A partial approach to the problem of deadlocks in processes with resources

Fernando Tricas and Joaquín Ezpeleta
Departamento de Informática e Ingeniería de Sistemas
Centro Politécnico Superior – Universidad de Zaragoza
María de Luna 3, 50015 Zaragoza (SPAIN)
E-mail: {ftricas,ezpeleta}@posta.unizar.es

Abstract

In the modeling and control of manufacturing systems one of the aims is to obtain a system that can run in an automatic way. This means that the software in charge of the monitoring and control of the system must be able to deal with a set of problematic situations (system failures, system deadlocks, overflow of storing facilities, etc.). In this paper we are going to concentrate on one of the previous problems: the deadlock problem. Deadlock problems are very usual in concurrent system where, even if the execution of each process is correct, the competition of a set of processes for the same set of resources can give rise to deadlocks. This problem is very common to all the environments where processes share resources (or communicate via messages): databases, operating systems, manufacturing systems, etc. The present work focuses on the deadlock problem for a class of systems that appears in Flexible Manufacturing Systems (FMS) environments. We are going to identify a class of nets that appears frequently in this environment; for this class (called here ES^3PR) we characterize deadlock problems to be related to structural elements of the Petri net model (siphons). We use this characterization in order to establish a control policy for deadlock prevention. The class of nets and results obtained in this paper generalizes some results presented in previous work [ECM95] and shows a way to improve the ones presented in [TM95].

Keywords: Deadlock Prevention, Flexible Manufacturing Systems, Petri Nets

1 Introduction

The software design for concurrent systems has to deal with the satisfaction of conditions of good behavior. One of these requirements is the absence of deadlocks; that is, the system must be able to ensure that each production order will be accomplished [SV89, VN92]. Our research has studied this problem in concurrent systems where a set of sequential processes compete for the use of a set of shared resources. This kind of systems appear in the modeling of FMS. The idea of process comes when we try to represent the fact that different types of parts must be processed following different processing sequences. In order to process a type of parts a set of resources are needed (robots, AGVs, machines, tools, ...) and even if different types of parts usually are processed following different sequences of steps, they can need to use some common resources. So, we have a set of concurrent processes that share a set of resources. This kind of problems are hard to manage without formal tools that help to represent and to demonstrate properties over the system.

Several models have been used to deal with this kind of problems, being Petri nets [Mur89] the one chosen in this paper. Petri net models have been widely used in FMS environment [VN92, DHP$^+$93, DAJ95, PX96]. In terms of Petri nets, the ability of the system to finish production orders can be seen as the study of absence of deadlocks. To deal with this problem, three different approaches are usually adopted. Deadlock recovery, deadlock prevention, and finally, the avoidance approach

This work fits in the prevention category. Once a model of the system has been constructed, the proposed method consists of the addition of some places that will control the entry of processes to the system in order to avoid dangerous situations. In previous work [ECM95] a theory of liveness (the deadlock freeness is a particular case) for these systems has been developed under a set of restrictions over the system. In this work the resources must be taken one copy at a time (for the same resource and if several copies are needed). The proposed policy is implemented via addition of elements of the Petri nets model (places and arcs) but is not optimal (there can be policies less restrictive that can wear to live models). In this context we can say that a control policy is optimal if it is minimally restrictive.

The organization of the rest of the paper is as follows: section 2 presents some notation and basic concepts and properties of Petri nets. In section 3, the class of nets we are interested in is presented. In section 4 the policy is presented and, finally in section 5, conclusions and work to be done are presented.

2 Notation and basic concepts

We assume the reader is familiar with basic Petri nets concepts. They can be found, for instance, in [Mur89]. In order to make the paper as selfcontained as possible the next section introduces, in a very compact way, a set of definitions that will be used later on.

Petri nets: A Petri net (or Place/Transition net) is a 3-tuple $\mathcal{N} = \langle P, T, W\rangle$ where P and T are two non-empty disjoint sets whose elements are called *places* and *transitions*. In a generic way, elements belonging to $P \cup T$ are called *nodes*. The set $W : (P \times T) \cup (T \times P) \rightarrow \mathbb{N}$ is the *flow relation*: if $W(x, y) > 0$, then we say that there is an arc from x to y, with weight or multiplicity $W(x, y)$. Ordinary nets are those where $W : (P \times T) \cup (T \times P) \rightarrow \{0, 1\}$. Given a net $\mathcal{N} = \langle P, T, W\rangle$ and a node $x \in P \cup T$, ${}^\bullet x = \{y \in P \cup T \mid W(y, x) > 0\}$ is the *pre-set* of x, while $x^\bullet = \{y \in P \cup T \mid W(x, y) > 0\}$ is the *post-set* of x. This notation is extended to a set of nodes as follows: given $X \subseteq P \cup T$, ${}^\bullet X = \bigcup_{x \in X} {}^\bullet x$, $X^\bullet = \bigcup_{x \in X} x^\bullet$. A *marking* is a mapping $M : P \longrightarrow \mathbb{N}$; in general, we use the multi-set notation for markings: $M = \sum_{p \in P} M(p).p$. When talking about a set of places $S \subseteq P$, $M(S) = \sum_{p \in S} M(p)$. The pair $\langle \mathcal{N}, M_0\rangle$, where $\mathcal{N}$ is a net and M_0 is an (initial) marking, is called a *marked Petri net* (or Place/Transition system). A transition $t \in T$ is *enabled* for a marking M iff $\forall\ p \in {}^\bullet t,\ M(p) \geq W(p, t)$; this fact will be denoted as $M[t\rangle$; when fired (in the usual way), this gives a new marking M'; this will be denoted as $M[t\rangle M'$. A marking M' is *reachable* from another marking M iff there exists a firing sequence $\sigma = t_1 t_2 ... t_n$ so that $M[t_1\rangle M_1[t_2\rangle M_2 ... M_{n-1}[t_n\rangle M'$, and is denoted as $M[\sigma\rangle M'$; given a set $X \subseteq T$, by $\sigma_{|X}$ we denote the projection of the firing sequence σ with respect to the set X. The set of markings reachable from M in $\mathcal{N}$ is denoted as $\mathcal{R}(\mathcal{N}, M)$. A Petri net is self–loop free iff $\neg(W(x, y) > 0 \wedge W(y, x) > 0)$. A self-loop free Petri net $\mathcal{N} = \langle P, T, W\rangle$ can be alternatively represented as $\mathcal{N} = \langle P, T, C\rangle$ where C is the net *flow matrix*: a $P \times T$ integer matrix so that $C = C^+ - C^-$ where $C^+[p, t] = W(t, p)$; $C^-[p, t] = W(p, t)$. Given $A \subseteq P$ and $B \subseteq T$, $C[A, B]$ is the submatrix of C whose rows and columns correspond to places in A and transitions in B.

A path in a Petri net $\mathcal{N} = \langle P, T, W\rangle$ is a sequence of nodes $x_1, x_2, \ldots, x_n$ from $P \cup T$ such that $x_{i+1} \in x_i{}^\bullet$, for all $i \in \{1, ..., n-1\}$. A path is simple iff for all $i \neq j$, $x_i \neq x_j$. A circuit is a simple path such that $x_1 = x_n$.

Some structural objects and properties of a marked Petri net: A set of places S is a *structural siphon* iff ${}^\bullet S \subseteq S^\bullet$, and it is a *structural trap* iff $S^\bullet \subseteq {}^\bullet S$. A marked Petri net $\langle \mathcal{N}, M_0\rangle$ is *deadlock-free* iff $\forall\ M \in \mathcal{R}(\mathcal{N}, M_0)$, $\{t \in T \mid M[t\rangle\} \neq \emptyset$. A transition t is *live* iff $\forall\ M \in \mathcal{R}(\mathcal{N}, M_0)$, $\exists\ M' \in \mathcal{R}(\mathcal{N}, M)$ so that $M'[t\rangle$. A marked net is live iff every transition is live. We will say that a transition t is *dead for a reachable marking* M iff there is no reachable marking from M that enables t. Notice that a transition is live iff it is not dead for each reachable marking. *Semiflows* are natural anullers of matrix C. Right and left natural anullers are called *T–semiflows* and *P–semiflows*, respectively. A semiflow is called *minimal* when its support is not a strict superset of the support of any other, and the greatest common divisor of its elements is one.

Some notations over sets: Let $A \subseteq B$ be two sets and let Y be an A-indexed vector. We denote by $Y_{A|B}$ the B-indexed vector defined as follows: $\forall x \in B$, $Y_{A|B}(x)$ = If $p \in A$ then $Y(x)$ else 0. As a particular case, given the A-indexed unit vector $\mathbb{1}$. We denote by $\mathbb{1}_{A|B}$ the B-indexed vector defined as follows: $\forall x \in B$, $\mathbb{1}_{A|B}(x)$ = If $p \in A$ then 1 else 0. Let $A \subseteq B$ be two sets and let Y be an B-indexed vector. We denote by $Y(A) = \sum_{a \in A} Y(a)$

Let us present a set of properties relating the Petri net behavior and siphons.

Proposition 2.1 ([Com72]) *Let $\langle\mathcal{N},M_0\rangle$ be a marked ordinary Petri net. Let $M\in\mathcal{R}(\mathcal{N},M_0)$ be a reachable marking such that a total deadlock occurs. Then there exists a siphon S such that $M(S)=0$.* □

Now we are going to concentrate on a class of nets that is wider than the class of ordinary nets, but that is a proper subset of the class of generalized nets. The weights of arcs belonging to the output of transitions are allowed to be greater than one but only ordinary (weight 1) arcs can be placed in a transition input arc. The following proposition extends proposition 2.1 to the considered less restrictive class of nets.

Proposition 2.2 *Let $\langle\mathcal{N},M_0\rangle$ be a marked generalized Petri net such that $\forall(p,t)\in P\times T$, $W(p,t)\leq 1$. Let $M\in\mathcal{R}(\mathcal{N},M_0)$ be a reachable marking such that a total deadlock occurs. Then there exists a structural siphon S such that $M(S)=0$.* □

The following proposition remembers the most important behavioral property of a siphon. When a marking such that the siphon is empty is reached, all the markings reachable from it will maintain the siphon empty. Therefore, no transition belonging to the output of the siphon will be able to be fired in the future.

Proposition 2.3 *Let $\langle\mathcal{N},M_0\rangle$ be a marked generalized Petri net such that $\forall(p,t)\in P\times T$, $W(p,t)\leq 1$. Let $M\in\mathcal{R}(\mathcal{N},M_0)$ and S a siphon such that $M(S)=0$. Then, $\forall M'\in\mathcal{R}(\mathcal{N},M)$, $M'(S)=0$.* □

We are going to describe in the next section the Petri net subclass we will obtain modeling concurrent processes sharing resources.

3 The class of the ES^3PR nets.

As stated previously, this paper is devoted to the study of a class of nets that appears in the modeling of Flexible Manufacturing Systems. Let us introduce in an informal way, and by means of an example, the new class of nets.

3.1 An intuitive approach.

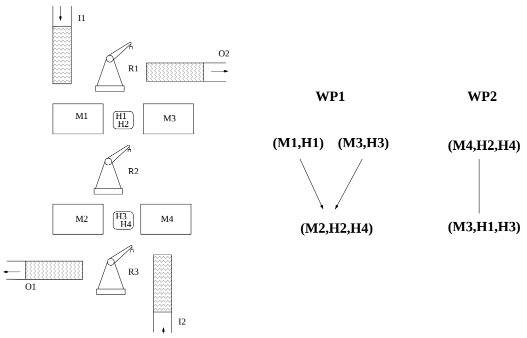
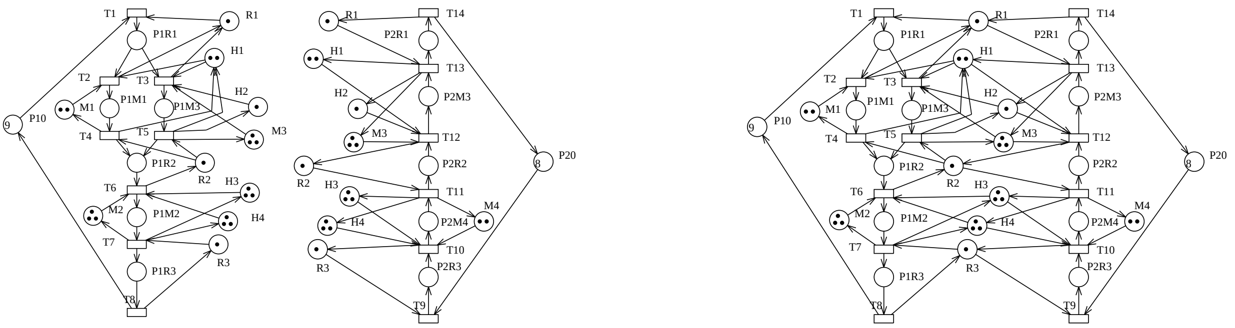


Figure 1: *Layout of an FMS where the parts are going to be processed.*

Figure 1 shows the layout of an hypothetical FMS cell composed of three robots ($R1$, $R2$ and $R3$) that are used for moving parts between machines ($M1$, $M2$, $M3$ and $M4$). These machines use different tools for their work ($H1$, $H2$, $H3$ and $H4$). Robot $R1$ can load parts from conveyor $I1$ to machines $M1$ and $M3$ and unload parts from machine $M3$ to conveyor $O2$. Machines $M1$ and $M3$ can use tools $H1$ and $H3$. Robot $R2$ can load and unload machines $M1$, $M2$, $M3$ and $M4$. Machines $M2$ and $M4$

can use tools $H2$ and $H4$. Finally, robot $R3$ can unload machine $M2$ to load exit conveyor $O1$ and can load machine $M4$ from conveyor $I2$. In this cell two types of parts must be processed according to their own process plans. Parts of type 1 arrive at the system by conveyor $I1$ and leave it by conveyor $O1$. They can be processed first in machine $M1$ or $M3$. $M1$ needs to take tool $H1$ to accomplish its work and $M3$ needs to take tools $H1$ and $H3$. After that, the part to be processed in machine $M2$ that need to use tools $H2$ and $H4$ (according to the process plan WP1). Parts of type 2 arrive to the system by conveyor $I2$ and leave it by conveyor $O2$. They first need to be processed in machine $M4$. This machine takes tools $H2$ and $H4$ to do the processing. Later, they are processed in machine $M3$ that needs to use tools $H1$ and $H3$ (according to process plan WP2). We are going to suppose that machines $M1$ and $M4$ can process two parts concurrently; machines $M2$ and $M3$ can manage three parts each and that the system has two tools $H1$ and two tools $H3$, one tool $H2$ and three $H4$. Each robot can hold a single part at a time.



(a) (b) (c)

Figure 2: *Petri nets that represent the processing of parts of type 1 and 2 and the net obtained by composition of them*

Let us pay a little more attention to the processing of parts of type 2. The processing sequence of a part of this type is as follows: in order to be processed, the part must be held by robot $R3$, loaded into machine $M4$ and processed, hold by robot $R2$ that unloads it from $M4$ and loads it into machine $M3$, and finally be unloaded from it to $O2$ by robot $R1$. In the Petri net of Figure 2–(b), the sequence of states in which a part of type 2 can stay during its processing in the system is modeled using places $P2R3, P2M4, P2R2, P2M3$ and $P2R1$. In this way, a token in place $P2R2$ models a part of type 2 that has already been processed in machine $M4$ (once robot $R3$ has loaded the part in it) and that is being held by robot $R2$ waiting to be loaded into machine $M3$. Notice that in order to reach this state, the sequence $T9T10T11$ has been fired. Moreover, in this state, the place $R2$ has empty marking; this marking represents the fact that robot $R2$ is carrying the part and that it cannot do any other thing while it is busy with the part. Firing of $T11$ has released one token to each one of $H3$, $H4$ and $M4$ modeling that machine $M4$ has finished and has been unloaded leaving the copies of tools $H3$ and $H4$ in their respective places in order to be used in the processing other parts. Finally, to fire $T12$ arcs from $M3$, $H1$ and $H2$ point out the need of available space in machine $M3$ and enough tools $H1$ and $H2$ to begin the next processing step of the part. If this transition fires, a token will be put in place $R2$, representing that the robot has finished and it can accomplish new movings.

Let us give now a more formal definition of the ES^3PR class. It is an extension of the ones presented in [BK90, HC94, XHC96, Jen96] in the sense that process plans with decisions are now allowed and of the ones presented in [ECM95, Jen96] in the way resources can be used and released. ES^3PR is a subclass of the class WS^3PSR, presented in [TM95] because in WS^3PSR several copies of a resource can be hold in a single processing step.

A net belonging to this class is composed of a set of state machines (each one models the states that can be reached by a type of part). These state machines use resources, modeled as monitors whose marking represents the availability of resources modeled by it. Each one of these resources can be shared between several state machines, modeling several processes that need to use a set of common resources. In this way, the marking of the system shows the state of the cell: the marking of every state machine shows the state of the parts of the type it models, while the marking of each place modeling resources represents the available copies of it in the considered state.

3.2 Definition of the ES³PR nets

The definition of the class of models studied is made in a constructive way similar to the one made in [ECM95, TM95].

Definition 3.1 *A Simple Sequential Process, S^2P, is an ordinary Petri net $\mathcal{N}=\langle P\cup\{p^0\},T,W\rangle$ where:*

1. $P\neq\emptyset$, $p^0\notin P$
2. *$\mathcal{N}$ is a strongly connected state machine.*
3. *All the circuits in $\mathcal{N}$ contain the place p^0.* □

The special place p^0 is called the idle state place. This place is used to introduce in the system the idea of idle state, that is, the state in which the system is not working. Note that the marking of this place limits the number of total tokens that are allowed to be in the system. Note also that choosing an adequate marking this place becomes implicit. In this way we can use this type of nets to model open systems, taking into account this fact. Addition of this place helps in simplifying the definition. This place also introduces (via condition 3) the inevitability of the idle state of the system. That is, unless it reaches a deadlock situation, every firing sequence related to a token of the net will carry the token to the idle state place. From the modeling point of view a S^2P represents the set of different processing sequences that a part can follow across the system. In this sense, the inevitability of the idle state place represents the fact that the processing of the part must be finished.

In Figure 2–(b) (do not consider places $R1, H1, H2, M3, R2, H3, H4, R3$ yet) we can see that the set of places from the Petri net modeling processing of parts of type 2: $P=\{P2R3,\ P2M4,\ P2R2,\ P2M3,\ P2R1\}$ and the place $p^0=P20$ with all the transitions form a S^2P. This net is trivially a strongly connected state machine and the only circuit it has contains place $P20$.

Places in P are called state places. Different states are configured as state machines; that is, decisions are allowed (a part can have to choose between several different paths).

Next, an Extended Simple Sequential Processes with Resources ES^2PR net is defined as a S^2P that uses resources in the states of the system that are not the idle one. In this class of nets, as said in the intuitive introduction, a process state can need the use of several resources simultaneously

To represent the resources in the model, we are going to add to the net a new set of places called P_R (Resource Places) and a set of arcs that join these new places with transitions from the S^2P. These arcs model the way this resources are hold and released to reach different processing states.

In Figure 2–(b) we can see that when we fire transitions $T9$ and $T10$ to mark place $P2M4$, places $M4$, $H3$ and $H4$ (that model resources) lost tokens; that is, to fire $T10$ the presence of, at least, one token in each one of these places is needed.

Definition 3.2 *An Extended Simple Sequential Process with Resources, ES^2PR, is a generalized self–loop free Petri net $\mathcal{N}=\langle P\cup\{p^0\}\cup P_R,T,W\rangle$, such that:*

1. *The subnet generated by the set $X=P\cup\{p^0\}\cup T$ is a S^2P.*
2. $P_R\neq\emptyset$ *and* $(P\cup\{p^0\})\cap P_R=\emptyset$.
3. $\forall t\in T,\ \forall p\in{}^\bullet t,\ W(p,t)=1$.
4. $\forall r\in P_R\ ,\exists|$ *minimal P–semiflow* Y_r *such that* $\{r\}=\|Y_r\|\cap P_R,\ p^0\notin\|Y_r\|,\ \|Y_r\|\cap P\neq\emptyset\wedge Y_r(r)=1$.
□

In the net depicted in Figure 2 we have $P_R=\{M3, M4, R1, R2, R3, H1, H2, H3, H4\}$ and $Y_{M3}=\mathbf{1}.M3 + 1.P2M3$, $Y_{M4} = \mathbf{1}.M4 + 1.P2M4$, $Y_{R1} = \mathbf{1}.R1 + 1.P2R1$, $Y_{R2} = \mathbf{1}.R2 + 1.P2R2$, $Y_{R3} = \mathbf{1}.R3 + 1.P2R3$, $Y_{H1} = \mathbf{1}.H1 + 1.P2M3$, $Y_{H2} = \mathbf{1}.H2 + 1.P2M3$, $Y_{H3} = \mathbf{1}.H3 + 1.P2M4$ and $Y_{H4} = \mathbf{1}.H4 + 1.P2M4$.

In order to simplify the notation we are going to name $P^0=\{p^0\}$. If X is a minimal T–semiflow of the net, F_X denotes the set defined as $F_X = p_0t_0p_1t_1...p_kt_k$ such that $\{t_0, t_1..., t_k\} = \|X\|$, $p_i \in {}^\bullet t_i \cap (P \cup P^0)$. Note that this is a well–defined set because in the class of nets we are studying $\forall t_i \in T$, ${}^\bullet t_i \cap (P \cup P^0)$ is a unique place. Moreover, let $\mathcal{N}$ be an ES^2PR. Let F be a circuit of the net. It induces the following sets over the net: $F_T = F \cap T$ and $F_P = F \cap P$. These two sets are the set of nodes of the path (transitions and places, respectively).

For example, in the net of Figure 2–(a) the T–invariant: $X_1 = T1 + T2 + T4 + T6 + T7 + T8$ induces the set: $F_{X_1} = \{P10, T1, P1R1, T2, P1M1, T4, P1R2, T6, P1M2, T7, P1R3, T8\}$. In the net of Figure 2–(a) the circuit F_{X_1} induces the sets: $F_{X_1T} = \{T1, T2, T4, T6, T7, T8\}$ and $F_{X_1P} = \{P10, P1R1, M1, P1R2, P1M2, P1R3\}$

Lemma 3.3 ([TE97]) *Let $\mathcal{N}$ be an ES^2PR*

- *Let F be a circuit of the net that does not contain places from P_R. Then the set F_T induces the following minimal T–semiflow of the net: $X_F = \mathbb{1}_{F_T|T}$*
- *Reversely, let X be a minimal T–semiflow of the net. The induced set F_X as defined previously is a circuit.* □

Semiflows are important because they induce invariant relations that are useful for reasoning about the behaviour. T–semiflows represent firing sequences that do not change the marking. Using the previous lemma, we can also see them as the actions completing the processing of a part. If we have a look to the physical system modeled with these nets, a T-semiflow represents one of the possible sequences of steps that a part can follow during the execution of its working plan. For example, in the net of the Figure 2–(a), T–semiflow X_1 models a part following the working plan $WP1$. In consequence, the set of T–semiflows represents the set of processing sequences in the system. The cyclical behaviour that T–semiflows reveal is a desirable property for the kind of systems we are modeling: it is related to the idea of finishing of parts that begin their processing In consequence, if $\mathcal{N}$ is an ES^2PR, and we denote $\mathcal{F} = \{F \mid$ circuit of $\mathcal{N}$ such that does not contain places of $P_R\}$ we can write the set of minimal T–semiflows as $\mathcal{T} = \{X_F \mid F \in \mathcal{F}\}$

In the net depicted in Figure 4–(a) we have: $\mathcal{T}^1 = \{X_{11} = tp1 + tp2 + tp31 + t1p4 + tp5,\ X_{12} = tp1 + tp2 + tp32 + t2p4 + tp5\}$ and $\mathcal{T}^2 = \{X_{21} = tq1 + tq2 + tq31 + t1q4,\ X_{22} = tq1 + tq2 + tq32 + t2q4\}$ For $\mathcal{T}^1$, the circuits are: $F_{X_{11}} = \{p0, tp1, p1, tp2, p2, tp31, p31, t1p4, p4, tp5\}$ and $F_{X_{12}} = \{p0,\ tp1,\ p1,\ tp2,\ p2,\ tp32,\ p32,\ t2p4,\ p4,\ tp5\}$.

Let us now introduce the class of initial markings that we are going to consider. These initial markings must be such that: 1) They represent an idle state for the system. This means that: a) The initial marking of places that represent states modeling some system activity must be 0 (no activity is being carried out). b) The initial marking of p^0 represents the maximal number of parts of the type modeled by the embedded S^2P that are allowed to be concurrently processed in the system. c) Considering the previous points, in the initial state no process is activated and then, no resource is used. In consequence, the initial marking of places in P_R corresponds to the capacity of the resources to process parts. Notice that it is then natural to assume that this capacity is greater than 0. 2) The system is well defined. We impose that from the initial marking, each minimal T–semiflow can be fired isolated from the rest of T–semiflows of the system (activating only one copy of the process). Let us remember here that a T–semiflow represents a processing sequence of the system. It makes sense to impose that each one of these sequences can be completed if it is carried out isolated from the rest of sequences. It would be a very poor system design if there were a processing sequence that could not be executed in an isolated way. The initial markings verifying the previous conditions will be called acceptable. We are going now to write them in a more formal way.

Definition 3.4 *Let $\mathcal{N}$ be an ES^2PR. The marking M_0 is acceptable for $\mathcal{N}$ iff: 1) $M_0(p^0)>0$. 2) $\forall p\in P$, $M_0(p)=0$. 3) $\forall r\in P_R$, $M_0(r)>0$. 4) All the minimal T–semiflows can be fired sequentially.* □

In the following only acceptable initial markings will be considered. Now, we are going to present a lemma that points out the relation of structural components with acceptable initial markings in ES^2PR nets.

Lemma 3.5 ([TE97]) *Let $\langle\mathcal{N},M_0\rangle$ be a marked ES^2PR and let X be a minimal T–semiflow of the net and F_X the induced circuit. If $\forall r\in P_R$, $M_0(r)\geq \max_{p\in\|Y_r\|} Y_r(p)$ then transitions in F_X can be fired sequentially.* □

Using this lemma, point 4 of definition can be substituted by: $\forall r\in P_R$, $M_0(r)\geq \max_{p\in\|Y_r\|} Y_r(p)$ Next we are going to present ES^3PR as nets composed of ES^2PR via shared resources. That is, after we have modeled a set of process sequences with ES^2PR, we need to represent the interaction between them; this interaction is restricted to the use of a set of common resources needed by several processes. Previously we need to present a formal definition of the way this composition can be made.

Definition 3.6 *Let $\mathcal{N}_i=\langle P_i,T_i,W_i\rangle$, be two generalized Petri nets. Then:*

1. *$\mathcal{N}_1$ and $\mathcal{N}_2$ are composable (via fusion of places) iff $T_1\cap T_2=\emptyset$ and $P_C=P_1\cap P_2\neq\emptyset$*
2. *The net $\mathcal{N}=\langle P,T,W\rangle$ where $P=P_1\cup P_2$, $T=T_1\cup T_2$, $W(p,t)$ = If $t\in T_1$ then $W_1(p,t)$ else $W_2(p,t)$ is named the composed net of $\mathcal{N}_1$ and $\mathcal{N}_2$ and it is represented by: $\mathcal{N}=\mathcal{N}_1\circ\mathcal{N}_2$.* □

This definition can be easily extended to more than two nets via sequential composition of the nets: first we can compose two of them, then we can compose the resulting one with a third one and so on.

Definition 3.7 *Let $i\in I_{\mathcal{N}}$ be a set of indexes, let $\mathcal{N}_i=\langle P_i,T_i,W_i\rangle$, be a set of generalized Petri nets. We denote by means of $\mathcal{N}=\bigcirc_{i\in I_{\mathcal{N}}}\mathcal{N}_i$ the net obtained as follows: if $card(I_{\mathcal{N}})=i_1$ then $\mathcal{N}=\mathcal{N}_{i_1}$ else $\mathcal{N}=\mathcal{N}_{i_1}\circ\bigcirc_{i\in(I_{\mathcal{N}}\setminus\{i_1\})}\mathcal{N}_i$.* □

We are going to use this definition to present ES^3PR nets.

Definition 3.8 *An ES^3PR, is the composition of a set of ES^2PR via shared resources.* □

In the following, when talking about an ES^3PR, $\mathcal{N}$, we will assume that it is of the the form $\mathcal{N}=\bigcirc_{i\in I_{\mathcal{N}}}\mathcal{N}_i$, being $\mathcal{N}_i$, $i\in I_{\mathcal{N}}$, the set of ES^2PR composing it. For this net it will be helpful to introduce the *set of indexes of a resource* as the set of indexes such that the resource belongs to the corresponding net. If $r\in P_R$, we can define the set of indexes of the resource as $I_r=\{i\in I_{\mathcal{N}}\mid r\in P_{R_i}\}$.

For example ES^3PR of Figure 2–(c) is composed of the two ES^2PR depicted in Figures 2–(a) and 2–(b) via the set of shared resources: $\{R1,R2,R3,M3,H1,H2,H3,H4\}$. For resource $R1$, $I_{R1}=\{1,2\}$ and for resource $M1$ the set $I_{M1}=\{1\}$.

This definition will be used to establish the acceptable marking for shared resources. Shared resources are the ones such that $card(I_r)>1$. The following definition extends the notion of acceptable marking for an ES^2PR to the new class of nets.

Definition 3.9 $\langle\mathcal{N},M_0\rangle$, $\mathcal{N}=\bigcirc_{i\in I_{\mathcal{N}}}\mathcal{N}_i$, $\mathcal{N}_i=\langle P_i\cup P_i^0\cup P_{R_i},T_i,W_i\rangle$ *is an acceptably marked ES^3PR iff 1) $\forall i\in I_{\mathcal{N}}$ $\forall p\in P_i\cup P_0^i$, $M_0(p)=M_{0_i}(p)$ and 2) $\forall i\in I_{\mathcal{N}}$ $\forall r\in P_{R_i}$, $M_0(r)=$ if $(card(I_r)=1)$ then $M_{0_i}(r)$ else $\max_{i\in I_r} M_{0_i}(r)$* □

In a composition of nets via shared resources the marking of places that are not common remains the same. For places that are common we can use the maximum initial marking for each place. We consider that this is a reasonable approach because if the model of some of the working processes shows an availability of copies for a resource greater than the availability of the same resource for a different plan it indicates that: 1) the maximum number of copies can be made available to the whole system and 2) this does not impose any restriction over the working processes except the ones related to the use of resources.

In the following, when speaking about an ES^3PR we are going to suppose that its initial marking is acceptable.

3.3 Some properties of ES³PR

In this short section we are going to show how T–semiflows for ES^3PR are.

Lemma 3.10 *Let $\mathcal{N} = \mathcal{N}_1 \circ \mathcal{N}_2$ be an ES^3PR obtained via composition of two ES^2PR and let Y^1, Y^2 such that $Y^{i^T} \cdot C_i = 0$, $Y \geq 0$. Then for $Y = Y^1_{(P^0_1 \cup P_{R_1} \cup P_1)|(P^0 \cup P_R \cup P)} + Y^2_{(P^0_2 \cup P_{R_2} \cup P_2)|(P_0 \cup P_R \cup P)}$, $Y^T \cdot C = 0$, $Y \geq 0$.*
Proof:
Trivial from the way the composition is made and basic properties of P–semiflows. □

This lemma can be generalized in a straightforward manner to the case of composition of any number of ES^2PR.
Example: For the net depicted in Figure 2–(c) the incidence matrix is:

$$
\begin{array}{c}
 \\ P10 \\ P1R1 \\ P1M1 \\ P1M3 \\ P1R2 \\ P1M2 \\ P20 \\ P1R3 \\ P2R3 \\ P2M4 \\ P2R2 \\ P2M3 \\ P2R1 \\ R1 \\ R2 \\ R3 \\ M1 \\ M2 \\ M3 \\ M4 \\ H1 \\ H2 \\ H3 \\ H4
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccccc} T1 & T2 & T3 & T4 & T5 & T6 & T7 & T8 & T9 & T10 & T11 & T12 & T13 & T14 \end{array} \\
\left(\begin{array}{cccccccccccccc}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & & & & & & \\
1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & & & & & & \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & & & & & & \\
0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & & & & & & \\
0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & & & & & & \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & & & & & & \\
 & & & & & & & & -1 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & 1 & -1 & 0 & 0 & 0 & 0 \\
 & & & & & & & & 0 & 1 & -1 & 0 & 0 & 0 \\
 & & & & & & & & 0 & 0 & 1 & -1 & 0 & 0 \\
 & & & & & & & & 0 & 0 & 0 & 1 & -1 & 0 \\
 & & & & & & & & 0 & 0 & 0 & 0 & 1 & -1 \\
-1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0
\end{array}\right)
\end{array}
$$

The structure of this matrix corresponds to the way we have composed the subsystems that constitute the whole system. It can be seen as a set of boxes with the following meaning: the first one on the left shows the flow relation for the nodes belonging to the S^2P for parts of type 1; the second one shows the flow relation for parts of type 2; finally the third one corresponds to the flow relation between nodes from P_R and transitions of the two ES^2PR that form the whole system. The last one can also be divided into two sub-matrices that correspond to the interaction between resources and each S^2P. Notice that for resources that are shared between the two subsystems the corresponding row has values different than 0 in both sub-matrices and that for resources only used in one of them, the row corresponding to the other subsystem is 0.

The following lemma helps us to see that the use of shared resources establishes a P–semiflow that takes into account the involved ES^2PR nets.

Lemma 3.11 ([TE97]) *Let $\mathcal{N} = \mathcal{N}_1 \circ \mathcal{N}_2$ be an ES^3PR obtained via composition of two ES^2PR. Then, $\forall r \in P_{R_1} \cup P_{R_2}$, $Y_r = Y_{r_1} + Y_{r_2}$ is a minimal P–semiflow.* □

This lemma can be extended to the composition of any number of ES^2PR. Finally, the next lemma establishes which ones are the T–semiflows of the composed net.

Lemma 3.12 ([TE97]) *Let $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_{\mathcal{N}}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P^0_i \cup P_{R_i}, T_i, W_i \rangle$ be an ES^3PR; the set of minimal T–semiflows of an ES^3PR is: $\mathcal{T} = \bigcup_{i \in I_{\mathcal{N}}} \mathcal{T}^i$*

The following lemma states that a minimal siphon that does not contain the support of a minimal P–semiflow cannot contain idle places. This serves also to describe a dangerous siphon as the composition of two parts: the resources of the siphon (S_R) and the process places of the siphon (S_P) (the meaning of the adjective "dangerous" will be understood later on).

Lemma 3.13 *Let $\mathcal{N}$ be an ES^3PR. Then, if S is a minimal siphon that does not contain the support of any P–semiflow $P^0 \cap S = \emptyset$, $P_R \cap S \neq \emptyset$, $P_P \cap S \neq \emptyset$.* □

3.4 Liveness Analysis of ES³PR Models.

In this section we are going to show some important behavioral properties ES^3PR nets. ES^3PR nets are a particular case of the ones presented in proposition 2.2 and then, for this class of nets it is true that if a total deadlock is reached, there exists an unmarked siphon. In the previously presented definitions we have taken an approach from the point of view of the system; we are going to show now some definitions from the resource point of view. The concept of holders of a resource is going to be presented as the set of process places whose marking implies that the resource has less tokens that in its initial marking.

Definition 3.14 (Holders of resource r) *Let $\langle \mathcal{N}, M_0 \rangle$, be a marked ES^3PR. For each resource $r \in P_R$ the set of holders of this resource is defined as: $\mathcal{H}(r) = \|Y_r\| \setminus \{r\}$.* □

Now we are going to particularize this definition to the ES^2PR subnets that compose a system. Then, the holders of the resource r in the subnet $\mathcal{N}_i$ are the holders of the ES^2PR with index i: $\mathcal{H}(r)_i = \mathcal{H}(r) \cap P_i$. For example in the net of Figure 4–(a) and for resource place $R1$, we have: $\mathcal{H}(R1) = \{p1, p2, p31, p32, q31\}$ and $\mathcal{H}(R1)_2 = \{q31\}$.

We are going to introduce now some technical lemmata that will help us to see that the existence of an unmarked siphon is a necessary and sufficient condition for the net to have a dead transition. The first of them remembers us that every transition can be fired at least once. This is a direct consequence of the initial marking we have chosen. The proof uses the idea that in a system, if we do not fire transitions from a subnet different from the one where transition t is, the system must behave as it were an independent system.

Lemma 3.15 *Let $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ be an ES^3PR; then, for all $t \in T$ there exists σ_t such that $M_0[\sigma_t\rangle M_t[t\rangle$*
Proof

Let us assume $t \in T_i$. Let F_t be a circuit that contains p_i^0 and t. Let X_{F_t} be the induced T–semiflow. Since M_0 is an acceptable marking the lemma holds trivially (remember lemmata 3.3, 3.12). □

The following lemma establishes that given a firing sequence that enables a transition t, the same sequence without firing of transitions of ES^2PR different than the one containing t also enables t. That is, if we have a firing sequence that enables a transition, the same sequence 'freezing' tokens that belong to other process types does not change the behavior of the transition.

Lemma 3.16 *Let $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ be a marked ES^3PR; Let $M \in \mathcal{R}(\mathcal{N}, M_0)$ and $j \in I_\mathcal{N}$ the index of an ES^2PR component such that $M(p_j^0) = M_0(p_j^0)$. If there exists a firing sequence σ such that $M[\sigma\rangle M'[t\rangle$, where $t \notin T_j$, then $\sigma_{|T \setminus T_j}$ verifies $M[\sigma_{|T \setminus T_j}\rangle M''[t\rangle$*
Sketch of the Proof:

If $\sigma_{|T \setminus T_j} = \sigma$, the lemma holds trivially. Let us assume now that $\sigma_{|T \setminus T_j} \neq \sigma$. If we remove all the transitions belonging to T_j from σ, we are removing constraints to the firing of the rest of transitions. (Because, at marking M, $\mathcal{N}_j$ is in its idle state, that is, using no resource). If we do not allow $\mathcal{N}_j$ to evolve, we have at least the same resources as if it evolved (and maybe more) to fire transitions of σ not belonging to T_j. Therefore $\sigma_{|T \setminus T_j}$ is also firable at marking M. (And t is at $M[\sigma_{|T \setminus T_j}\rangle M''$). □

Let us now prove that if there is a dead transition, the initial marking cannot be reached again.

Lemma 3.17 *Let* $\langle\mathcal{N},M_0\rangle$ *be a marked* ES^3PR, $M\in\mathcal{R}(\mathcal{N},M_0)$ *such that* $t\in T$ *is a dead transition for* M. *Then* $M_0\notin\mathcal{R}(\mathcal{N},M)$.
Proof:

Let σ_t *a firing sequence such that* $M_0[\sigma_t\rangle M$. *Let us suppose that there exists* σ' *such that* $M[\sigma'\rangle M_0$ *then by lemma 3.15* t *is not dead, that contradicts the hypothesis.* □

The following lemmata characterize the relation among partial deadlock (in the sense that a part of the system is blocked but no total deadlock occurs) and siphons in ES^3PR nets. As stated in [ECM95], this property is not true for the general class of Petri nets.

Lemma 3.18 *Let* $\langle\mathcal{N},M_0\rangle$, $\mathcal{N}=\bigcirc_{i\in I_\mathcal{N}}\mathcal{N}_i$, $\mathcal{N}_i=\langle P_i\cup P_i^0\cup P_{R_i},T_i,W_i\rangle$ *be an* ES^3PR. *Let* $M\in\mathcal{R}(\mathcal{N},M_0)$ *and let* $t\in T$ *be a dead transition for* M. *Then, there exists* $M'\in\mathcal{R}(\mathcal{N},M)$ *and two subsets* $I\subseteq I_\mathcal{N}$ *and* $H\subset I_\mathcal{N}$ *such that* $I_\mathcal{N}=I\cup H, I\cap H=\emptyset, I\neq\emptyset$ *and: 1)* $\forall h\in H, M'(p_h^0)=M_0(p_h^0)$*; 2)* $\forall i\in I, M'(p_i^0)<M_0(p_i^0)$ *and* $\{p^\bullet \mid p\in P$ *and* $M'(p)>0\}$ *is a set of dead transitions.*
Proof:

Let us consider $I\subseteq I_\mathcal{N}$, *the set of all indexes so that* $\forall M''\in\mathcal{R}(\mathcal{N},M)$, $M''(p_i^0)<M_0(p_i^0)$ $(M''(p_i^0)>M_0(p_i^0)$ *is not possible because of conservativeness of the class of nets). Note that* $I\neq\emptyset$ *(If* $I=\emptyset$ *then* $M_0\in\mathcal{R}(\mathcal{N},M)$ *and this is not possible because* t *is a dead transition for* M *–lemma 3.17–). We denote* $H=I_\mathcal{N}\setminus I$.

Then, we can ensure that there exists $M_H\in\mathcal{R}(\mathcal{N},M)$ *such that* $\forall h\in H$, $M_H(p_h^0)=M_0(p_h^0)$. *From this marking, and moving only processes corresponding to indexes in* I, *and considering that every* $\mathcal{N}_i$ *is a state machine that contains the place* p_i^0 *the net reaches a marking that verifies condition 2) in hypothesis (if not* $M_0\in\mathcal{R}(\mathcal{N},M)$, *contradiction with lemma 3.17).* □

This lemma shows that, when there exists a dead transition, there is a set of ES^2PR that will never reach the initial marking. Remembering that each ES^2PR represents a type of process, that means that there are some of them that cannot reach the initial state. We can interpret this as a set of deadlocked processes. Now we are going to introduce the definition of extended siphon for composition of ES^3PR systems. If we have a marked ES^3PR and a siphon of it, and we compose it with an ES^2PR, we can extend the siphon to the composed net just adding to the original one the holders of the resources of the siphon in the added ES^2PR, that will be denoted as S_R (following notation used for resource places).

Definition 3.19 (Extended siphon) *Let* $\langle\mathcal{N},M_0\rangle$, $\mathcal{N}=\bigcirc_{i\in I_\mathcal{N}}\mathcal{N}_i$, $\mathcal{N}_i=\langle P_i\cup P_i^0\cup P_{R_i},T_i,W_i\rangle$ *be a marked* ES^3PR*; let* $I\subseteq I_\mathcal{N}$ *and* $H\subset I_\mathcal{N}$ *such that* $I_\mathcal{N}=I\cup H$ *and* $I\cap H=\emptyset$, $I\neq\emptyset$*; let* $\mathcal{N}_I=\bigcirc_{i\in I}\mathcal{N}_i$ *and let* $h\in H$. *Let* S *be a siphon of* $\mathcal{N}_I$. *We define for the composed net* $\mathcal{N}_I\circ\mathcal{N}_h$ *the following set:* $ES(S,h)=S\cup\mathcal{H}(S_R)_h$ □

Of course, the definition can be extended in a natural way to the iterated composition with several ES^2PR.

Note 3.20 *This definition can be extended to sets in the natural way: let* $U\subseteq H$, *then* $ES(S,U)=\bigcup_{u\in U}ES(S,u)$. □

When we are composing a set of ES^3PR this definition provides a way to extend sets of places that are siphons of one of the systems to be composed to sets of places of the whole system. The way to act is to add, for each resource of the siphon, the holders it has in the other systems. The intuitive idea behind this is that, in the composed system, ${}^\bullet S$ has been augmented with transitions from other ES^3PR. Then, to 'close' the siphon, we need to add new places to it. One conceptually simple way to reach it is by adding holders of resources. Note that siphons obtained in this way need not to be minimal. For example, in net of Figure 2–(b), the set $S=\{P2R1,R1\}$ is a siphon that can be extended to a siphon of net of Figure 2–(c) as follows: $S_R=\{R1\}$, $\mathcal{H}(S_R)_1=\{P1R1\}$ and $ES(S,1)=\{P2R1,R1,P1R1\}$. The following propositions can be used to show that the defined set is a siphon of the whole net. First we are going to introduce a proposition that will help us to check if the new defined set is a siphon of the composed net using information about the subnets.

Proposition 3.21 ([TE97]) *Let $\mathcal{N}_i = \langle P_i, T_i, W_i \rangle, i \in \{1,2\}$ be two composable Petri nets and let $\mathcal{N} = \langle P, T, W \rangle = \mathcal{N}_1 \circ \mathcal{N}_2$ be the net obtained by composition of the two nets. Let $S \subseteq P$ and $S_i = S \cap P_i$. S is a siphon of $\mathcal{N}$ iff S_i is a siphon of $\mathcal{N}_i$, $i \in \{1,2\}$.*

□

The following proposition shows that the defined extended siphon is a siphon of the composed net.

Proposition 3.22 ([TE97]) *Let $\langle \mathcal{N}, M_0 \rangle$ $\mathcal{N} = \bigcirc_{i \in I_{\mathcal{N}}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ be a marked ES^3PR; let $I \subseteq I_{\mathcal{N}}$ and $H \subset I_{\mathcal{N}}$ such that $I_{\mathcal{N}} = I \cup H$ and $I \cap H = \emptyset$, $I \neq \emptyset$; let $\mathcal{N}_I$ be the subnet obtained via composition of nets with indexes in I, let S be a siphon of this net. Then $ES(S,H)$ is a siphon of the net $\mathcal{N}$.* □

Once we have shown that, given an ES^3PR and a siphon of it we can augment the system adding new ES^2PR and we are able to construct a siphon for the augmented system, we are going to use this property to demonstrate that, given an ES^3PR with a dead transition, we can reach a marking with an empty siphon.

Theorem 3.23 *Let $\langle \mathcal{N}, M_0 \rangle$ be an ES^3PR. Let $M \in \mathcal{R}(\mathcal{N}, M_0)$ and let $t \in T$ be a dead transition for M. Then there exists a siphon S and $M' \in \mathcal{R}(\mathcal{N}, M)$ so that $M'(S) = 0$.*
Sketch of the Proof:

Let M', I and H be the marking and sets given in lemma 3.18. Let $\mathcal{N}_I$ be the net obtained by composition of subnets $\{\mathcal{N}_i \mid i \in I\}$. From this net we can drop the tokens contained in $\{p_i^0 \mid i \in I\}$. At this point $\mathcal{N}_I$ is a net with a total deadlock and so, there exists S_I siphon of this net that at marking $M'_I = 0$.

Then, using proposition 3.22 we can construct a set S that is a siphon of the net $\mathcal{N}$. This siphon is empty because S_I is and by construction $S \setminus S_I$ are places without tokens. □

4 A control policy for deadlock prevention

Deadlock problems are not desirable in automated systems; in consequence, many research in the field of FMS models has been devoted to this problem. When deadlock situations can be reached in a system, and if one is able to characterize them, it is important to impose restrictions in order the system can avoid them (deadlock prevention/avoidance problem) or it can recover from such undesired situations (deadlock recovery problem). Our attention is going to concentrate on the deadlock prevention/avoidance problem. As we have proved in previous sections, in the class of systems we are dealing with, deadlock problems are characterized by empty siphons. A control policy for deadlock prevention can be easily established: constraint the system evolutions in such a way that no siphon can be emptied.

In order to use the liveness characterization to establish a control policy we mustt take into account that there are three kinds of markings (Figure 3 sketches a reachability graph for an ES^3PR). The first class of markings contains those markings from which M_0 is reachable. These markings are not involved in deadlock problems (the shadowed zone in the figure). The second class is composed of those that contain an empty siphon (and perhaps more than one). They can be easily eliminated controlling the system in such a way that no siphon can be emptied (depicted as black dots outside the shadowed zone). Finally, the third class is composed by the markings that have no empty siphons, but from which a marking with an empty siphon will be reached in an inevitable way (depicted as white dots in the figure). These last markings are the most problematic ones. An optimal policy (optimal here means that it allows as much concurrency as possible maintaining the system liveness) will avoid only markings of type two and type three.

We are going to show how the liveness characterization can be used in order to obtain a live model. Unfortunately, we are going to cut not only markings of type two and three, but also some markings of type one (and then, the proposed control policy is not optimal).

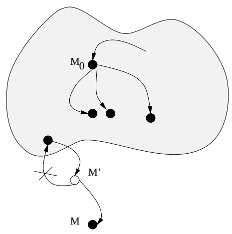


Figure 3: *M' does not empty any siphon but from it the grey zone cannot be reached.*

It is not necessary to control every siphon in the net. If a minimal siphon is the support of a P–semiflow, it cannot become insufficiently marked because we have chosen an acceptable initial marking. As a consequence of this we can state that the only problematic siphons are the minimal siphons that are not the support of any P–semiflow.

To illustrate the following definitions and results we are going to use the very simple ES^3PR depicted in Figure 4–(a). Since idle places do not participate in "dangerous" siphons, they are not depicted in the figure for the shake of simplicity. For the net of Figure 4–(a), the minimal siphons are: $S1 = \{p31, p32, q31, p2, r1\}$, $S2 = \{q31, q32, p4, r1, r2, r3\}$, $S3 = \{q32, p4, q2, r2\}$, $S4 = \{p31, q32, r3\}$ plus the two trivial ones involving the idle place.

4.1 Some definitions

First we are going to define the set of places that are responsible of unmarking of siphons.

Definition 4.1 (Places that stole tokens from a siphon) *Let $\langle \mathcal{N}, M_0\rangle$, $\mathcal{N} = \bigcirc_{i \in I_{\mathcal{N}}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i\rangle$ be a marked ES^3PR; let S be a siphon; we define the set of places that stole tokens from it as $Th_S = \{p \in P \mid \exists r \in S_R$ such that $p \in \mathcal{H}(r)$ and $p \not\in S\}$* □

This set represents the places of the net that use resources of the siphon S but do not belong to it. This places are the ones that contribute to empty the siphon. In the net we are using as example we have: $Th_{S1} = \{p1\}$, $Th_{S2} = \{p1, p2, p31, p32, q1, q2\}$ and $Th_{S3} = \{q1\}$.

The next lemma states that, if there are tokens in the places that stole tokens from a siphon, the siphon has less tokens than it had in the initial marking. The proof is based in the fact that, for places in Th_S, the corresponding resources have less tokens than they had in the initial marking. As these resources are part of the siphon, the lemma holds. □

Lemma 4.2 ([TE97]) *Let $\langle \mathcal{N}, M_0\rangle$, $\mathcal{N} = \bigcirc_{i \in I_{\mathcal{N}}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i\rangle$ be a marked ES^3PR; let S be a siphon and let $M \in \mathcal{R}(\mathcal{N}, M_0)$ such that $M(Th_S) > 0$. Then $M(S) < M_0(S)$*

To avoid undesired states we are going to propose a control policy consisting in locking the entry of processes in zones previous to the ones where problems can arise. Let $\langle \mathcal{N}, M_0\rangle$, $\mathcal{N} = \bigcirc_{i \in I_{\mathcal{N}}} \mathcal{N}_i$, and let S be a siphon of this net. Then we are going to denote $I_S = \{i \in I_{\mathcal{N}} \mid S \cap P_i \neq \emptyset\}$; that is, the set of indexes corresponding to ES^2PR that use resources from S. For example, for nets in Figure 4–(a) and siphon $S1$, $I_{S_1} = \{1, 2\}$(assuming $\mathcal{N}_1$ be the ES^2PR on the left part of the figure and $\mathcal{N}_2$ be the second one). If $\mathcal{N}$ is an ES^3PR, we define the set of dangerous siphons of this net as: $\mathcal{S} = \{S$ minimal siphon $\mid S$ is not the support of any P–semiflow $\}$. The net of Figure 4–(a) has got three minimal P–semiflows: $P1 = \{p31, q32, r3\}$, $P2 = \{p4, q1, q2, q32, r2\}$ and $P3 = \{p1, p2, p31, p32, q31, r1\}$ (plus the two trivial ones involving idle places that are not depicted). In consequence the siphons that are not support of a P–semiflow are $\mathcal{S} = \{S1, S2, S3\}$.

The following definition presents the concept of *need of resources of a place with respect to a siphon* as the sum of weights that the invariants containing resources from the siphon for this place. This weight represents for a resource the number of tokens that have disappeared from the resource place when the place is marked. This number gives us the idea of how many tokens are lost by the siphon if a token reaches the place.

Table 1: PNR values for siphons of the example.

	$NR(p,S1)$	$PNR(p,S1)$	$NR(p,S2)$	$PNR(p,S2)$	$NR(p,S3)$	$PNR(p,S3)$
p1	1	1	1	3	0	0
p2	0	0	2	3	0	0
p31	0	0	2	2	0	0
p32	0	0	3	3	0	0
p4	0	0	0	0	0	0
q1	0	0	1	2	1	1
q2	0	0	2	2	0	0
q31	0	0	0	0	0	0
q32	0	0	0	0	0	0

Definition 4.3 *Let* $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ *be an* ES^3PR*; let* $r \in P_R$ *and let* Y_r *be its associated P–semiflow. Let* $S = S_P \cup S_R$ *be a minimal siphon and let* $p \in P_i$*. The need of resources of* p *with respect to siphon* S *is defined as:* $NR(p,S) =$ *If* $p \in S$ *then* 0 *else* $\sum_{r \in S_R} Y_r(p)$. □

In table 1 we can see the values of this parameter for the net of Figure 4. The next definition presents a subpath of a circuit. It is useful because the part of the circuit considered is the one that a token marking place p must follow to complete the circuit and to reach the idle place.

Definition 4.4 *Let* $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ *be a marked* ES^3PR*; let* X *be a minimal T–semiflow and* F_X *the induced circuit. Let* $p \in \|F_X\| \cap P$*. We define the following path:* $SUC(p, F_X) = pt_p \ldots p_k t_k$*. That is, the sub-path of* F_X *that begins in place* p*.* □

Based in previous definitions we are going to define the *potential need of resources for a place p with respect to a siphon* as the maximum of the need of resources for the places that are in any path from p to the idle state. This number gives us an idea about the maximum quantity of tokens that can be stolen from the siphon if we have a token in p and we want to reach the corresponding idle place.

Definition 4.5 *Let* $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ *be a marked* ES^3PR*; let* $S = S_P \cup S_R$ *be a minimal siphon and let* $p \in P_i$*. The potential need of resources for this place with respect to siphon* S *is defined as:* $PNR(p,S) = \max_{X \in \mathcal{T}^i, q \in SUC(p, F_X) \cap P_i} NR(q,S)$ □

For the example of Figure 4–(a) values for NR and PNR can be seen in Table 1.

Lemma 4.6 *Let* $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i \rangle$ *be a marked* ES^3PR*. Let* X *be a minimal T–semiflow and* F_X *the induced circuit. If* $q \in \|SUC(p, F_X)\|$ *, then* $PNR(q,S) \leq PNR(p,S)$.
Proof:
As $PNR(p,S)$ *is defined as a maximum over paths that begin in* p*, then for each* q *that is in these paths* $PNR(q,S)$ *will be at most like* $PNR(p,S)$*.* □

In order to control the system we are going to add to it a new place for each dangerous siphon, with an initial marking ensuring that no state that leads in an inevitable way to a deadlock can be reached. Once these places have been added, the system evolution is controlled as follows. Let us consider a dangerous siphon S and let us assume that a new process is activated (via firing of a transition $t \in P^{0\bullet}$) such that during the execution of the process a transition belonging to ${}^\bullet Th_S$ can be fired (firing a minimal T–semiflow that contains the transition t in its support). That means that such an execution can decrease the marking of the siphon S. The aim of the added place associated to the siphon is to ensure that the set of active processes cannot empty the siphon.

Therefore, we are going to add to the initial Petri net a place and some arcs for each dangerous siphon. Let P_C be the set of control places we are going to add. This set is related to the set $\mathcal{S}$ by means of a bijection $B : P_C \longrightarrow \mathcal{S}$. In the following definition we are going to introduce the controlled Petri net model corresponding to a given ES^3PR.

Definition 4.7 *Let $\mathcal{N}$ be a marked ES^3PR. The net $\langle \mathcal{N}^C, {M_0}^C\rangle = \langle P \cup P_R \cup P^0 \cup P_C, T, W^C, {M_0}^C\rangle$*
The controlled system $\langle \mathcal{N}^C, {M_0}^C\rangle = \langle P \cup P_R \cup P^0 \cup P^C, T, W^C, {M_0}^C\rangle$ is defined as follows:

1. *P_C is such that there exists a bijection $B : P_C \longrightarrow \mathcal{S}$*
2. *W^C is such that:*
   - $W^C[P \cup P^0 \cup P_R, T] = W$
   - $\forall t \in {P^0}^\bullet$, $\forall p_C \in P_C$, $W^C(p_C, t) = PNR(t^\bullet \cap P, B(p_C))$.
   - $\forall t \in T \setminus {P^0}^\bullet$, $\forall p_C \in P_C$, $W^C(t, p_C) = PNR({}^\bullet t \cap P, B(p_C)) - PNR(t^\bullet \cap P, B(p_C)$.
3. - $M_0^C[P \cup P^0 \cup P_R] = M_0$.
   - $M_0^C(P_C) = M_0(B(p_C)) - 1$. □

Note 4.8 *Notice that M_0^C is well defined. It can be seen as a consequence of good definition of marking for ES^3PR where $M(S) > 0$, $\forall S$ siphon of the net.*

The places that we would need to add to control the system would be the three ones depicted in Figure 4–(b). For example, place $PS2$ has an arc to $tp1$ because for the place $p1$ the potential need of resources is 3. From $tp31$ there is an arc to $PS2$ of weight 1 because in place $p31$, PNR decreases one unit. After that, one more arc of weight 2 from transition $t1p4$ shows that PNR goes to 0. In $t2p4$ the arc weight is 3 and reflects that in this transition the need of resources decreases to 0 here. For the subnet that models processing type 2, the PNR at the beginning is of 2 and we need to put an arc from place $PS2$ to transition $tq1$ with that weight. Two arcs with weights equal to 2 show that tokens enter the siphon $S2$ at transitions $tq31$ and $tq32$ and that no variation in the need of resources from siphon $S2$ exists until these transitions are reached.

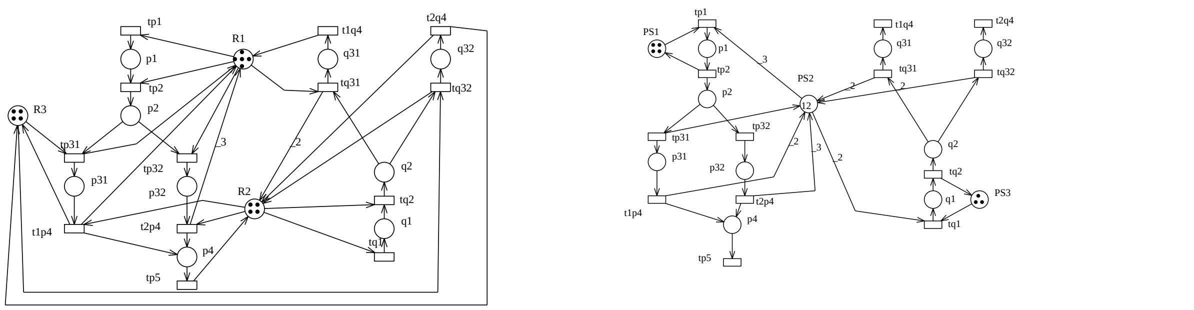


Figure 4: *The original net and the controlled one (without resources).*

Now we are going to prove that this policy is correct in the sense that it makes the system live. The following lemma indicates that T–semiflows can be fired in the new model. The proof is based in: if X is a left annuller of C it is easy to see that it also is an annuller of C^C for all the components belonging to C. For the added places this is true by the way PNR has been constructed and they way it is used to define the control policy.

Lemma 4.9 ([TE97]) *Let $\langle \mathcal{N}, M_0\rangle$, $\mathcal{N} = \bigcirc_{i \in I_\mathcal{N}} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i\rangle$ be a marked ES^3PR; let $X \in \mathcal{T}$. Then $X \in \mathcal{T}^C$. Moreover, the transitions in F_X can be fired sequentially.*
easy to see that it also is an annuller of C^C for all the components belonging to C. For the added places this is true by the way PNR has been constructed and they way it is used to define the control policy. □

The following lemma shows that the firing sequences of the controlled system are also firing sequences for the original one.

Lemma 4.10 *Let $\langle \mathcal{N}^C, M_0^C\rangle$ be the controlled system of the ES^3PR system $\langle \mathcal{N}, M_0\rangle$ and let σ be a firing sequence of $\langle \mathcal{N}^C, M_0^C\rangle$. Then σ is also a firing sequence of $\langle \mathcal{N}, M_0\rangle$.*
Proof:

The only difference between them are the control places. In $\mathcal{N}$ they do not restrict the firing sequences from $\mathcal{N}^C$. $\square$

We are going to show now an invariant relation that control places introduce in the system. This will be used later to show that the policy really controls the system. The proof can be done by induction over the length of the firing sequences. It is based in the way we have controlled the system: control places marking represents the number of tokens that can enter in selected parts of the system.

Lemma 4.11 ([TE97]) *Let $\langle \mathcal{N}, M_0\rangle$ be a marked ES^3PR and let $\langle \mathcal{N}^C, M_0^C\rangle$ be the controlled system as defined in 4.7. Let $p_C \in P_C$. It defines the following invariant relation over reachable markings:*
$\forall M^C \in \mathcal{R}(\mathcal{N}^C, {M_0}^C)$, $\sum_{p\in P} PNR(p, B(p_C)).M^C(p) + M^C(p_C) = M_0^C(p_C)$

The next lemma shows that no siphon can reach the empty marking with this control policy.

Lemma 4.12 *Let $\langle \mathcal{N}, M_0\rangle$ be a marked ES^3PR and let $\langle \mathcal{N}^C, M_0^C\rangle$ be the controlled system as defined in 4.7. Then, $\forall S \in \mathcal{S}$, $M'(S) \neq 0$*
Proof:

Let us suppose that $\exists S \in \mathcal{S}$ such that $M^C(S) = 0$. Then $\sum_{p\in Th_S} M^C(p).NR(p,S) = M_0^C(S) (> 0)$.
$\sum_{p\in Th_S} M^C(p).NR(p,S) \leq \sum_{p\in Th_S} M^C(p).PNR(p,S) \leq \sum_{p\in P} M^C(p).PNR(p,S)$
$\leq \sum_{p\in P} M^C(p).PNR(p,S) + M^C(p_S) = M_0^C(p_S)$ *that contradicts the definition of* $M_0^C(p_S) = M_0^C(S) - 1$ $\square$

Finally, the next proposition joins the previously presented results to prove that the obtained net is live.

Proposition 4.13 *Let $\langle \mathcal{N}, M_0\rangle$, $\mathcal{N} = \bigcirc_{i\in I_N} \mathcal{N}_i$, $\mathcal{N}_i = \langle P_i \cup P_i^0 \cup P_{R_i}, T_i, W_i\rangle$ be a marked ES^3PR. Let $\langle \mathcal{N}^C, M_0^C\rangle$ be the controlled net as defined in 4.7 and $M^C \in \mathcal{R}(\mathcal{N}^C, {M_0}^C)$. Then, $\forall t \in T$, t is not a dead transition for M^C.*
Proof:

Let us assume that $t \in T_i$. The proof is made by induction over the number of tokens in the system that are not in the idle state. Let K_{M^C} be that number. If $K_{M^C} = 0$, then $M^C = M_0^C$, and taking into account that $M_0^C(P_S) \geq W(P_S, t)$, $\forall t \in {P_S}^\bullet$) the proposition holds. Let us assume that $K_{M^C} > 0$; since no siphon from the initial net can be emptied in M^C, there exists $t' \in T \setminus {P^0}^\bullet$ such that $M^C \mid_{\mathcal{N}} [t'\rangle$, and then $M^C[t'\rangle M^{C'}$ (remember that $t' \notin {P^0}^\bullet$ and that implies that $t' \notin {P_C}^\bullet$). We can apply the same reasoning to the token generated by firing of t (if any in $P \setminus P^0$) and iterate the procedure until we reach a marking $M^{C''}$ such that $K''_{M^C} = K_{M^C} - 1$. Then, by induction, the proposition is true. $\square$

Then, we can conclude that the controlled system is live.

5 Conclusions and future work

This work has focused on the study of the liveness property for a class of systems that appears in the modeling of Flexible Manufacturing Systems. The paper extends previous work to a wider class of systems.

The presented approach has some drawbacks: it is not optimal (some good states together with the states leading to deadlocks), the class of nets cannot model the grant of more than one copy of the same resource at a time and, finally, the approach is based on the computation of minimal siphons which can be expensive.

To solve these drawbacks some ideas are being investigated. To obtain better policies two approaches can be studied: the first one is to select the parts of the net that the control places supervise in order to reduce the number of good states forbidden. A second line of work is to cut only states that contain unmarked siphons: in this way some states that in the original net were previous to states with unmarked siphons become states with empty siphons in the controlled net. An iterative solution is a natural way to deal with this kind of problems. In order to extend the class of nets we can concentrate on two main lines: reducing some restrictions in the way the resources can be taken and allowing the processing sequences to be modeled with more general classes of nets different from state machines. Some work presented here can be easily extended to allow weighted arcs from resource places to transitions. In order to do that, the weighted arcs can be simulated by means of ordinary arcs; in this way, some sufficient conditions for the control of the initial net can be obtained. Another promising research is to study 'ad-hoc' techniques to obtain the siphons of the net.

6 Acknowledgments

The authors are grateful to the anonymous referees whose comments have helped us to improve the contents of this paper.

References

[BK90] Z.A. Banaszak and B.H. Krogh. Deadlock avoidance in flexible manufacturing systems with concurrently competing process flows. *IEEE Transactions on Robotics and Automation*, 6(6):724–734, December 1990.

[Com72] F. Commoner. Deadlocks in Petri nets. *Applied Data Research*, 1972.

[DAJ95] A.A. Desrochers and R.Y. Al-Jaar. *Application of Petri Nets in Manufacturing Systems: Modeling, Control and Performance Analysis.* IEEE Press, 1995.

[DHP+93] F. Dicesare, G. Harhalakis, J.M. Proth, M. Silva, and F.B. Vernadat. *Practice of Petri Nets in Manufacturing.* Chapman & Hall, 1993.

[ECM95] J. Ezpeleta, J.M. Colom, and J. Martínez. A Petri net based deadlock prevention policy for flexible manufacturing systems. *IEEE Transactions on Robotics and Automation*, 11(2):173–184, April 1995.

[HC94] Fu-Shiung Hsieh and Shi-Ching Chang. Dispatching-driven deadlock avoidance controller synthesis for flexible manufacturing systems. *IEEE Transactions on Robotics and Automation*, 10(2):196–209, April 1994.

[Jen96] M.D. Jeng. A Petri net synthesis theory for modeling flexible manufacturing systems. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, 27(2):169–183, April 1996.

[Mur89] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, April 1989.

[PX96] J.M. Proth and X. Xie. *Petri Nets. A Tool for Design and Management of Manufaturing Systems.* John Wiley & Sons, 1996.

[SV89] M. Silva and R. Valette. Petri nets and flexible manufacturing. Technical Report GISI-RR-90-06, Dep. Ingeniería Eléctrica e Informática, Universidad de Zaragoza, 1989.

[TE97] F. Tricas and J. Ezpeleta. A partial approach to the problem of deadlocks in processes with resources. Technical Report GISI–RR-97-05, Departamento de Informática e Ingeniería de Sistemas – Universidad de Zaragoza, September 1997.

[TM95] F. Tricas and J. Martínez. An extension of the liveness theory for concurrent sequential processes competing for shared resources. In *Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, pages 4119–4124, Vancouver, British Columbia, Canada, October 1995. IEEE Systems, Man and Cybernetics Society.

[VN92] N. Viswanadham and Y. Narahari. *Performance Modeling of Automated Manfacturig Systems.* Prentice-Hall, 1992.

[XHC96] Ke-Yi Xing, Bao-Sheng Hu, and Hao-Xun Chen. Deadlock avoidance policy for petri-net modeling of flexible manufacturing systems with shared resources. *IEEE Transactions on Automatic Control*, 41(2):289–295, February 1996.