Optimal Speed Allocation and Sensitivity Analysis of Hybrid Stochastic Petri Nets*

†Fabio Balduzzi, ‡Alessandro Giua, †Giuseppe Menga

† Dipartimento di Automatica ed Informatica, Politecnico di Torino

Corso Duca degli Abruzzi 24 , 10129 Torino, Italy

(balduzzi, menga)@polito.it

‡ Dipartimento di Ingegneria Elettrica ed Elettronica, Università di Cagliari

Piazza d'Armi, 09123 Cagliari

giua@diee.unica.it

ABSTRACT

In this paper we present a method for performance evaluation of Hybrid Stochastic Petri Nets based on sensitivity analysis and parametric linear programming techniques and we show how this approach can be used for optimization. The problem is addressed by determining an optimal firing speed allocation for the continuous transitions obtained by solving a sequence of linear programming problems aimed at optimizing a certain performance index. The primary advantage of our approach is that it gives rise to a dynamic firing speeds allocation that is based on global state information rather than local information. This original formulation allows us to easily solve conflicts, evaluate performance measures and perform gradient estimation very efficiently.

1 INTRODUCTION

We consider in this paper *Hybrid Stochastic Petri Nets* (HSPN), a model in which places and transitions may be either continuous or discrete. This model, presented in [4], combines the hybrid framework proposed by Alla and David [2] with the generalized stochastic Petri nets of Ajmone et al. [1]. The main differences with a similar model presented by Trivedi and Kulkarni [6] concern the maximum firing speed (MFS) of continuous transitions that we assume to be constant.

In the HSPN framework a net consists of continuous places holding fluid, discrete places containing a non–negative integer number of tokens, and transitions, either discrete or continuous. Enabled continuous and discrete transitions may fire according to their firing speeds or time delays, respectively.

*Published as: F. Balduzzi, G. Menga, A. Giua, "Optimal Speeds Allocation and Sensitivity Analysis for Hybrid Stochastic Petri Nets," *Proc. 1998 IEEE Int. Conf. on Systems, Man, and Cybernetics* (San Diego, California), pp. 656-662, October, 1998.

In a previous work [4] the authors have shown that HSPNs are well suited for modelling automated manufacturing systems characterized by unreliable machines, buffers of finite capacity and general service time distributions and routing policies, where the continuous transitions model the production of the machines. Continuous firing of these transitions corresponds to a continuous production at rates determined by the current values of their instantaneous firing speeds (IFS). In [4] the focus was on conflict resolution policies, i.e. on the computation of IFSs, seen as the decisions that a plant operator must take in order to optimize the process. This can be done solving a linear programming problem (LPP) of the form $\max_{\mathbf{v}} \{\mathbf{c}^T\mathbf{v} \mid \mathbf{A}\mathbf{v} \leq \mathbf{b},\ \mathbf{v} \geq 0\}$ where:

- the set of admissible IFS vectors $\mathbf{v}$ can be characterized by the feasible solutions of a linear constraint set $\mathcal{S} = \{\mathbf{v} \mid \mathbf{A}\mathbf{v} \leq \mathbf{b}\}$. The MFSs of continuous transitions appear in the right–hand side vector $\mathbf{b}$. In the coefficients matrix $\mathbf{A}$ appear the arc weights of continuous transitions, e.g. the fluid routing coefficients, and (eventually) the fixed ratios that may be imposed among IFSs.
- the different objective functions $J = \mathbf{c}^T\mathbf{v}$ to be maximized can be associated to different conflict resolution policies.

The constraint set $\mathcal{S}$ is a function of the current marking m of the net, because it is characterized by the marking of the discrete places and by the set of non–empty continuous places, i.e. it is characterized by the *macro–state* of the net. This formulation leads to a myopic procedure which generates a piecewise optimal control policy during each time interval in which the macro–state remains constant. As the system evolves through a sequence of macro–states upon the occurrence of the *macro–events*, the myopic procedure will be called repeatedly.

In this paper we further exploit the linear algebraic formalism underlying this model, to show that we can naturally apply in the HSPN framework those sensitivity analysis techniques that pertain to LPPs. The optimal basis approach, i.e. the simplex method, is adopted in this paper to solve LPPs.

Sensitivity analysis or *postoptimal analysis* serves as a tool for obtaining information about the degrees of freedom in the problem. Specifically it 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, see for instance [9] and [5]. However even though these parameters may change, when the perturbations are within a certain range, the current set of basic variables may remain unchanged. This invariance of the set of basic variables is a desirable property because it allows one to compute the gradient of the objective function with respect to these parameters. The maximum range of individual perturbation is called the *allowable range*.

The main motivation of this paper is to provide a tool for sensitivity analysis of HSPNs. In fact in optimizing a performance of an HSPN one needs to compute the sensitivity of the performance. This is achieved by exploiting some results of parametric linear programming techniques to make gradient evaluation of the optimal firing speeds of the continuous transition. The proposed formulation can also be applied to real systems, i.e. manufacturing systems, to obtain the sensitivity without changing the values of the parameters.

2 DEFINITION OF HSPNs

We recall the Petri net formalism used in this paper. For a more comprehensive introduction to place/transition Petri nets see [7], while the common notation and semantics for GSPNs can be found in [1]. The first approach towards continuous Petri nets was carried out by Alla and David and then extended to hybrid nets in [2]. The HSPN model we use follows [4].

An HSPN is a structure $N = (P, T, Pre, Post, \mathcal{F})$. 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).

$$Pre : \begin{cases} P_d \times T \to \mathbb{N} \\ P_c \times T \to \mathbb{R}^+ \cup \{0\} \end{cases}$$

and

$$Post : \begin{cases} P_d \times T \to \mathbb{N} \\ P_c \times T \to \mathbb{R}^+ \cup \{0\} \end{cases}$$

are the *pre-* and *post-incidence functions* that specify the arcs. We require (*well-formed nets*) that for all $t \in T_c$ and for all $p \in P_d$, $Pre(p,t) = Post(p,t)$. The function $\mathcal{F}$ is defined for continuous and discrete timed transitions so that $\mathcal{F} : T \setminus T_I \to \mathbb{R}^+$. We associate to a continuous transition $t_i \in T_c$ its *maximum firing speed* (MFS) $V_i = \mathcal{F}(t_i)$. We associate to a deterministic timed transition $t_i \in T_D$ its (constant) firing delay $\delta_i = \mathcal{F}(t_i)$. We associate to an exponentially distributed timed transition $t_i \in T_E$ its average firing rate $\lambda_i = \mathcal{F}(t_i)$, i.e. the average firing delay is $\frac{1}{\lambda_i}$, where λ_i is the parameter of the corresponding exponential distribution.

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 $\mathrm{C}(p,t) = Post(p,t) - Pre(p,t)$. The restriction of C to P_X and T_Y ($X, Y \in \{c, d\}$) is denoted C_{XY}. Note that by the well-formedness hypothesis $C_{dc} = 0$.

A *marking*

$$m : \begin{cases} P_d \to \mathbb{N} \\ P_c \to \mathbb{R}^+ \cup \{0\} \end{cases}$$

is a function that assigns to each discrete place a non-negative number of tokens, represented by black dots and assigns to each continuous place a fluid volume; m_p denotes the marking of place p. A discrete transition t is enabled at m if for all $p \in {}^\bullet t$, $m_p \geq Pre(p,t)$. An enabled discrete transition t fires (after the associated delay) yielding the marking $m' = m + C(\cdot, t)$.

A continuous transition t is enabled at m if for all $p \in {}^{(d)}t$, $m_p \geq Pre(p,t)$. Note that the enabling of a continuous transition does not depend on the marking of its continuous input places. We distinguish *strongly enabled* and *weakly enabled* continuous transitions. A transition $t_i \in T_c$ is strongly enabled at $m(\tau)$ if for all places $p \in {}^{(c)}t$, $m_p(\tau) > 0$. Then it may fire with an *instantaneous firing speed* (IFS) $v_i(\tau) = V_i$. A transition $t_i \in T_c$ is weakly enabled at $m(\tau)$

Figure 1: The HSPN model of a service (buffer and machine).

at time τ if for some $\bar{p} \in {}^{(c)}t$, $m_{\bar{p}}(\tau) = 0$. Thus its IFS may result $v_i(\tau) < V_i$ because it cannot remove more fluid from place $\bar{p}$ than the quantity entered in $\bar{p}$ by other transitions. Moreover if $t_i \in T_c$ is not enabled at m_p at time τ then $v_i(\tau) = 0$.

We can now define the macro–behavior of a net. A *macro–event* occurs when: (a) either a discrete transition fires, thus changing the discrete marking and enabling/disabling a continuous transition; (b) 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 in time of consecutive macro–events; the interval of time $\Delta_k = [\tau_k, \tau_{k+1})$ is called a *macro–period.* We will assume that the IFS of continuous transitions are piecewise constant during a macro–period. Thus the discrete marking and the IFS vector during a macro–period define a *macro–state* that correspond to the *invariant behavior states* of [2].

Let $v_i(\tau)$ be the IFS of each transition $t_i \in T_c$. We can write the equation which governs the evolution in time of the marking of a place $p \in P_c$ as

$$\frac{\mathrm{d}m_p}{\mathrm{d}\tau} = \sum_{t_i \in T_c} \mathrm{C}(p, t_i) \cdot v_i(\tau) \tag{1}$$

Indeed Equation (1) holds assuming that at time τ no discrete transition is fired and that all speeds $v_i(\tau)$ are continuous in τ. The evolution in time of the marking of a place $p \in P_d$ is governed by the common enabling and firing rules defined in [1].

Example 1. In Figure 1 we have represented the HSPN model of a manufacturing system 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. ■

3 FIRING SPEEDS AND DYNAMICS

The computation of an admissible IFS vector of continuous and hybrid nets is not trivial. Our approach makes use of linear inequalities to define the set of all admissible firing speed vectors $\mathcal{S}$. Each set $\mathcal{S}$ corresponds to a particular system macro-state, hence our optimization scheme can only be *myopic* [3], in the sense that it generates a piecewise optimal solution.

Definition 2 (admissible IFS vectors)**.** Let N be an HSPN, with n_c continuous transitions, incidence matrix $\mathbf{C}$, and current marking $m(\tau)$. Let $T_{\mathcal{E}}(m) \subset T_c$ ($T_{\mathcal{N}}(m) \subset T_c$) be the subset of continuous transitions enabled (not enabled) at $m(\tau)$, while $P_{\mathcal{E}} = \{p \in P_c \mid m_p = 0\}$ is the subset of continuous places that are empty. Any *admissible IFS vector* $\mathbf{v}(\tau) = [v_1, \cdots, v_{n_c}]^T$ is a feasible solution of the following linear set:

$$\begin{cases} (a) \quad V_j - v_j(\tau) \geq 0 & \forall t_j \in T_{\mathcal{E}}(m) \\ (b) \quad v_j(\tau) \geq 0 & \forall t_j \in T_{\mathcal{E}}(m) \\ (c) \quad v_j(\tau) = 0 & \forall t_j \in T_{\mathcal{N}}(m) \\ (d) \quad \sum_{t_j \in T_{\mathcal{E}}} \mathbf{C}(p, t_j) \cdot v_j(\tau) \geq 0 & \forall p \in P_{\mathcal{E}}(m) \end{cases} \tag{2}$$

Thus the total number of constraints that define this set is $2\text{card}\,\{T_{\mathcal{E}}(m)\} + \text{card}\,\{T_{\mathcal{N}}(m)\} + \text{card}\,\{P_{\mathcal{E}}(m)\}$. The set of all feasible solutions is denoted $\mathcal{S}(N, m)$. ■

Constraints of the form (2.a), (2.b), and (2.c) follow from the enabling rules. Constraints of the form (2.d) follow from (1), because if a place is empty its fluid content cannot decrease. Additional constraints may be added to the linear set (2) to require a fixed ratio among IFSs. As an example, constraints of the form $v_i = s v_j$ assign a fixed ratio between the IFSs of transitions t_i and t_j.

Each vector $\mathbf{v} \in \mathcal{S}$ 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 one according to a given objective. We have considered in [4] linear objective functions of the form $J = \mathbf{c}^T \mathbf{v}$ to be maximized. By a suitable choice of the cost vector $\mathbf{c}$ it is possible to: maximize the sum over all flow rates, maximize the throughput of a given set of transitions, assign global priorities to the transition firings.

Example 3. The constraint set associated to the net shown in Figure 1 from the given marking is:

$$\begin{cases} v_1 & \leq 5 \\ v_2 & \leq 5 \\ v_3 & \leq 4 \\ -v_1 + v_2 + v_3 & \leq 0 \\ v_1 - v_2 - v_3 & \leq 0 \\ v_1, v_2, v_3 & \geq 0 \end{cases} \tag{3}$$

We take as objective function to be maximized $J = v_2 + v_3$, representing the overall output flow. ■

4 SENSITIVITY ANALYSIS OF HSPNs

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 IFSs of transitions that are not enabled are 0. Let $I_t = \{\alpha_1, \ldots, \alpha_k\}$ be the set of indices of the enabled continuous transitions and

$I_p = \{\alpha_{k+1}, \dots, \alpha_\ell\}$ be the set of indices of the empty continuous places. Thus we can write:

$$\begin{array}{l} \max \sum_{j \in I_t} c_j \; v_j \qquad \text{s.t.} \\ \left\{ \begin{array}{lcl} v_{\alpha_1} + s_1 & = & V_{\alpha_1} \\ \dots & & \\ v_{\alpha_k} + s_k & = & V_{\alpha_k} \\ \sum_{j \in I_t} \mathbf{C}(p_{\alpha_{k+1}}, t_j) \; v_j - s_{k+1} & = & 0 \\ \dots & & \\ \sum_{j \in I_t} \mathbf{C}(p_{\alpha_\ell}, t_j) \; v_j - s_\ell & = & 0 \end{array} \right. \\ v_j, s_j \geq 0 \end{array} \tag{4}$$

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

$$\max_{\mathbf{x}} \{\mathbf{c}^T \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}, \; \mathbf{x} \geq 0\} \tag{5}$$

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

In this work the *simplex method* will be used to solve LPPs. 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 basic solution $\mathbf{x}^o$, the corresponding optimal basis $\mathcal{B}$ (a set of ℓ indices), and $\mathbf{A}_\mathcal{B}$ the optimal basis matrix obtained by taking only those columns of $\mathbf{A}$ whose indices are in $\mathcal{B}$. An optimal basic solution $\mathbf{x}^o$ can always be written as:

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

The variables with index in $\mathcal{B}$ are the basic variables while the others, whose index 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 4. For the net described in Example 3 we consider

$$\begin{array}{l} \max v_2 + v_3 \qquad \text{s.t.} \\ \left\{ \begin{array}{ll} v_1 + s_1 & = 5 \\ v_2 + s_2 & = 5 \\ v_3 + s_3 & = 4 \\ v_1 - v_2 - v_3 & = 0 \end{array} \right. \end{array}$$

and define $\mathbf{x} = [v_1, v_2, v_3, s_1, s_2, s_3]^T$. Note that we have packed together the last two inequalities of (3). 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 Figure 2 in the plane $v_1 = 5$. Two of these are basic solutions: $\mathbf{v}_{(A)} = [5, 1, 4]^T$ and $\mathbf{v}_{(B)} = [5, 5, 0]^T$. Point (A) is a non–degenerate solution with basic variables v_1, v_2, v_3, s_2 and basis $\mathcal{B}_A = \{1, 2, 3, 5\}$. Point (B) is a degenerate solution with two optimal basis: $\mathcal{B}_{B1} = \{1, 2, 3, 6\}$, with basic variables v_1, v_2, v_3, s_3, and $\mathcal{B}_{B2} = \{1, 2, 5, 6\}$, with basic variables v_1, v_2, s_2, s_3. Furthermore we observe that in (B) there is also another basis, with basic variables v_1, v_2, s_1, s_3, which is not optimal. ■

Figure 2: Feasible region for the net considered in Example 1.

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 (5) 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.

The perturbed model

The following perturbed LPP is treated:

$$\max_{\mathbf{x}} \{\mathbf{c}^T\mathbf{x} \mid \mathbf{A}(\mathbf{q})\mathbf{x} = \mathbf{b}(\mathbf{q}),\ \mathbf{x} \geq 0\} \tag{6}$$

where $\mathbf{q} = [q_0, \ldots, q_p]^T$ is a vector of uncertain parameters. The nominal value is denoted $\overline{\mathbf{q}}$. For a given value of $\mathbf{q}$, the optimal solution of (6) is

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

We compute with the simplex method an optimal solution in $\overline{\mathbf{q}}$ and the corresponding optimal basis $\mathcal{B}$. The sensitivity of the basic variables $\mathbf{x}_{\mathcal{B}}(\overline{\mathbf{q}})$ 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

$$\frac{\partial \mathbf{x}_{\mathcal{B}}(\overline{\mathbf{q}})}{\partial q_i} = \mathbf{A}_{\mathcal{B}}^{-1}(\overline{\mathbf{q}}) \left(\frac{\partial \mathbf{b}(\overline{\mathbf{q}})}{\partial q_i} - \frac{\partial \mathbf{A}_{\mathcal{B}}(\overline{\mathbf{q}})}{\partial q_i} \mathbf{x}_{\mathcal{B}}(\overline{\mathbf{q}}) \right) \tag{7}$$

while the non–basic variables $\mathbf{x}_{\mathcal{N}}(\overline{\mathbf{q}})$ do not change. It is only required first order differentiability of $\mathbf{A}_{\mathcal{B}}^{-1}(\overline{\mathbf{q}})$ and $\mathbf{b}(\overline{\mathbf{q}})$ with respect to q_i. For simplicity in this presentation we make the following assumptions:

1. Only one parameter q_i varies at a time, that is $\mathbf{q} = \overline{\mathbf{q}} + \lambda \mathbf{e}_i$, where $\mathbf{e}_i$ is the i–th canonical basis vector. Under this assumption the sensitivity given by (7) can be regarded as function of λ in the allowable range.

2. Matrix $\mathbf{A}$ and vector $\mathbf{b}$ are linear functions of the parameter λ. Then we can write:

$$\mathbf{A}_{\mathcal{B}}(\lambda) = \mathbf{A}_{\mathcal{B}} + \lambda \mathbf{A}_{\mathcal{B}}^*$$
$$\mathbf{b}(\lambda) = \mathbf{b} + \lambda \mathbf{b}^*$$

where $\mathbf{A}_{\mathcal{B}} = \mathbf{A}_{\mathcal{B}}(\overline{\mathbf{q}})$, $\mathbf{b} = \mathbf{b}(\overline{\mathbf{q}})$.

3. The variation of each parameter q_i influences only one column, say the j-th, of matrix $\mathbf{A}_{\mathcal{B}}(\lambda)$. Then
$$\mathbf{A}_{\mathcal{B}}(\lambda) = \mathbf{A}_{\mathcal{B}} + \lambda \mathbf{A}^*_{\mathcal{B}} = \mathbf{A}_{\mathcal{B}} + \lambda \mathbf{a}^* \mathbf{e}_j^T$$

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

Perturbation of the right–hand side vector

We assume that the right–hand side constant vector $\mathbf{b}$ varies linearly with the parameter $\lambda \in \mathbb{R}$, that is $\mathbf{b}(\lambda) = \mathbf{b} + \lambda \mathbf{b}^*$. In the HSPN framework, this perturbation corresponds to changes in the entries of the vector $\mathbf{V} = [V_{\alpha_1}, \ldots, V_{\alpha_k}]^T$, 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 $\mathbf{b}^* = \mathbf{e}_i$ for $i = 1, \ldots, k$. 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 $\mathbf{b}^* = \mathbf{e}_i - \mathbf{e}_j$.

Let $\mathbf{x}^o$ be an optimal basic solution of (5) and $\mathcal{B}$ an associated optimal basis. The perturbed optimal solution $\mathbf{x}^o(\lambda)$ has basic components:

$$\mathbf{x}^o_{\mathcal{B}}(\lambda) = \mathbf{A}_{\mathcal{B}}^{-1} \mathbf{b}(\lambda) = \mathbf{A}_{\mathcal{B}}^{-1}(\mathbf{b} + \mathbf{b}^* \lambda) = \mathbf{x}^o_{\mathcal{B}} + \lambda \mathbf{x}^*_{\mathcal{B}} \tag{8}$$

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

$$J(\lambda) = \mathbf{c}_{\mathcal{B}}^T \mathbf{x}^o_{\mathcal{B}}(\lambda) = \mathbf{c}_{\mathcal{B}}^T \mathbf{x}^o_{\mathcal{B}} + \lambda \mathbf{c}_{\mathcal{B}}^T \Delta \mathbf{x}_{\mathcal{B}} = J + \lambda J^*. \tag{9}$$

Equations (8) and (9) 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 non–negativity of the basic variables, $\mathbf{x}^o_{\mathcal{B}}(\lambda) \geq \mathbf{0}$, and the bounds for the parameter λ can be computed as follows:

$$\underline{\lambda}_{\mathcal{B}} = \begin{cases} -\infty & \text{if} \quad I^+ = \emptyset \\ \max_{i \in I^+} \left\{ -\frac{\beta_i}{\beta_i^*} \right\} \end{cases} \tag{10}$$

and

$$\overline{\lambda}_{\mathcal{B}} = \begin{cases} +\infty & \text{if} \quad I^- = \emptyset \\ \min_{i \in I^-} \left\{ -\frac{\beta_i}{\beta_i^*} \right\} \end{cases} \tag{11}$$

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

Much attention has been devoted in the literature [8], [5] to the case in which the optimal solution $\mathbf{x}^o$ of the nominal LPP is unique. In this case $\mathbf{x}^o$ 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 point. As the basis changes, the gradient $\Delta J = J^*$ of the objective function 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 [10] 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 5. Let us consider again the net in Example 4. There are two optimal basic solutions, (A) and (B), and three optimal basis. 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 gradient 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 6. *Let* $\mathbf{x}_0^{nd}$ *and* $\mathbf{x}_0^d$ *be the optimal basic solutions of the LPP (5), and let the perturbed solutions take the form given by Equation (8). Let* $\mathbf{x}_0^{nd}$ *be a non–degenerate 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* $\mathbf{x}_0^d$ *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 gradient* ΔJ *of the objective function (9) is continuous and constant over all the interval* $\Lambda = \Lambda_{\mathcal{B}_1} \cup \Lambda_{\mathcal{B}_2} \cup \Lambda_{\mathcal{B}_3}$.

Perturbation of the matrix coefficients

We assume that the basis matrix $\mathbf{A}_\mathcal{B}$ varies linearly with the parameter $\lambda \in \mathbb{R}$, according to $\mathbf{A}_\mathcal{B}(\lambda) = \mathbf{A}_\mathcal{B} + \lambda \mathbf{A}^*_\mathcal{B} = \mathbf{A}_\mathcal{B} + \lambda \mathbf{a}^* \mathbf{e}_j^T$, i.e. we assume that only the j–th column of $\mathbf{A}_\mathcal{B}$ may vary. The results we present here also hold when a single row of $\mathbf{A}_\mathcal{B}$ varies with the parameter λ. Nevertheless this case is less relevant in the context of HSPN. In fact perturbations of matrix $\mathbf{A}$ correspond in the HSPN framework to variations of the arc–weights between continuous places and transitions, as it can be seen from Equation (4). 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.

Let $\mathbf{x}^o$ be an optimal basic solution of (5) and $\mathcal{B}$ an associated optimal basis. We recall the matrix equality:

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

Then the perturbed optimal solution $\mathbf{x}^o(\lambda)$ has basic components:

$$\mathbf{x}_{\mathcal{B}}^o(\lambda) = \mathbf{A}_{\mathcal{B}}^{-1}(\lambda)\mathbf{b} = \mathbf{x}_{\mathcal{B}}^o - \frac{\lambda}{1+v\lambda}\mathbf{x}_{\mathcal{B}}^* \tag{12}$$

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

$$\mathbf{r}(\lambda) = \mathbf{c}_{\mathcal{B}}^T\mathbf{A}_{\mathcal{B}}^{-1}(\lambda)\mathbf{A} - \mathbf{c} = \mathbf{r}^o - \frac{\lambda}{1+v\lambda}\mathbf{r}^* \tag{13}$$

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

$$J(\lambda) = \mathbf{c}_{\mathcal{B}}^T\mathbf{x}_{\mathcal{B}}^o(\lambda) = \mathbf{c}_{\mathcal{B}}^T\mathbf{x}_{\mathcal{B}}^o - \frac{\lambda}{1+v\lambda}\mathbf{c}_{\mathcal{B}}^T\mathbf{x}_{\mathcal{B}}^*. \tag{14}$$

Equations (12-14) 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: (1) non–singularity of the basis matrix, i.e., $1 + v\lambda > 0$, (2) non-negativity of the basic variables, $\mathbf{x}_{\mathcal{B}}^o(\lambda) \geq \mathbf{0}$ and (3) non-negativity of the relative cost coefficients $\mathbf{r}(\lambda) \geq \mathbf{0}$, i.e. the optimality condition. The bounds for the parameter λ can be computed as follows. Let:

$$\mathbf{y} = \begin{bmatrix} 1 \\ \mathbf{x}^o \\ \mathbf{r} \end{bmatrix}, \quad \mathbf{y}^* = \begin{bmatrix} 0 \\ \mathbf{x}^* \\ \mathbf{r}^* \end{bmatrix}$$

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:

$$\underline{\lambda}_{\mathcal{B}} = \begin{cases} -\infty & \text{if} \quad I^+ = \emptyset \\ \max_{i\in I^+} \left\{-\frac{y_i}{vy_i - y_i^*}\right\} \end{cases} \tag{15}$$

and

$$\overline{\lambda}_{\mathcal{B}} = \begin{cases} +\infty & \text{if} \quad I^- = \emptyset \\ \min_{i\in I^-} \left\{-\frac{y_i}{vy_i - y_i^*}\right\} \end{cases} \tag{16}$$

¿From Equations (12) and (14) 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 j-th column vector of $\mathbf{A}$, say $\mathbf{a}_j = \lambda\mathbf{a}^*$, is a non–linear function of the parameter λ and for each value of $\lambda \in \Lambda_{\mathcal{B}}$, for $\lambda \neq -\frac{1}{v}$, it can be easily computed as

$$\Delta J(\lambda) = -\frac{1}{(1+v\lambda)^2}\mathbf{c}_{\mathcal{B}}^T\mathbf{x}_{\mathcal{B}}^* \tag{17}$$

Figure 3: A re–entrant production line.

Figure 4: Feasible regions for the net in Figure 3.

An example: re–entrant production lines.

In this section we consider a simple HSPN which represents a re-entrant production line, as shown in Figure 3, that will clarify our development. 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:

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

Now solving for $J = v_2 + v_3$, we obtain the optimum firing speeds 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 LPPs, 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 $V_i + \Delta V_i$.

Now let $V_1 = 5$, $V_2 = 5$ and $V_3 = 4$. In Figure 4 we have shown the feasible regions in the plane $v_1 = 5$ for the LPP considered in this example and for values of $\alpha \in [0,1]$. The thin lines labelled by the different values of α represent the fourth constraint. Note that for $\alpha = 0$ we obtain the same results already developed in Example 4 where we have two optimal basic solutions $\mathbf{v}_{(A)} = [V_1, V_2 - V_3, V_3]^T$ and $\mathbf{v}_{(B)} = [V_1, V_2, 0]^T$, i.e. points (A) and (B), and the optimal value of the objective function J is equal to V_2. For $0 < \alpha < \frac{V_3}{V_2}$ there is a unique non–degenerate optimal basic solution (point (C)), $\mathbf{x}^o_{\mathcal{B}_C} = [v_1, v_2, v_3, s_3]^T$, where s_3 is the slack variable associated to third constraint, which can be analytically computed as

$$\mathbf{x}^o_{\mathcal{B}_C} = [V_1, V_2, V_1 - (1-\alpha)V_2, V_3 - V_1 + (1-\alpha)V_2]^T$$

with an associated optimal basis $\mathcal{B}_C = \{1, 2, 3, 6\}$, which yields an optimal objective function value equal to $V_1 + \alpha V_2$. For $\alpha = \frac{V_3}{V_2}$ we have a degenerate optimal basic solution (point (D)). Finally for $\alpha > \frac{V_3}{V_2}$ the fourth constraint becomes redundant and the unique optimal basic solution (point (D)) is simply given by $\mathbf{x}^o_{\mathcal{B}_D} = [V_1, V_2, V_3, V_1 - (1-\alpha)V_2 - V_3]^T$ with optimal basis $\mathcal{B}_D = \{1, 2, 3, 7\}$ and optimal objective function value equal to $V_2 + V_3$. Therefore we will only consider perturbations of the parameter α for $\alpha \in (0, \frac{V_3}{V_2})$ which yield non–trivial sensitivity analysis for the objective function J.

Now computing the bounds for the parameter α to obtain the allowable range $\Lambda_{\mathcal{B}_C}$ of the optimal basis $\mathcal{B}_C = \{1, 2, 3, 6\}$, we must consider $I^+ = \{3\}$ and $I^- = \{4\}$, where $v = 0$ and $\mathbf{x}^*_{\mathcal{B}_C} = [0, 0, -V_2, V_2]^T$. Then it does 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 gradient of the objective function J with respect to the perturbation of the reworking factor α by making use of Equation (17). In this simple case it does result $\Delta J(\alpha) = V_2$ which is constant over the interval $\Lambda_{\mathcal{B}}$.

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 $\Lambda_{\mathcal{B}_C}$ of the design parameter V_1 as follows:

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

within which the IFS vector and the objective function vary linearly with λ. As a numerical example if $\alpha = .5$ then we have:

$$\begin{aligned}
\Delta V_1 &= [V_1 - 2, 5, V_1 + 1.5] && \text{for} \quad V_1 \\
\Delta V_2 &= [V_2 - 3, V_2 + 5] && \text{for} \quad V_2 \\
\Delta V_3 &= [V_3 - 1.5, +\infty] && \text{for} \quad V_3
\end{aligned}$$

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

5 CONCLUSION

Automated manufacturing system can be naturally modeled by HSPNs. Optimal speed allocation and sensitivity analysis have been 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. Our main contribution has been to offer a simple tool for making sensitivity analysis of an HSPN based on parametric linear programming techniques used in LP problems to efficiently evaluate gradients of the performance measures, such as machine utilization and system throughput, with respect to the design parameters, i.e. the MFSs, and to the structural parameter, i.e. the routing coefficients. Our future goal will be to apply this hybrid model along with its sensitivity analysis tool for the optimal design and control of discrete event dynamic processes.

References

[1] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, G. Franceschinis, *Modelling with Generalized Stochastic Petri Nets*, Wiley Series in Parallel Computing, John Wiley and sons, 1995.

[2] H. Alla, R. David, J. Le Bail, "Hybrid Petri Nets", *ECC'91 European Control Conference*, (Grenoble, France), pp. 1472–1477, July 1991.

[3] F. Balduzzi, G. Menga, "A State Variable Model for the Fluid Approximation of Flexible Manufacturing Systems," *Proc. 1998 IEEE Int. Conf. on Robotics and Automation*, (Leuven, Belgium), pp. 1172–1178, May 1998.

[4] F. Balduzzi, A. Giua, G. Menga, "Hybrid Stochastic Petri Nets: Firing Speed Computation and FMS Modelling", *WODES'98, Proc. of the Fourth Workshop on Discrete Event Systems*, (Cagliari, Italy), to appear.

[5] T. Gal, *Postoptimal analysis, parametric programming and related topics*, Mac-Graw Hill Inc., 1979.

[6] V.G. Kulkarni, K.S. Trivedi, "FSPNs: Fluid Stochastic Petri Nets" *Lecture Notes in Computer Science*, Vol. 691, M. Ajmone Marsan (ed.), *Proc. 14th International Conference on Applications and Theory of Petri Nets*, Springer-Verlag, Heidelberg, pp. 24–31, 1993.

[7] T. Murata, "Petri Nets: Properties, Analysis and Applications", *Proceedings IEEE*, Vol. 77, No. 4, pp. 541–580, 1989.

[8] K.G. Murty, *Linear and Combinatorial Programming*, John Wiley and Sons, 1976.

[9] J. Renegar, "Some Perturbation Theory for Linear Programming", *Mathematical Programming*, No. 65, pp. 73–91, 1991.

[10] R. Suri, B.R. Fu, "Using Continuous Flow Models to Enable Rapid Analysis and Optimization of Discrete Production Lines - A Progress Report", *Proc. of the 19th Annual NSF Grantees Conference on Design and Manufacturing Systems Research*, 1993.