol{c} = \boldsymbol{r}^o - \frac{\lambda}{1 + v\lambda} \boldsymbol{r}^* \tag{20}$$

where $\boldsymbol{r}^o = (\boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{A})^T - \boldsymbol{c}$, $\boldsymbol{r}^* = (\boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{a}^* \boldsymbol{e}_j^T \boldsymbol{A}_{\mathcal{B}}^{-1} \boldsymbol{A})^T$. Finally, the optimal value of the objective function is given by

$$J(\lambda) = \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}^o_{\mathcal{B}}(\lambda) = \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}^o_{\mathcal{B}} - \frac{\lambda}{1 + v\lambda} \boldsymbol{c}_{\mathcal{B}}^T \boldsymbol{x}^*_{\mathcal{B}}. \tag{21}$$

Equations (19)–(21) hold only when the parameter λ belongs to a certain interval $\Lambda_{\mathcal{B}} = [\underline{\lambda}_{\mathcal{B}}, \overline{\lambda}_{\mathcal{B}}]$ wherein the optimal basis $\mathcal{B}$ remains unchanged. This requires the following: 1) nonsingularity of the basis matrix, i.e., $1 + v\lambda > 0$; 2) nonnegativity of the basic variables, $\boldsymbol{x}^o_{\mathcal{B}}(\lambda) \geq \boldsymbol{0}$; and 3) nonnegativity of the relative cost coefficients, $\boldsymbol{r}(\lambda) \geq \boldsymbol{0}$, i.e., the optimality condition. Note that the first condition can be written as $\det(\boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T) \neq 0$. Moreover, it holds $\det(\boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T) = (1 + v\lambda) \det \boldsymbol{A}_{\mathcal{B}}$. Since our interest is in the behavior around $\boldsymbol{q} = \overline{\boldsymbol{q}}$, condition (1) becomes that $\det(\boldsymbol{A}_{\mathcal{B}} + \lambda \boldsymbol{a}^* \boldsymbol{e}_j^T) \neq 0$ has the same sign as $\det \boldsymbol{A}_{\mathcal{B}}$

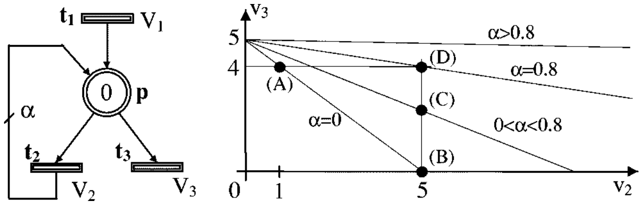

Fig. 5. FOHPN model of a re-entrant service and its feasible regions.

and this condition is equivalent to $1+v\lambda>0$. The bounds for the parameter λ can be computed as follows. Let us define

$$\boldsymbol{y}=\begin{bmatrix}1\\ \boldsymbol{x}_{\mathcal{B}}^{o}\\ \boldsymbol{r}^{o}\end{bmatrix}\qquad \boldsymbol{y}^{*}=\begin{bmatrix}0\\ \boldsymbol{x}_{\mathcal{B}}^{*}\\ \boldsymbol{r}^{*}\end{bmatrix}$$

and let us consider the following sets of indices: $I^{+}=\{i\geq 1|(vy_i-y_i^{*})>0\}$ and $I^{-}=\{i\geq 1|(vy_i-y_i^{*})<0\}$. Then, we can easily find

$$\underline{\lambda}_{\mathcal{B}}=\begin{cases}-\infty, & \text{if}\quad I^{+}=\emptyset\\ \max\limits_{i\in I^{+}}\left\{-\dfrac{y_i}{vy_i-y_i^{*}}\right\}\end{cases}\tag{22}$$

and

$$\overline{\lambda}_{\mathcal{B}}=\begin{cases}+\infty, & \text{if}\quad I^{-}=\emptyset\\ \min\limits_{i\in I^{-}}\left\{-\dfrac{y_i}{vy_i-y_i^{*}}\right\}.\end{cases}\tag{23}$$

From (19) and (21), we observe that the optimum IFS vector and the objective function do not vary linearly with the parameter λ within the allowable interval $\Lambda_{\mathcal{B}}=[\underline{\lambda}_{\mathcal{B}},\overline{\lambda}_{\mathcal{B}}]$ as it does happen if the perturbations of the matrix coefficients are made infinitesimally small. Therefore, the gradient of the objective function with respect to the jth column vector of $\boldsymbol{A}$, say $\boldsymbol{a}_j=\lambda\boldsymbol{a}^{*}$, is a nonlinear function of the parameter λ. In particular, if $v\neq 0$, for each value of $\lambda\in\Lambda_{\mathcal{B}}$ such that $\lambda\neq-(1/v)$, the derivative of the objective function with respect to the parameter λ can be easily computed as

$$\frac{dJ(\lambda)}{d\lambda}=-\frac{1}{(1+v\lambda)^2}\boldsymbol{c}_{\mathcal{B}}^{T}\boldsymbol{x}_{\mathcal{B}}^{*}.\tag{24}$$

Note that in the case of $v=0$, the objective function $J(\lambda)$ varies linearly with the parameter λ within the allowable interval $\Lambda_{\mathcal{B}}$.

D. Example: Sensitivity Analysis for a Re-Entrant Service

In this section, we consider a simple FOHPN which represents a re-entrant service, as shown in Fig. 5, that will clarify our developments. In this net transition, t_1 models the production of a machine whose maximum production rate is bounded by the MFS V_1, while the maximum outflow rates cannot exceed V_2 and V_3, respectively. The routing coefficient α, with $0\leq\alpha\leq 1$, represents the percentage of parts that are required to be reworked on the machine (reworking factor).

From the given marking, being place p empty, the constraint set associated to this net is

$$\begin{cases}v_1+s_1 & =V_1\\ v_2+s_2 & =V_2\\ v_3+s_3 & =V_3\\ -v_1+v_2(1-\alpha)+v_3+s_4 & =0.\end{cases}\tag{25}$$

Now solving for $\max\{v_2+v_3\}$ subject to (25), we obtain the optimum firing speed allocation (production rates) which maximizes the machine utilization. As discussed in the previous sections, this LP formulation allows us to make sensitivity analysis, that is, we can make perturbations of the elements of the LPP, e.g., the reworking factor α, the maximum machine production rate V_1, and the maximum outflow rates V_2 and V_3, to perform optimization. First, we consider the case in which α is changed to $\alpha+\Delta\alpha$ and then the case in which V_i are changed to $V_i+\Delta V_i$.

Let $V_1=5$, $V_2=5$, and $V_3=4$. In Fig. 5, we have shown the feasible regions in the plane $v_1=5$ for this LPP. The thin lines labeled by the different values of α represent the fourth constraint. Note that for $\alpha=0$, we obtain the same results already developed in Example 18, where we have two optimal basic solutions $\boldsymbol{v}_{(A)}=[V_1,\,V_2-V_3,\,V_3]^T$ and $\boldsymbol{v}_{(B)}=[V_1,\,V_2,\,0]^T$, i.e., points (A) and (B), and the optimal value of the objective function J is equal to V_2. For $0<\alpha<(V_3/V_2)$, there is a unique nondegenerate optimal basic solution [point (C)]

$$\boldsymbol{x}_{\mathcal{B}_C}^{o}=[V_1,\;V_2,\;V_1-(1-\alpha)V_2,\;V_3-V_1+(1-\alpha)V_2]^T$$

with an associated optimal basis $\mathcal{B}_C=\{v_1,\,v_2,\,v_3,\,s_3\}$, which yields an optimal objective function value equal to $V_1+\alpha V_2$. For $\alpha=(V_3/V_2)$, we have a degenerate optimal basic solution [point (D)]. Finally, for $\alpha>(V_3/V_2)$, the fourth constraint becomes redundant and the unique optimal basic solution [point (D)] is simply given by $\boldsymbol{x}_{\mathcal{B}_D}^{o}=[V_1,\;V_2,\;V_3,\;V_1-(1-\alpha)V_2-V_3]^T$ with optimal basis $\mathcal{B}_D=\{v_1,\,v_2,\,v_3,\,s_4\}$ and optimal objective function value equal to V_2+V_3. Therefore, we will only consider perturbations of the parameter α for $\alpha\in(0,\,(V_3/V_2))$, which yield nontrivial sensitivity analysis for the objective function $J=v_2+v_3$.

Now computing the bounds for the parameter α to obtain the allowable range $\Lambda_{\mathcal{B}_C}$ for the optimal basis $\mathcal{B}_C$, we must consider $I^{+}=\{3\}$ and $I^{-}=\{4\}$, where $v=0$ and $\boldsymbol{x}_{\mathcal{B}_C}^{*}=[0\;0\;-V_2\;V_2]^T$. Then, it follows

$$\Lambda_{\mathcal{B}_C}=\left[(1-\alpha)-\frac{V_1}{V_2},\,(1-\alpha)+\frac{V_3-V_1}{V_2}\right]$$

within which we can calculate the partial derivative of the objective function J with respect to the reworking factor α by making use of (24). In this simple case, it does result $(\partial J/\partial\alpha)=V_2$, which is constant over the interval $\Lambda_{\mathcal{B}_C}$.

Now let us suppose that the MFS V_1 is perturbed, that is, V_1 changes to $V_1+\lambda$. Then, applying the method developed in the previous sections, we compute the characteristic interval $\Lambda_{\mathcal{B}_C}$ for the design parameter V_1 as follows:

$$\Lambda_{\mathcal{B}_C}=[\max((1-\alpha)V_2-V_1,\,-V_1),\;V_3-V_1+(1-\alpha)V_2]$$

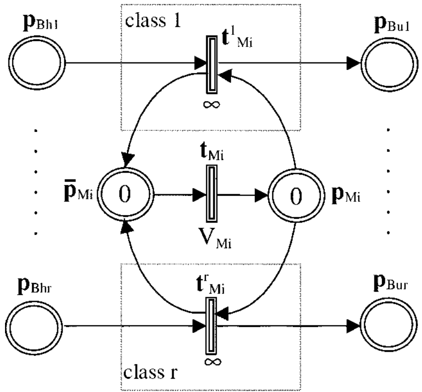

Fig. 6. The FOHPN model of a multiclass machine.

within which the IFS vector and the objective function vary linearly with λ. As a numerical example, if $\alpha = 0.5$, then we have

$$\begin{array}{ll} \Delta V_1 = [V_1 - 2.5,\ V_1 + 1.5], & \text{for } V_1 \\ \Delta V_2 = [V_2 - 3,\ V_2 + 5], & \text{for } V_2 \\ \Delta V_3 = [V_3 - 1.5,\ +\infty], & \text{for } V_3 \end{array}$$

which represent the allowable right-hand side ranges for the basis $\mathcal{B}_C$ to remain unchanged.

V. Modeling Manufacturing Systems with FOHPN

We show in this section how FOHPN's can be used to model manufacturing systems by means of first-order fluid approximations. Indeed, fluid models are well studied and documented in the literature, and the readers are referred to Chen and Mandelbaum [11] for references on the *fluid approximation theory*.

A. Notation and Machine Model

We consider an FMS consisting of a set of n single-server stations among which different classes of continuous flows (fluids) are circulated and processed as in [5]. A machine M_i is represented in an FOHPN by a continuous transition $t_{Mi} \in T_c$, whose firing corresponds to a continuous production at a rate v_{Mi}. A buffer B_j is represented in an FOHPN by a continuous place $p_{Bj} \in P_c$, whose marking represents the current buffer content. Parts of different classes are routed from machines to buffers and vice versa according to their production cycles. A transition associated to the routing, say, from M_i to B_j, is denoted $t_{Mi,\,Bj}$. The occurrence of discrete fail/repair events is modeled by discrete transitions $t_{f,\,Mi}$ and $t_{r,\,Mi}$.

To describe multiclass machines and buffers, it may be necessary to impose synchronization constraints among continuous transitions. As an example, let us consider a multiclass single-input/single-output machine M_i, where parts of class q arrive from buffer B_{hq} and after being processed are routed to buffer B_{uq}. Such a machine is modeled by the net in Fig. 6. Here, the continuous transition t^q_{Mi} ($q = 1, \cdots, r$) represents the flow of parts of class q machined by M_i, and transition t_{Mi}, whose MFS is V_{Mi}, represents the total flow of parts processed by M_i. We assume that the production of any part class is not singularly bounded, i.e., the MFS of each t^q_{Mi} is ∞, while we assume that the machine has an overall maximum production rate, denoted V_{Mi}.

Such a machine is described by the following set of equations:

$$\begin{cases} v_{Mi} \le V_{Mi} & \text{(a)} \\ v_{Mi} = \sum_{q=1}^{r} v^q_{Mi} & \text{(b)} \end{cases} \tag{26}$$

Equation (26b) warrants comment. It is derived from two inequalities

$$\begin{cases} v_{Mi} \ge \sum_{q=1}^{r} v^q_{Mi} \\ v_{Mi} \le \sum_{q=1}^{r} v^q_{Mi}. \end{cases}$$

The first inequality is imposed by the empty continuous place p_{Mi} that has an input arc from t_{Mi} and an output arc to each t^p_{Mi}. The second one is imposed by the empty continuous place $\overline{p}_{Mi}$, that is complementary to place p_{Mi}. These two places form a structure that we call *zero-capacity buffer*. Note that removing v_{Mi}, (26) can be simplified to

$$\sum_{q=1}^{r} v^q_{Mi} \le V_{Mi}. \tag{27}$$

B. Example: Network Layout

We consider the model of an open production system consisting of two shaping machines and an assembly machine with two classes of parts flowing through, as shown in Fig. 7. Parts of classes 1 and 2, coming from external independent sources, are queued in buffers B_1 and B_2, which are both feeding machine M_1, and then start the processing at machine M_1.

The arrival flow of parts of class 1 may be controlled by the plant operator within the range $[V'_{\text{in}\,1},\, V_{\text{in}\,1}]$; the arrival flow of parts of class 2 may be controlled within the range $[0,\, V_{\text{in}\,2}]$. Buffer B_2 has a finite capacity C_{B2} while buffer B_1 has an unlimited capacity. At the exit of machine M_1 parts of class 2 are ready to enter the assembly machine M_a, while parts of class 1 flow into the buffer B_3 with finite capacity C_{B3}, then to machine M_2, where after the processing some parts may require to be reworked on the same machine (parameter α). At the exit of machines M_1 and M_2, parts of both classes are respectively collected in the buffers B_{a1} and B_{a2} with unlimited capacity and then are packed together by the assembly machine according to a specified production mix (parameter β). The maximum machines production rates are denoted V_{M1}, V_{M2}, and V_{Ma}. Since machines are unreliable, we must also take into account a certain failure model.

Although this model may seem quite simple, it captures the key difficulties of common control problems arising in manufacturing systems, such as dynamic scheduling and routing policies as well as production rate selection. Problems of parts routing, admission, and service rate selection have been deeply studied in recent years. In fact, for a simpler production network model than the one proposed here (e.g., a tandem two-station network with two part classes—parts of class 2 visit machine M_1 only,

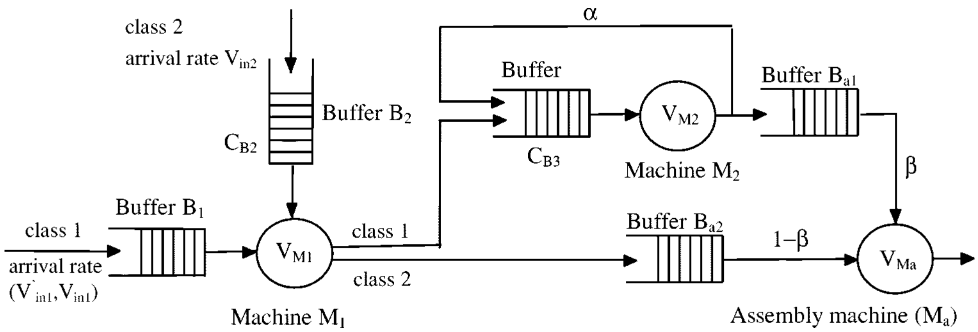

Fig. 7. A production network.

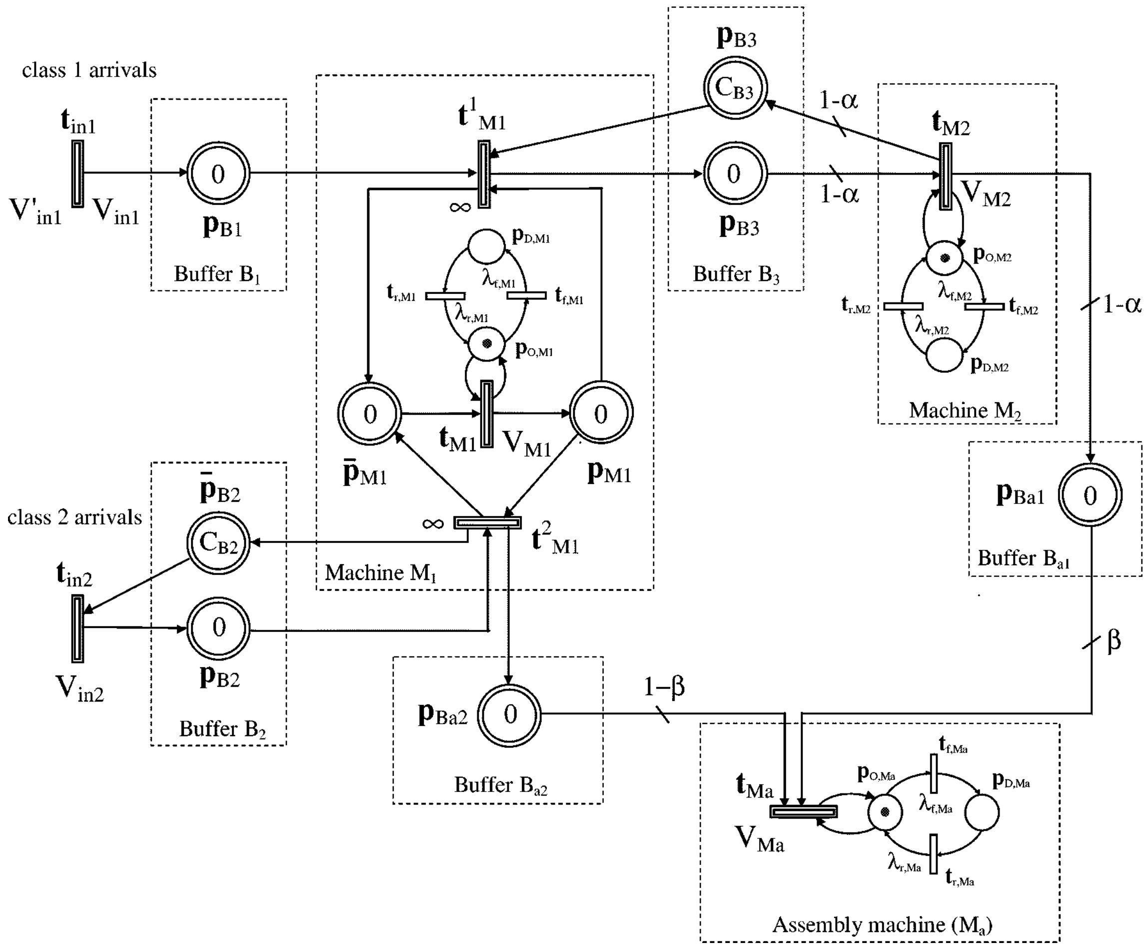

Fig. 8. FOHPN model of the production network in Fig. 7.

while parts of class 1 visit both machines in sequence), the determination of an explicit solution is still an open problem (see Chen *et al.* [12], Wein [23], Phillis and Zhang [24]).

C. Example: PN Model

Let us now model the production network depicted in Fig. 7 by using in a modular way some elementary manufacturing components described by basic FOHPN models.

The FOHPN model of the production system under consideration is shown in Fig. 8, where the initial marking shown assumes that all buffers are initially empty and that the machines are operational. There are a few points we would like to discuss.

- Machine M_1 is a multiclass machine. Transition t^1_{M1} and t^2_{M1} represent the processing of parts of classes 1 and 2, respectively. The overall processing is represented by transition t_{M1}.
- Machines M_2 and M_a are single class machines, represented by a single transition (t_{M2} and t_{Ma}, respectively).
- Buffers of unlimited capacity are represented by continuous places, i.e., place p_{B1} represents the buffer B_1, while finite buffers are represented by a couple of continuous places, i.e., place p_{B2} is the buffer and place $\overline{p}_{B1}$ is the co-buffer.
- The failure model of the machines is represented by two discrete places and two discrete transitions. As an ex-

ample, for machine M_1 place $p_{O,\,M1}$ is marked when the machine is operative, place $p_{D,\,M1}$ is marked when the machine is down. The fail/repare of the machine correspond to the firing of transitions $t_{f,\,M1}$ and $t_{r,\,M1}$.

- The input flow of parts of class 1 is represented by the continuous transition $t_{\text{in}\,1}$ characterized by an mfs $V'_{\text{in}\,1}$ and an MFS $V_{\text{in}\,1}$. This represents the fact that the plant operator may take some control actions on the external flow of parts of class 1 but cannot block the arrival of parts. Although $V'_{\text{in}\,1} > 0$, transition $t_{\text{in}\,1}$ is mfs-free (see Definition 8), thus an admissible IFS always exists for this net.
- The input flow of parts of class 2 is represented by the continuous transition $t_{\text{in}\,2}$. The plant operator may choose to discard parts that arrive, hence this transition has an mfs equal to zero.
- Some parts of class 1 after being processed by machine M_2 may require to be reworked at the same machine according to a given reworking factor α. This fact is represented in the FHOPN model by the weights $1-\alpha$ on arcs $(p_{B3},\, t_{M2})$, $(t_{M2},\, \overline{p}_{B3})$, and $(t_{M2},\, p_{Ba2})$. If we assume that defective parts are scrapped, rather than being reworked, then arcs $(p_{B3},\, t_{M2})$ and $(t_{M2},\, \overline{p}_{B3})$ will have a unitary weight. Note that machine M_2 can be considered as a re-entrant line, and it can be modeled as depicted in Fig. 3(a).
- Parts of both classes are finally assembled by machine M_a according to a given production mix. The mix factors of part of classes 1 and 2 are denoted β and $1-\beta$ and are represented as weights on arcs $(p_{Ba1},\, t_{Ma})$ and $(p_{Ba2},\, t_{Ma})$, respectively.

D. *Numerical Examples*

In this section, we highlight the main steps followed by an FOHPN simulator and show how to solve production control problems and make sensitivity analysis by means of the FOHPN framework.

First of all, we have to define the control problem that we want to solve in terms of a given performance measure that has to be optimized. Then, at the occurrence of the macro-events, a linear programming solver is invoked to provide the optimal machines production rates, i.e., the instantaneous firing speeds of the continuous transitions, according to the constraints defined by the current macro-state. At each step, sensitivity analysis can be done in order to make adjustment on the optimal myopic solution that represents the reference values for the machine production rates within the next macro-state. The marking evolution over several macro-states can be represented by a phase diagram.

In a first example (problem LP1), we assume that the goal is the maximization of the system outflow within a macro-period, i.e., the maximization of the production rate of machine M_a corresponding to the throughput of the network. We show that among all possible optimal solutions it is also possible to choose one (by applying the global priority algorithm) that minimizes the buffer content. We also give examples of sensitivity analysis with respect to machine production rates, the reworking factor α, and the production mix factor β.

In a second example (problem LP2), we choose to maximize machines utilization and study the marking evolution during few macro-periods by constructing the corresponding phase diagram.

The numerical values used in the examples are: $V_{\text{in}\,1} = 5$, $V'_{\text{in}\,1} = 2$, $V_{\text{in}\,2} = 4$, $V_{M1} = 7$, $V_{M2} = 5$, $V_{Ma} = 7$, $\alpha = 0.2$, $\beta = 0.8$. Let us define the instantaneous firing speed vector $\boldsymbol{v} = [v_{\text{in}\,1}\, v_{\text{in}\,2}\, v^1_{M1}\, v^2_{M1}\, v_{M2}\, v_{Ma}]^T$ and let $J = \boldsymbol{c}^T\boldsymbol{v}$ be the performance function to be optimized. For problem LP1, we set $\boldsymbol{c} = [0\ 0\ 0\ 0\ 0\ 1]^T$, and for problem LP2, we set $\boldsymbol{c} = [0\ 0\ 1\ 1\ 1\ 1]^T$. The initial marking $\boldsymbol{m}(\tau_0)$ of the net shown in Fig. 8 represents an initial macro-state in which all machines are operational and all buffers are empty. Such a marking has discrete component

$$\begin{aligned}\boldsymbol{m}^d(\tau_0) = \boldsymbol{m}^d_0 &= [m_{p0,\,M1}\ m_{pD,\,M1}\ m_{p0,\,M2}\\ &\qquad m_{pD,\,M2} m_{p0,\,Ma}\ m_{pD,\,Ma}]^T\\ &= [1\ 0\ 1\ 0\ 1\ 0]^T\end{aligned}$$

and continuous component

$$\begin{aligned}\boldsymbol{m}^c(\tau_0) &= [m_{pB1}\ m_{pB2}\ m_{\overline{p}B2}\ m_{pB3}\ m_{\overline{p}B3}\\ &\qquad m_{pBa1}\ m_{pBa2}\ m_{pM1}\ m_{\overline{p}M1}]^T\\ &= [0\ 0\ C_{B2}\ 0\ C_{B3}\ 0\ 0\ 0\ 0]^T.\end{aligned}$$

To this initial macro-state, we can associate the following set of constraints in standard form (nonnegativity constraints are omitted):

$$\mathcal{S}_0\colon \begin{cases} v_{\text{in}\,1} & +s_1 & = & V_{\text{in}\,1} & .1\\ v_{\text{in}\,1} & -s_2 & = & V'_{\text{in}\,1} & .2\\ v_{\text{in}\,2} & +s_3 & = & V_{\text{in}\,2} & .3\\ v^1_{M1} + v^2_{M1} & +s_4 & = & V_{M1} & .4\\ v_{M2} & +s_5 & = & V_{M2} & .5\\ v_{Ma} & +s_6 & = & V_{Ma} & .6\\ v^1_{M1} - v_{\text{in}\,1} & +s_7 & = & 0 & .7\\ v^2_{M1} - v_{\text{in}\,2} & +s_8 & = & 0 & .8\\ (1-\alpha)v_{M2} - v^1_{M1} & +s_9 & = & 0 & .9\\ \beta v_{Ma} - (1-\alpha)v_{M2} & +s_{10} & = & 0 & .10\\ (1-\beta)v_{Ma} - v^2_{M1} & +s_{11} & = & 0 & .11 \end{cases} \tag{28}$$

1) Problem LP1: Maximization of the System Outflow: The first control problem we consider is the maximization of the system outflow. For the macro-state corresponding to the initial marking shown in Fig. 8, this control problem translates into the following constrained optimization problem:

$$\max_{\boldsymbol{v}}\,[v_{Ma}] \quad \text{s.t.} \quad \boldsymbol{v} \in \mathcal{S}_0 \tag{29}$$

where $\mathcal{S}_0$ is given by (28). The solver provides the following optimal solution: $J^o = 5$ with $\boldsymbol{v}^o = [4\ 3\ 4\ 3\ 5\ 5]^T$. The optimal basis is $\mathcal{B} = \{v_{\text{in}\,1},\ v_{\text{in}\,2},\ v^1_{M1},\ v^2_{M1},\ v_{M2},\ v_{Ma},\ s_1,\ s_2,\ s_3,\ s_6,\ s_{11}\}$.

a) Sensitivity of the Machine Production Rates: To obtain information about the network bottlenecks, we can perform sensitivity analysis with respect to the machine production rates.

For the ith constraint ($i = 1, \cdots, 6$), we compute the *dual prices* J_i^* and the allowable ranges, as described in Section IV-B

$$\Lambda_{\mathcal{B}}[V_{\text{in}\,1}] = [4, \infty), \qquad J_1^* = \frac{\partial J}{\partial V_{\text{in}\,1}} = 0$$

$$\Lambda_{\mathcal{B}}[V'_{\text{in}\,1}] = [0, 4], \qquad J_2^* = \frac{\partial J}{\partial V'_{\text{in}\,1}} = 0$$

$$\Lambda_{\mathcal{B}}[V_{\text{in}\,2}] = [3, \infty), \qquad J_3^* = \frac{\partial J}{\partial V_{\text{in}\,2}} = 0$$

$$\Lambda_{\mathcal{B}}[V_{M1}] = [5, 8], \qquad J_4^* = \frac{\partial J}{\partial V_{M1}} = 0$$

$$\Lambda_{\mathcal{B}}[V_{M2}] = [3.75, 6.25], \qquad J_5^* = \frac{\partial J}{\partial V_{M2}} = 1$$

$$\Lambda_{\mathcal{B}}[V_{Ma}] = [5, \infty), \qquad J_6^* = \frac{\partial J}{\partial V_{M2}} = 0.$$

Consider $\gamma \in \{V_{\text{in}\,1}, V'_{\text{in}\,1}, V_{\text{in}\,2}, V_{M1}, V_{M2}, V_{Ma}\}$. As long as γ varies within its allowable range $\Lambda_{\mathcal{B}}[\gamma]$, the optimal basis of the nominal model remains unchanged. In particular, we observe that the dual price associated to constraint (28.5) is $J_5^* = 1$ and all other dual prices are 0. Thus, in this configuration, machine M_2 represents the bottleneck of the system. If we increase the maximum production rate of machine M_2—i.e., the maximum firing speed V_{M2} of transition t_{M2}—within its allowable range, we can proportionally increase the value of the objective function J. Within this range, the partial derivative of the objective function with respect to V_{M2} is $(\partial J/\partial V_{M2}) = J_5^* = 1$ and the partial derivative of the optimal basic solution with respect to V_{M2} is $(\partial \boldsymbol{v}/\partial V_{M2}) = [0.8\ \ -0.8\ \ 0.8\ \ -0.8\ \ 1\ \ 1]^T$. These two derivatives are constant within the allowable range $\Lambda_{\mathcal{B}}[V_{M2}]$.

b) Sensitivity of the Reworking/Scrapping Factor: We now consider sensitivity analysis with respect to the reworking factor (parameter α). We first observe that if we define $v'_{M2} = (1-\alpha)v_{M2}$, the constraint set (28) can be rewritten, changing constraints 5, 9, and 10 as follows:

$$\begin{cases} v'_{M2} & +s_5 & = & (1-\alpha)V_{M2} & .5\text{a} \\ v'_{M2} - v^1_{M1} & +s_9 & = & 0 & .9\text{a} \\ \beta v_{Ma} - v'_{M2} & +s_{10} & = & 0 & .10\text{a} \end{cases} \quad (30)$$

This shows that the net in Fig. 8 is equivalent to a net with no reworking factor and where transition t_{M2} has MFS $(1-\alpha)V_{M2} = 4$. In this case, the sensitivity analysis with respect to α reduces to the sensitivity analysis with respect to a right-hand side coefficient and can be carried out as previously described.

Let us consider instead a more interesting case. We assume that parts of class 1 are scrapped after failing their processing at machine M_2. In this case, the arcs (p_{B3}, t_{M2}) and $(t_{M2}, \overline{p}_{B3})$ in the net depicted in Fig. 8 will have unitary weight, while the parameter α in the arc (t_{M2}, p_{Ba1}) is now called *scrapping factor*. These changes correspond to rewriting the constraint set (28) changing constraint 9 as

$$\{\, v_{M2} - v^1_{M1} \quad +s_9 \quad = \quad 0 \qquad .9\text{b} \qquad (31)$$

The solver provides the following optimal solution: $J^o = 5$ with $\boldsymbol{v}^o = [5\ \ 2\ \ 5\ \ 2\ \ 5\ \ 5]^T$ and optimal basis $\mathcal{B} = \{v_{\text{in}\,1}, v_{\text{in}\,2}, v^1_{M1}, v^2_{M1}, v_{M2}, v_{Ma}, s_1, s_2, s_3, s_6, s_{11}\}$.

A perturbation of parameter α changes just one element of matrix $\boldsymbol{A}$, namely the element $a = -(1-\alpha)$ in constraint 10, in the column corresponding to variable v_{M2}. Thus, we define $\alpha(\lambda) = \alpha + \lambda$. In this particular case, the parameter v introduced in Section IV-C is $v = 0$. Therefore, the objective function $J(\lambda)$ and the optimal basic solution $\boldsymbol{v}(\lambda)$ vary linearly with the parameter λ as long as it remains within the allowable range

$$\Lambda_{\mathcal{B}}[\lambda] = [-0.32, 0]$$

but for physical reasons we should consider $\Lambda_{\mathcal{B}}(\lambda) = [-0.2, 0]$, since $0 \leq \alpha \leq 1$. Within this range, the derivative of the objective function with respect to λ [by applying (24)] is $(dJ(\lambda)/d\lambda) = -6.25$ and the partial derivative of the optimal basic solution with respect to λ is $(\partial \boldsymbol{v}/\partial\lambda) = [0\ 0\ 0\ 0\ 0\ -6.25]^T$. Thus, we may obtain a better performance by reducing the percentage of scrapped parts that fail their processing at machine M_2.

c) Sensitivity of the Production Mix: We now consider sensitivity analysis with respect to the production mix factor (parameter β). We write $\beta(\lambda) = \beta + \lambda$, and we observe that a perturbation of β changes two elements of matrix $\boldsymbol{A}$ given by (28) (constraints 10 and 11) in the column corresponding to variable v_{Ma}. In this case, the parameter v introduced in Section IV-C is $v = 1.25$, and therefore, the optimal basic solution and the objective function do not vary linearly with the parameter λ. Sensitivity analysis provides the following allowable range for λ:

$$\Lambda_{\mathcal{B}}[\lambda] = [-0.1333, 0]$$

showing that β cannot be increased if the optimal basis has to remain unchanged. Applying (24), we obtain the derivative of the objective function with respect to λ as $(dJ(\lambda)/d\lambda)|_{\lambda=0} = -6.25$. In this case, we may obtain a better performance by reducing the factor β, i.e., by changing the production mix so as to increase the ratio of parts of class 2 with respect to parts of class 1.

As an example, consider $\beta = 0.7$, i.e., $\lambda = -0.1$. By solving LPP (28) for the updated value of β, we obtain $J(\lambda)|_{\lambda=-0.1} = 5.7143$ (while $J^o = J(\lambda)|_{\lambda=0} = 5$) and $\boldsymbol{v}(\lambda)|_{\lambda=-0.1} = [5\ 2\ 5\ 2\ 5\ 5.7143]^T$.

d) Maximum Outflow with Minimal Buffers Content: The control problem defined by (29) admits more than one optimal basic solutions, e.g., $[4\ \ 1\ \ 4\ \ 1\ \ 5\ \ 5]^T$, $[4\ \ 4\ \ 4\ \ 1\ \ 5\ \ 5]^T$ and $[5\ \ 4\ \ 5\ \ 1\ \ 5\ \ 5]^T$ are other optimal basic solutions. Thus, the plant operator may use the global priority algorithm given in Definition 15 to derive a control law in order to minimize, in a second step, the overall buffers content by maximizing the overall buffer outflows

$$J' = [v_{Ma} - v_{\text{in}\,1} - v_{\text{in}\,2}]$$

subject to the constraint set (28) with the additional constraint: $v_{Ma} = 5$. As a solution, we obtain $J'^o = 0$ and $\boldsymbol{v}'^o = [4\ 1\ 4\ 1\ 5\ 5]^T$ that allows all buffers to have their content equal to 0.

2) Problem LP2: Maximization of Machines Utilization: The second problem we consider is the maximization of the machines utilization. We show how to derive an optimal control policy that myopically maximizes the machines utilization, and we describe the developments within the first four macro-periods: $[\mathrm{MP}_0]$ all buffers are empty; $[\mathrm{MP}_1]$ machine M_2 breaks down; $[\mathrm{MP}_2]$ buffer B_3 becomes full; $[\mathrm{MP}_3]$ machine M_2 gets repaired.

$[\mathrm{MP}_0]$*:* The first macro-period, of length Δ_0, starts at time τ_0. The initial marking has discrete and continuous components

$$\begin{aligned}\boldsymbol{m}^d(\tau_0) &= \boldsymbol{m}_0^d = [1\ 0\ 1\ 0\ 1\ 0]^T\\ \boldsymbol{m}^c(\tau_0) &= [0\ 0\ C_{B2}\ 0\ C_{B3}\ 0\ 0\ 0\ 0]^T\end{aligned}$$

i.e., all machines are operational and all buffers are empty. The set of admissible IFS is defined by the constraint set $\mathcal{S}_0$ given by (28). We solve the following constrained linear optimization problem:

$$\max_{\boldsymbol{v}} \left[v_{M1}^1 + v_{M1}^2 + v_{M2} + v_{Ma}\right] \qquad \text{s.t.} \quad \mathcal{S}_0. \qquad (32)$$

The solver provides the following optimal solution: $J_0 = 17$ and $\boldsymbol{v}_0 = [4\ 3\ 4\ 3\ 5\ 5]^T$, which represents an optimal control policy to be adopted during the first macro-period. In particular, throughout the interval Δ_0, B_{a2} is increasing at a rate equal to 2, while the other buffers content are constant and equal to 0. We assume that machine M_2 fails, i.e., transition $t_{f,M2}$ fires, at time τ_1. This macro-event ends the current macro-period.

$[\mathrm{MP}_1]$*:* The second macro-period, of length Δ_1, starts at time τ_1. At the beginning of the macro-period, the marking has discrete and continuous components

$$\begin{aligned}\boldsymbol{m}^d(\tau_1) &= \boldsymbol{m}_1^d = [1\ 0\ 0\ 1\ 1\ 0]^T\\ \boldsymbol{m}^c(\tau_1) &= [0\ 0\ C_{B2}\ 0\ C_{B3}\ 0\ 2\Delta_0\ 0\ 0]^T\end{aligned}$$

i.e., all machines—but M_2—are operational and all buffers—but B_{a2}—are empty. The set of admissible IFS is defined by the new constraint set $\mathcal{S}_1$, that is obtained from $\mathcal{S}_0$ in (28) removing constraint 11 because buffer B_{a2} is not empty and changing constraint 5 to

$$\{\, v_{M2} \quad = \quad 0 \qquad .5'$$

because machine M_2 is down. We solve the following constrained linear optimization problem:

$$\max_{\boldsymbol{v}} \left[v_{M1}^1 + v_{M1}^2 + v_{M2} + v_{Ma}\right] \qquad \text{s.t.} \quad \mathcal{S}_1. \qquad (33)$$

The solver provides the following optimal solution: $J_1 = 7$ and $\boldsymbol{v}_1 = [3\ 4\ 3\ 4\ 0\ 0]^T$. This solution means that the failure of machine M_2 forces machine M_a to produce at a rate $v_{Ma} = 0$, thus increasing the content of buffers B_3 and B_{a2}. In particular, throughout the interval Δ_1, the content of buffer B_{a2} is increasing at a rate equal to 4 and the content of buffer B_3 is increasing at a rate equal to 3. All other buffers are empty. Since buffer B_3 has a finite capacity C_{B3}, it will reach its maximum level after an interval of time

$$\Delta_1 = \frac{C_{B3}}{v_{M1}^1 - (1-\alpha)v_{M2}} = \frac{C_{B3}}{3}.$$

If we assume that no discrete transition fires before, at time $\tau_2 = \tau_1 + \Delta_1$, buffer B_3 will be full and this macro-event ends the current macro-period.

$[\mathrm{MP}_2]$: The third macro-period, of length Δ_2, starts at time τ_2. At the beginning of the macro-period, the marking has discrete and continuous components

$$\begin{aligned}\boldsymbol{m}^d(\tau_2) &= \boldsymbol{m}_2^d = [1\ 0\ 0\ 1\ 1\ 0]^T\\ \boldsymbol{m}^c(\tau_2) &= [0\ 0\ C_{B2}\ C_{B3}\ 0\ 0\ (2\Delta_0 + 4\Delta_1)\ 0\ 0]^T\end{aligned}$$

i.e., all machines—but M_2—are operational, buffer B_3 is full, buffer B_{a2} is not empty, all other buffers are empty. The set of admissible IFS is defined by the new constraint set $\mathcal{S}_2$, that is obtained from $\mathcal{S}_0$ in (28) removing constraint 11 because buffer B_{a2} is not empty, and changing constraints 5 and 9 to:

$$\begin{cases} v_{M2} & = \quad 0 \qquad .5''\\ (1-\alpha)v_{M2} - v_{M1}^1 \quad -s9 & = \quad 0 \qquad .9''\end{cases}$$

because machine M_2 is down and buffer B_3 is full (i.e., place $\overline{p}_{B3}$ is empty). We solve the following constrained linear optimization problem:

$$\max_{\boldsymbol{v}} \left[v_{M1}^1 + v_{M1}^2 + v_{M2} + v_{Ma}\right] \qquad \text{s.t.} \quad \mathcal{S}_2. \qquad (34)$$

The solver provides the following optimal solution: $J_2 = 4$, and $\boldsymbol{v}_2 = [2\ 4\ 0\ 4\ 0\ 0]^T$. Throughout the interval Δ_2, the content of buffer B_1 is increasing at a rate equal to 2 and the content of buffer B_{a2} is increasing at a rate equal to 4. Buffer B_3 is full, while all other buffers are empty. We assume that machine M_2 is repaired, i.e., transition $t_{f,M2}$ fires, at time τ_3. This macro-event ends the current macro-period.

$[\mathrm{MP}_3]$: The fourth macro-period, of length Δ_3, starts at time τ_3. At the beginning of the macro-period, the marking has discrete and continuous components

$$\begin{aligned}\boldsymbol{m}^d(\tau_3) &= \boldsymbol{m}_3^d = [1\ 0\ 1\ 0\ 1\ 0]^T\\ \boldsymbol{m}^c(\tau_3) &= [2\Delta_2\ 0\ C_{B2}\ C_{B3}\ 0\ 0\ (2\Delta_0 + 4\Delta_1 + 4\Delta_2)\ 0\ 0]^T\end{aligned}$$

i.e., all machines are operational, buffer B_3 is full, buffers B_1 and B_{a2} are not empty, all other buffers are empty. The set of admissible IFS is defined by the new constraint set $\mathcal{S}_3$, that is obtained from $\mathcal{S}_0$ in (28) removing constraints 7 and 11 because buffers B_1 and B_{a2} are not empty, and changing constraint 9 to

$$\{\,(1-\alpha)v_{M2} - v_{M1}^1 \quad -s9 \quad = \quad 0 \qquad .9'''$$

because buffer B_3 is full. We solve the following constrained linear optimization problem:

$$\max_{\boldsymbol{v}} \left[v_{M1}^1 + v_{M1}^2 + v_{M2} + v_{Ma}\right] \qquad \text{s.t.} \quad \mathcal{S}_3. \qquad (35)$$

The solver provides the following optimal solution: $J_3 = 17$ and $\boldsymbol{v}_3(\tau_3) = [2\ 3\ 4\ 3\ 5\ 5]^T$. Throughout the interval Δ_3, the content of buffer B_1 is decreasing at a rate equal to 2 and the content of buffer B_{a2} is increasing at a rate equal to 2. Buffer B_3 is full, and all other buffers are empty.

Macro-Behavior and Phase Diagram: In the previous evolution, the myopic optimal control policy

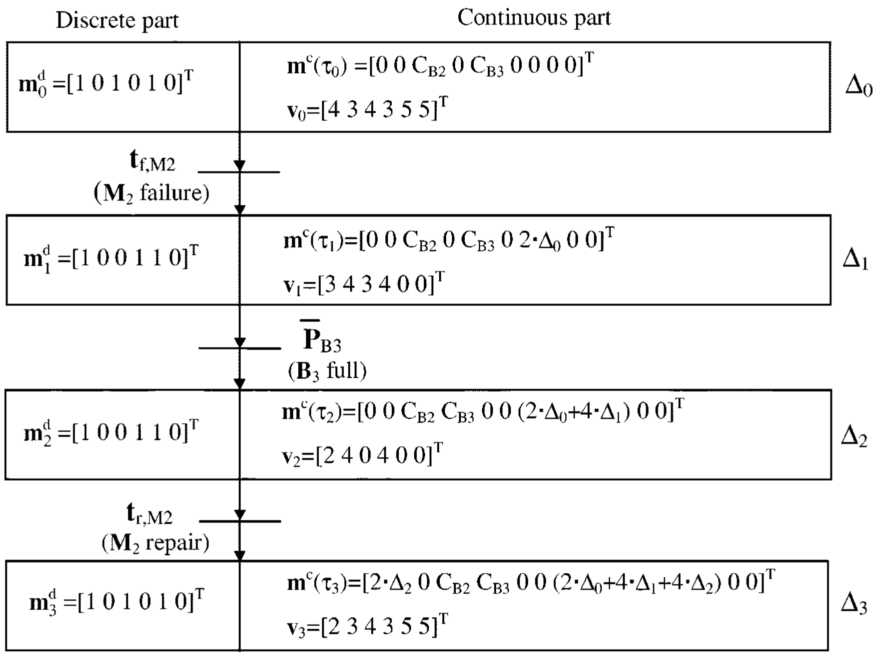

Fig. 9. Phase diagram of the FOHPN in Fig. 8.

$\mathcal{V}^o = \{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots\}$ that allows the maximization of the machines utilization is defined as follows:

$$\begin{array}{llll} J_0 = 17, & \boldsymbol{v}_0 = [4\ 3\ 4\ 3\ 5\ 5]^T & \text{throughout} & \Delta_0 \\ J_1 = 7, & \boldsymbol{v}_1 = [3\ 4\ 3\ 4\ 0\ 0]^T & \text{throughout} & \Delta_1 \\ J_2 = 4, & \boldsymbol{v}_2 = [2\ 4\ 0\ 4\ 0\ 0]^T & \text{throughout} & \Delta_2 \\ J_3 = 17, & \boldsymbol{v}_3 = [2\ 3\ 4\ 3\ 5\ 5]^T & \text{throughout} & \Delta_3 \\ \quad\cdots & & & \end{array}$$

The developments discussed so far can be graphically shown, as in Fig. 9, by means of the phase diagram of the net with regard to the first four macro-periods.

VI. Conclusions

We have considered in this paper FOHPN's, and we have set up a linear algebraic formalism to study the first-order continuous behavior of this model, thus showing how its control can be framed as a conflict resolution policy that aims to optimize a given objective function. Assuming that the instantaneous firing speeds of continuous transitions are piecewise constant, we have shown that the set of all possible behaviors of the net during a macro-state can be represented by a convex set defined by linear inequalities. The computation of the instantaneous firing speed—and the associated problem of conflict resolution—can be seen as the net counterpart of a performance optimization with global or local objective functions.

Sensitivity analysis techniques have been also proposed in this paper to obtain information about the degrees of freedom that can be exploited when making performance optimization or optimal design of the system parameters configuration. Finally, we have discussed in depth a case study of a realistic manufacturing system with three machines and five buffers. In this system, parts may fail their processing and thus may have to be sent back for reworking or may be scrapped, and we have introduced a parameter to take into account the possibility of manufacturing different finished products according to an arbitrary production mix. We have shown examples of how different control policies may be enforced by different objective functions, of sensitivity analysis with respect to different design parameters (machine production rates, reworking or scrapping factor, production mix factor), and of evolution over more than one macro-period.

Acknowledgment

The authors would like to thank F. Della Croce and C. Seatsu for their useful comments and valuable discussions.

References

[1] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis, *Modeling with Generalized Stochastic Petri Nets*, ser. in Parallel Computing. New York: Wiley, 1995.

[2] H. Alla and R. David, "Continuous and hybrid Petri nets," *J. Circuits, Syst., Comput.*, vol. 8, no. 1, pp. 159–188, 1998.

[3] ——, "A modeling and analysis tool for discrete event systems: Continuous Petri net," *Perform. Eval.*, vol. 33, pp. 175–199, 1998.

[4] A. Amrah, N. Zerhouni, and A. El Moudni, "On the control of manufacturing lines modeled by controlled continuous Petri nets," *Int. J. Syst. Sci.*, vol. 29, no. 2, pp. 127–137, 1998.

[5] F. Balduzzi and G. Menga, "A state variable model for the fluid approximation of flexible manufacturing systems," in *Proc. 1998 IEEE Int. Conf. on Robotics and Automation*, Leuven, Belgium, May 1998, pp. 1172–1178.

[6] F. Balduzzi, A. Giua, and G. Menga, "Hybrid stochastic Petri nets: Firing speed computation and FMS modeling," in *Proc. 4th Workshop on Discrete Event Systems (WODES'98)*, Cagliari, Italy, Aug. 1998, pp. 432–438.

[7] ——, "Optimal speed allocation and sensitivity analysis of hybrid stochastic Petri nets," in *Proc. 1998 IEEE Conf. on Systems, Man and Cybernetics*, San Diego, CA, Oct. 1998, pp. 656–662.

[8] F. Balduzzi, A. Giua, G. Menga, and C. Seatsu, "A linear state variable model for first-order hybrid Petri nets," in *Proc. 14th IFAC World Congress* Bejing, China, July 1999, pp. 205–210.

[9] R. E. Burkard and F. Rendl, "Lexicographic bottleneck problems," *Oper. Res. Lett.*, vol. 10, pp. 303–308, 1991.

[10] R. Champagnat, Ph. Esteban, H. Pingaud, and R. Valette, "Modeling and simulation of a hybrid system through Pr/Tr PN-DAE model," in *Proc. 3rd Int. Conf. on the Automation of Mixed Processes*, Reims, France, Mar. 1998, pp. 131–137.

[11] H. Chen and A. Mandelbaum, "Discrete flow networks: Bottleneck analysis and fluid approximations," *Math. Oper. Res.*, vol. 16, pp. 408–446, 1991.

[12] H. Chen, P. Yang, and D. D. Yao, "Control and scheduling in a two-station queuing network: Optimal policies and heuristics," *Queueing Syst.*, vol. 18, pp. 301–332, 1994.

[13] R. David and H. Alla, "Réseaux de Petri hybrides," in Modélization et Commande des Systèmes Dynamiques Hybrids, Hermés, Ed. Paris, France, to be published.

[14] I. Demongodin and N. T. Koussoulas, "Differential Petri nets: Representing continuous systems in a discrete-event world," *IEEE Trans. Automat. Contr.*, vol. 43, no. 4, pp. 573–579, 1998.

[15] T. Gal, *Postoptimal Analysis, Parametric Programming and Related Topics*. New York: McGraw-Hill, 1979.

[16] A. Giua and E. Usai, "High-level hybrid Petri nets: A definition," in *Proc. 35th Conf. on Decision and Control*, Kobe, Japan, Dec. 1996, pp. 148–150.

[17] T. Murata, "Petri nets: Properties, analysis and applications," *Proc. IEEE*, vol. 77, pp. 541–580, Apr. 1989.

[18] K. G. Murty, *Linear and Combinatorial Programming*. New York: Wiley, 1976.

[19] J. Renegar, "Some perturbation theory for linear programming," *Math. Programming*, pp. 73–91, 1991.

[20] R. Suri and B. R. Fu, "Using continuous flow models to enable rapid analysis and optimization of discrete production lines—A progress report," in *Proc. 19th Annu. NSF Grantees Conf. on Design and Manufacturing Systems Research*, 1993.

[21] ——, "On using continuous flow lines to model discrete production lines," *J. Discrete Event Dynamic Syst.*, vol. 4, pp. 129–169, 1994.

[22] K. S. Trivedi and V. G. Kulkarni, "FSPNs: Fluid stochastic Petri nets," in *Proc. 14th Int. Conf. on Applications and Theory of Petri Nets*. ser. Lecture Notes in Computer Science, M. Ajmone Marsan, Ed. Heidelberg: Springer-Verlag, 1993, vol. 691, pp. 24–31.

[23] L. M. Wein, "Optimal control of a two-station Brownian network," *Math. Oper. Res.*, vol. 15, pp. 215–242, 1990.

[24] R. Zhang and Y. A. Phillis, "Multiple control policies of two-station production network with two types of parts using fuzzy logic," in *Proc. 1998 IEEE Int. Conf. on Robotics and Automation*, Leuven, Belgium, May 1998, pp. 2759–2764.

Fabio Balduzzi (M'98) received the Laurea degree (*cum laude*) in electronic engineering and the Ph.D. degree in computer and systems engineering from Politecnico di Torino, Torino, Italy, in 1994 and 1999, respectively.

He is a Research Assistant with the Department of Control and Computer Science at Politecnico di Torino. His research interests include the field of discrete event and hybrid systems theory and application, especially in the area of control of flexible manufacturing systems.

Dr. Balduzzi is co-author of one of the six Best Conference Papers at the 1999 IEEE International Conference on Robotics and Automation. He was member of the ETFA'99 International Program Committee.

Alessandro Giua received the Laurea degree in electric engineering from the University of Cagliari, Italy, in 1988, and the M.S. and Ph.D. degrees in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1990 and 1992.

He was a Visiting Researcher at the University of Zaragoza (Spain) in 1992, and at INRIA Roquencourt (France) in 1995. Since 1993, he has been with the Department of Electrical and Electronic Engineering of the University of Cagliari, where he is now an Associate Professor of Automatic Control. His current research interests include control engineering, discrete event systems, hybrid systems, automated manufacturing, and Petri nets.

Prof. Giua is a Member of the Editorial Board of the *Journal of Discrete Event Dynamic Systems* and is an Associate Editor of the *European Journal of Control*. He is member of the WODES Steering Committee (workshop series on discrete event systems).

Giuseppe Menga received the Laurea degree in electronic engineering from Politecnico di Torino, Turin, Italy, in 1967. He was a Fellowship Researcher at the University of Colorado in 1973, and a Senior Postdoctoral Associate at the NASA AMES Research Center from 1974 to 1975.

From 1970 to 1980, he was an Associate Professor at Politecnico di Torino, and since 1980, he has been Full Professor of Automatic Control at Politecnico di Torino. He is author of more than 100 international scientific papers or book chapters in the fields of automatic control, automation, robotics and software engineering. He has also published an academic book titled, *Theory of Discrete Event Systems* UTET 1998.

At the national level, Dr. Menga has been Coordinator of one of the research lines of the first special project on Information of the Italian National Research Council (CNR), he has also been President of the local chapter and National Vice President of the Italian professional association on Automation (ANIPLA) in the 1980's. At the international level, he has been member of the Advisory Committee and Technical Editor of the IEEE Robotics and Automation Society; for the same society, he served as General Chairman of the International IEEE Robotics and Automation Conference held, the first time in its history outside the US, in Nice, France, 1992. As professional activities, in 1978, he founded SYCO, a consulting company in the fields of robotics, automation, and software engineering.