16

Diagnosis of Discrete Event Systems with Petri Nets

Dimitri Lefebvre
GREAH – University Le Havre
France

1. Introduction

Modern technological processes include complex and large scale systems, where faults in a single component have major effects on the availability and performances of the system as a whole. For example manufacturing systems consists of many different machines, robots and transportation tools all of which have to correctly satisfy their purpose in order to ensure and fulfil global objectives. In this context, a failure is any event that changes the behaviour of the system such that it does no longer satisfy its purpose. Failure events lead to fault states (Rausand et al., 2004). Faults can be due to internal events as to external ones, and are often classified into three subclasses : plant faults that change the dynamical input – output properties of the system, sensor faults that result in substantial errors during sensors reading, and actuator faults when the influence of the controller to the plant is disturbed (Blanke et al., 2003).

In order to limit the effects of the faults on the system, diagnosis is used to detect and isolate the failures. Diagnosis is often associated with control reconfiguration, that adapts the controller to the faulty situation such that it continues to satisfy its goal. Fault diagnosis and controller reconfiguration are carried out by supervision systems. This chapter only consider problems related to the diagnosis of systems. Diagnosis includes distinct stages:

1. The fault detection decides whether or not a failure event has occurred. This stage also concerns the determination of the time at which the failure occurs.
2. The fault isolation find the component that is faulty.
3. The fault identification identifies the fault and estimates also its magnitude.

Diagnosis is usually discussed according to the model type used, with component based analysis that uses architectural and structure graph models, with continuous variables systems described by differential or difference equations and transfer functions, with discrete event systems represented by automata or Petri nets and with hybrid dynamical systems that combine continuous and discrete event behaviours (Blanke et al., 2003). Component based methods uses qualitative methods (Rausand et al., 2004) as failure modes and effect analysis (Blanke, 1996) and bi-partite graphs to investigate the redundancies included in the set of constraints and measurements for diagnosis purposes (Cordier et al., 2000; Patton et al., 1999). Fault diagnosis of continuous variables systems is usually based on residual generation and evaluation with parity space approaches or observation,

Source: Petri Net,Theory and Applications, Book edited by: Vedran Kordic, ISBN 978-3-902613-12-7, pp. 534, February 2008, I-Tech Education and Publishing, Vienna, Austria

identification and parameters estimation techniques (Gertler, 1998; Patton et al., 1989). The behaviour of discrete event dynamical systems (DES) is described by sequences of input and output events. In contrast to the continuous systems only abrupt changes of the signal values are considered with DES. In that case, the state of the art is different in comparison with continuous approaches and only few results are available for diagnosis. The problem has been originally investigated with observation methods for automata (Sampath et al., 1995) developed in connection with the supervisory control theory (Ramadge et al., 1987).
This chapter focus on diagnosis of DES modelled with Petri nets (PN) where failures are represented with some particular transitions. The problem is to detect and isolate the firing of the failure transitions in a given firing sequence. The firings of the failure transitions are assumed to be unobserveable and must be estimated according to complete or partial marking measurements that are eventually disturbed by measurement errors. Several problems are related : firing sequences estimation, sensor selection, delay from failure event to detection, and also diagnosers complexity. Let us notice that this study is limited to the methods that represent the faulty behaviours according to the firing of failure transitions and that assume that the state (i.e. the marking vector) of the system is partially or totally measurable. In a alternative way, faults can be also considered as forbidden states. In that case, the observation of the state has been investigated in order to design controllers with forbidden marking specifications (Giua et al., 2002). Asynchronous diagnosis by means of PN unfolding techniques and hidden state history reconstruction obtained from alarm observations was also investigated (Benvenist et al., 2003). These approaches are not developped in this chapter.
The chapter is divided into six sections. Section two states the problem and introduces the notations. Section three is about state space methods that are based on a partial expansion of the reachability graph of the PN under consideration. Section four concerns structural methods that investigate the causality relationships characterized by incidence matrix. Section five is about algebraic methods inspired from coding theory in finite fields of integer numbers. The section six sums up the results and is a tentative of synthesis of the different approaches.

2. Problem statement, motivations and notations

A dynamical system with input u and output y is subject to some faults f. Basically, the diagnosis problem is to find the fault f from a given sequence of input – output couples (U, Y) with:

$$U = (u(0), u(1),\ldots,u(k))$$
$$Y = (y(0), y(1),\ldots,y(k)) \quad (1)$$

where k stands for time $t = k.\Delta t$, and Δt represents the sampling period of sensors. In the next Δt will be omitted and time t will be referred as k as long as there is no ambiguity. It is commonly assumed that no inspection of the process is possible. As a consequence the diagnosis is only based on available measurement data. Moreover the diagnosis problem is usually considered under real time constraints. As long as DES are considered the signals are not real-valued but belong to a discrete value set.

The motivations for the diagnosis of DES is obvious as long as DES occur naturally in the engineering practice. Many actuators like switches, valves and so on, only jump between discrete states. Binary signals are mainly used with numerical systems and logical values "true" and "false" are often used as input and output signals. Alarm sensors that indicate that a physical quantity exceeds a prescribed bound are typical systems with only two logical states. Moreover, in several systems also the internal state is discrete valued. As an example, robot encoders are discrete valued even if the number of discrete state is large enough to produce smooth trajectories. At last, one must keep in mind that a given dynamical system can always be considered as a DES system or as a continuous variable system according to the purpose of the investigation. As long as supervision problems are considered, a rather broad view on the system behaviour can be adopted that is based on discrete signals. On the contrary, if signals have to remain in a narrow tolerance band, the following approaches do no longer fit and one has to adopt a continuous point of view (Blanke et al., 2003).

2.1 Ordinary Petri nets

An ordinary PN with n places and q transitions is defined as < P, T, Pre, Post > where P = $\{P_i\}$ is a non-empty finite set of n places, T = $\{T_j\}$ is a non-empty finite set of q transitions, such that $P \cap T = \varnothing$. Pre: $P \times T \rightarrow \{0, 1\}$ is the pre-incidence application and $W_{PR} = (w^{PR}_{ij}) \in \{0, 1\}^{n \times q}$ with w^{PR}_{ij} = Pre (P_i, T_j) is the pre-incidence matrix. Post: $P \times T \rightarrow \{0, 1\}$ is the post-incidence application and $W_{PO} = (w^{PO}_{ij}) \in \{0, 1\}^{n \times q}$ with w^{PO}_{ij} = Post (P_i, T_j) is the post-incidence matrix. The PN incidence matrix W is defined as $W = W_{PO} - W_{PR} \in Z_3^{n \times q}$ with $Z_3 \in \{-1, 0, 1\}$ and w_i stands for the i^{th} column of W (Askin et al., 1993; Cassandras et al., 1999; David et al., 1992). $M = (m_i) \in (Z^+)^n$ is defined as the marking vector and $M_I \in (Z^+)^n$ as the initial marking vector, with Z^+ the set of non negative integer numbers. A firing sequence $\sigma = T_i.T_j\ldots T_k$ is defined as an ordered series of transitions that are successively fired from marking M to marking M' (i.e. M [σ > M') such that equation (2) is satisfied:

$$\sigma : M \xrightarrow{T_i} M_1 \xrightarrow{T_j} M_2 \rightarrow \cdots \xrightarrow{T_k} M' \tag{2}$$

A sequence σ can be represented by its characteristic vector (i.e. Parikh vector) $X = (x_j) \in (Z^+)^q$ where x_j stands for the number of times T_j has occurred in sequence σ (David et al., 1992). Marking M' resulting from marking M with the execution of sequence σ is given by (3):

$$\Delta M = M' - M = W.X \tag{3}$$

The reachability graph R(PN, M_I) is the set of markings M such that a firing sequence σ exists from M_I to M. A sequence σ is said to be executable for marking M_I if there exists a couple of markings (M, M') $\in$ R(PN, M_I) such that M [σ > M'.

2.2 Problem statement and notations

The objective of diagnosis problem is to identify the occurrence and type of failure events, based on observable traces generated by the system. For this purpose, let us define $\Delta_F = \{F_k\}$ the set of K distinct faults that may affect the system. A label $L \in \Delta = \{N\} \cup \Delta_F$ is associated

to each transition. As a consequence $T = T_F \cup T_N$ with T_F the set of "failure" transitions and T_N the set of "normal" transitions. The firing of transitions is usually unobservable. L = N is interpreted as a "normal" behavior, and $L = F_k$ means that fault F_k has occurred. Starting from an initial state, the system may evolve according to a "normal" behavior by firing "normal" transitions or according to a faulty behavior by firing a sequence with one or several "failure" transitions.

Let us define $\theta = \{\theta_k\} \subset T^b$ be a list of b groups of fault transitions $\theta_k \subset T$ (or eventually single failure transitions). We define $B(\theta) = (b_{kj}) \in \{0, 1\}^{b \times q}$ such that $b_{kj} = 1$ if $T_j \in \theta_k$, else $b_{kj} = 0$. Let us also consider $X_\theta = B(\theta).X \in (Z^+)^b$ the firing vector to be estimated. In other words, the k^{th} row of matrix $B(\theta)$ characterizes θ_k, and the sum of firing occurrences in the k^{th} subset of transitions (i.e. the k^{th} entry of $X(\theta)$) has to be estimated from the measurement of the observable markings. To define a list θ of transitions subsets is interesting in case of non discernable faults. When the faults $\{F_k\}_{k = 1,...K}$ must be detected and located, then the list $\theta = \{\{T_{F1}\},\ldots,\{T_{FK}\}\}$ with K singletons $\{T_{F1}\},\ldots,\{T_{FK}\}$ is used. When the faults $\{F_k\}_{k = 1,...K}$ must be detected but not isolated (i.e non discernable faults) $\theta = \{T_{F1},\ldots,T_{FK}\}$ with a single subset $\{T_{F1},\ldots,T_{FK}\}$ is defined.

The set P is also divided into the set $P_O = \{P'_i\}$ of c observable places and the set P_U of n - c unobservable ones: $P = P_O \cup P_U$. Vector $M_O \in (Z^+)^c$ is defined as $M_O = C(P_O).M$ with $C(P_O) = (c_{ij}) \in \{0, 1\}^{c \times n}$, such that $c_{ij} = 1$ if $P_j \in P_O$ and $P_j = P'_i$, else $c_{ij} = 0$. Only the marking M_O of the observable places is assumed to be measured. Let us also define $W_O = C(P_O).W \in (Z_3)^{c \times q}$, $w_O(j)$ as the j^{th} column of matrix W_O, and ΔM_O according to (4):

$$\Delta M_O = C(P_O).W.X = W_O.X \tag{4}$$

Petri nets are asynchronous models. As a consequence, two distinct transitions are never simultaneously fired and the following basic assumption can be considered: there always exists a marking measurement between two consecutive firings in a given firing sequence.

The preceding hypothesis is necessary because the firing of a transition will be undetectable if it does not have any observable influence on the marking variation. For example, the marking of the cycle $\{P_2, T_3, P_3, T_4\}$ in PN1 (figure 1) is not modified if there is no intermediate observation for the sequence of firings $\sigma = T_3.T_4$. Moreover the marking of a given place is not modified if a transition in the preset and another one in the post – set are both fired between two consecutive observations. For example, the marking of place P_1 in PN1 remains unchanged after the execution of sequence $\sigma = T_2.T_1$. According to the preceding hypothesis, the firing sequences that are considered in the following can always be separated into sub-sequences of size $1 : X \in \{0, 1\}^q$, and $||X|| \leq 1$.

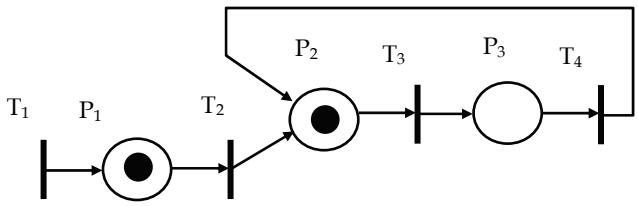


Fig. 1. Example PN1 of Petri net with cycles

3. State space methods for the diagnosis of DES

3.1 Partial expansion of reachability graph and indetermined cycles

Fault diagnosis based on state space approach and on partial expansion of the reachability graph was first formulated with automata (Sampath et al., 1995). Sampath et al. introduce the study of indeterminate cycles in automata and state that a langage is diagnosable if and only if the diagnoser satisfies the following condition : there is no F_k - failure indeterminate cycle for all failure types.

The investigation of indeterminate cycles was then extended to PN with finite reachability graph (Ushio et al., 1998). The considered PN are live (i.e. for any $T_j \in T$, and for all $M \in R(PN, M_I)$ there exists a sequence σ executable from M that includes transition T_j) and safe (i.e. for all $M \in R(PN, M_I)$, $M \in \{0, 1\}^n$) with some places that are observable and other not. Transitions are usually assumed to be unobservable. The diagnosability of the system is based on the study of indetermined cycles included in the observable part of the labelled reachability graph $R(PN, T_F, M_I, P_O)$ (Ushio et al., 1998). A cycle is called "determined" if it contains at least one observable state that results with no ambiguity from a normal firing sequence, or from a F_k - failure firing sequence (i.e. a firing sequence that contains a F_k - failure transition). Characterisation of the cycles is obtained according to label propagation and range functions that tell us how to assign the fault labels and how to estimate all the next possibly diagnoser states from an initial state. Starting from an observable initial marking, the diagnoser detects and isolates a failure transition in a given firing sequence from measurement of the successive observable states visited by the system.

The notion of diagnosability is defined as the inherent property of the system that when a failure occurred, we can always infer its type, no matter how the system evolves after the failure. The resulting diagnosers are "delayed" (i.e. multi-steps diagnoser) in the sense that the occurrence of intermediate events may be necessary to detect and isolate the faults. The number of intermediate events is upper bounded according to the maximal size of the determined cycles. In (Chung et al., 2003) some transitions are assumed to be observable in order to increase the database used by the diagnoser. An algorithm, based on linear programming, of polynomial complexity in the worst case for computing a sufficient condition of diagnosability has been also proposed (Wen et al., 2005).

Let us consider the Petri net named PN2 in figure 2 as an example. All transitions are supposed to be unobservable. The transition T_1 represents a failure event F. Other transitions are assumed to represent normal events.

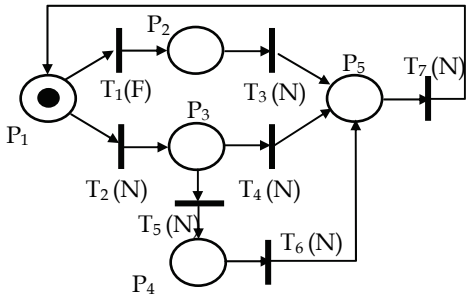


Fig. 2. Example PN2 of Petri net

If the set of observable places is given by $P_{O1} = \{P_1, P_4, P_5\}$, the observable part of the labelled reachability graph R(PN2, $\{T_1\}$, $(1, 0, 0, 0, 0)^T$, P_{O1}) is worked out as in figure 3a. This diagnoser has an indetermined cycle so the system is not diagnosable (figure 3a, on the left). If $P_{O2} = \{P_1, P_3\}$, the observable part of the labelled reachability graph R(PN2, $\{T_1\}$, $(1, 0, 0, 0, 0)^T$, P_{O2}) is worked out as in figure 3b. This diagnoser has no indetermined cycle so the system is diagnosable.

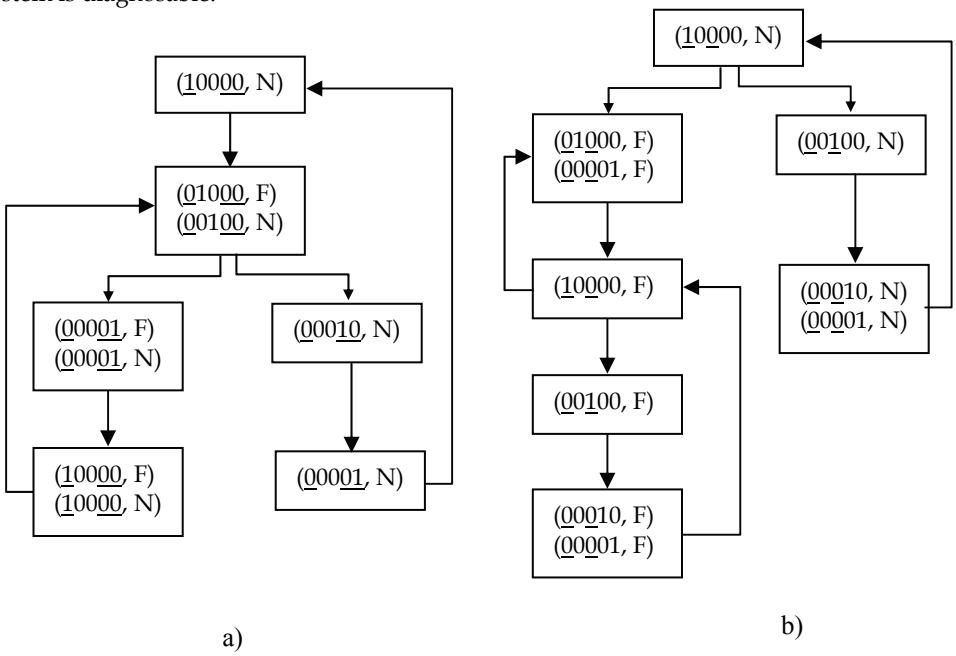


Fig. 3. Two partial expansions of the reachability graph for PN2

a) R(PN2, $\{T_1\}$, $(1, 0, 0, 0, 0)^T$, P_{O1}) ; b) R(PN2, $\{T_1\}$, $(1, 0, 0, 0, 0)^T$, P_{O2})

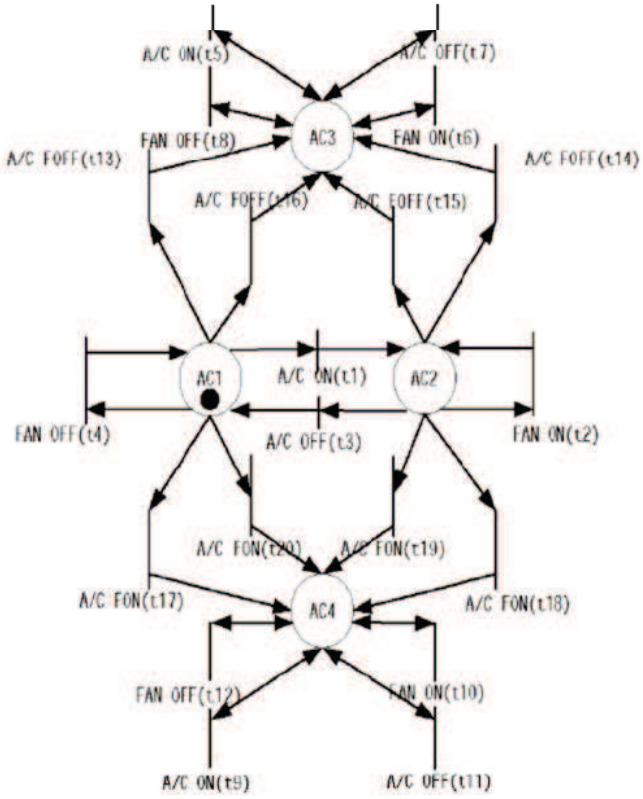
As a conclusion, let us notice that the preceding method is efficient to evaluate the diagnosability of a system but not suitable to design diagnosers. The reason is that the partial expansion of the reachability graph must be worked out for all diagnoser candidates. Such a computation is time consuming so that it cannot be adapted for sensor selection problems in case of large scale systems.

3.2 Application

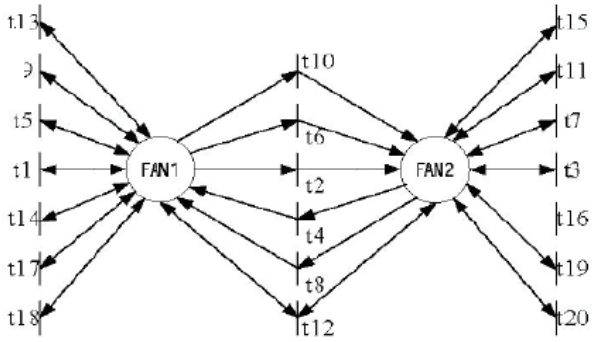
State space method have been used to state the diagnosability of an automatic temperature control system (ATC) for automobile applications (Wen et al., 2005). The PN models of ATC has 3 components (figure 4a-b-c):

a) The pump model has four unobservable states. The places ACI and AC2 stand for pump off and pump on respectively. The places AC3 and AC4 stand for pump failed off and pump failed on respectively.
b) The fan model has two unobservable states : FAN1 and FAN2 stand for fan off and fan on respectively.
c) The controller has four observable states and four events. The state Cl represents both the pump and fan are off. State C2 represents that the pump turns on first, while the fan

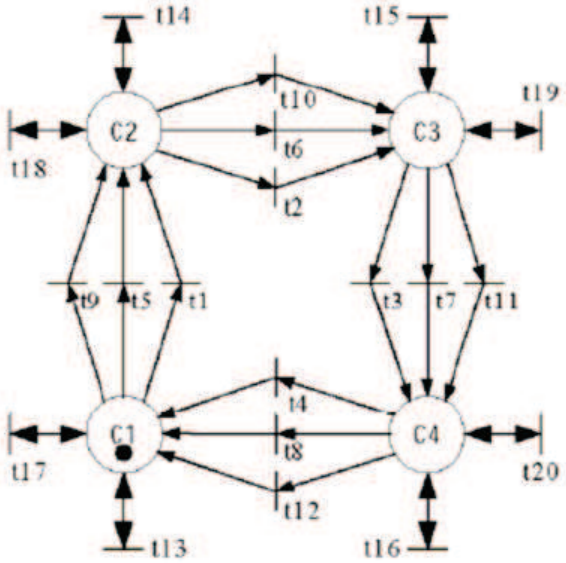
is in off. State C3 represents that the pump turns on, and the fan turns on. State C4 represents that the pump turns off first, while the fan is still working.



a) Pump



b) Fan



c) Controller

Fig. 4. PN3 model of an automatic temperature control system (Wen et al., 2005)

Transition	Event Type (Fail type)	Sensor Map
t_1	A/C ON (N)	H to H
t_2	Fan ON (N)	H to L
t_3	A/C OFF (N)	L to H
t_4	Fan OFF (N)	H to H
t_5	A/C ON (F1)	H to H
t_6	Fan ON (F1)	H to H
t_7	A/C OFF (F1)	H to H
t_8	Fan OFF (F1)	H to H
t_9	A/C ON (F2)	H to H
t_{10}	Fan ON (F2)	H to L
t_{11}	A/C OFF (F2)	L to L
t_{12}	Fan OFF (F2)	L to H
t_{13}	A/C FOFF (F1)	H to H
t_{14}	A/C FOFF (F1)	H to H
t_{15}	A/C FOFF (F1)	L to H
t_{16}	A/C FOFF (F1)	H to H
t_{17}	A/C FON (F2)	H to H
t_{18}	A/C FON (F2)	H to H
t_{19}	A/C FON (F2)	L to L
t_{20}	A/C FON (F2)	H to L

Table 1. Transitions and sensor map of the ATC (Wen et al., 2005)

There are two failure types. Failure types F1 and F2 stand for pump fails off and pump fails on respectively. It is assumed that the system has one temperature sensor. The set of outputs is L = {low} and H={high} according to the temperature in the cabin of the vehicle. The meaning of the transitions and sensor map are listed in table 1. For example, S (H to H) means that the reading of the cabin sensor changes from High to High. The study of the indetermined cycles in observable part of reachability graph and the investigation of the transitions (events) with the same observable projection (for example T_1, T_5, and T_9 represent the same observable projection $\{e_l\}$ where $\{e_l\}$ depicts that the controller state is "pump on" and it's sensor reading changes from High to High) is useful to state that this system is diagnosable.

4. Diagnosis based on structural approaches

4.1 Event detectability

Another diagnosis approach for DES has been developed according to event detectability of interpreted PN (Alcaraz-Mejia et al., 2003; Ramirez-Trevino et al., 2004). An interpreted PN is event detectable when any pair of transitions can be distinguished from each other by the observation of the input - output symbols of the interpreted PN (inputs are defined according to the events associated with the transitions and outputs are defined according to the measurements of the observable markings). Preliminary results have been obtained according to the additive independence of columns of the output matrix (Ichikawa et al., 1988). A characterization of event detectability has been established as a consequence, when all columns of matrix $W_O = C(P_O).W \in (Z_3)^{c \times q}$ are not zero and different from each other (Alcaraz-Mejia et al., 2003). Input - output diagnosability in finite number of steps has been derived as a consequence. An interpreted PN is input - output diagnosable in r steps if any marking M resulting immediately from the firing of a fault transition is distinguishable from any other marking M′ by firing any sequence with r transitions (Alcaraz-Mejia et al., 2003; Ramirez-Trevino et al., 2004). Several structural characterizations of input - output diagnosability have been provided: necessary and sufficient conditions related to input - output relationships between places, sufficient conditions when the normal behaviour of the interpreted PN is event detectable (Alcaraz-Mejia et al., 2003; Ramirez-Trevino et al., 2004; Ramirez-Trevino et al., 2007).

In order to illustrate event diagnosability, let us consider again PN2 in figure 2. On one hand, if the set of observable places is given by $P_{O1} = \{P_1, P_4, P_5\}$, event detectability is worked out according to matrix W_{O1}:

$$C(P_{O1}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad W_{O1} = C(P_{O1}).W = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & -1 \end{pmatrix} \tag{5}$$

System (5) is not event detectable because columns 1 and 2, and also columns 3 and 4 of matrix W_{O1} are identical.

On the other hand, if the set of observable places is given by $P_{O2} = \{P_1, P_3\}$, event detectability is worked out according to matrix W_{O2}:

$$C(P_{O2}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad W_{O2} = C(P_{O2}).W = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \end{pmatrix} \tag{6}$$

However system (6) is diagnosable for fault F_1, it is not event detectable because columns 3 and 6 are zero and columns 4 and 5 are identical. As a consequence, input - output diagnosers cannot be derived.

In (Aramburo-Lizarraga et al., 2005) the condition of event detectability is relaxed over parts of the model where the faults are not expected; thus the diagnoser handles a reduced model. Moreover, a method for splitting the global model into communicating modules is proposed that leads to the design of a set of distributed diagnosers. A framework concerning DES diagnosis based on PN and event detectability approach can also be found in (Ramirez-Trevino et al., 2007) where the authors introduce a bottom-up modelling methodology that avoids tuning phases and state combinatory found in finite state automata approaches.

4.2 Minimal sets of observable places for single step diagnosis

Fault diagnosis is strongly related to the problem of sensor selection that leads to the determination of minimal sets (for inclusion) of observable places in order to detect and identify the firing of some particular "failure" transitions. In this context, places are assumed to have a physical meaning so that direct relationships exist between places, state variables and sensors. The problem is to decide the number and location of the places to be observed (i.e. the state variables to be measured) in order to estimate the firings of some transitions (i.e. to detect and isolate some faults). Such sets of places are named "minimal sets of observable places" (Lefebvre 2004; Lefebvre et al., 2007).The problem that is solved is to give necessary and sufficient conditions in order to decide if the unbiased observation of the marking variation for a set of places P_O leads to immediate estimation of $X(\theta)$.

The subset of places $P_O \subset P$ is called a set of observable places (SOP) for θ, if $X(\theta)$ can be estimated exactly (i.e. with no error) and immediately (i.e. with no delay) from the unbiased measurement of ΔM_O between two consecutive observations. The subset of places $P_O \subset P$ is called a minimal set of observable places (MSOP) for θ, if P_O is a SOP for θ, and if there is no subset of places $P' \subset P_O$, $P' \neq P_O$ that is also a SOP for θ.

A SOP for θ provides enough information to detect and isolate a firing in θ before the occurrence of any other event and a MSOP is a minimal SOP for inclusion. According to basic assumption in section 2.b, P_O is a SOP for θ means that for any vector $X \in \{0, 1\}^q$ such that $||X|| \leq 1$, the unbiased measurement of $\Delta M_O = C(P_O).W.X \in (Z_3)^c$ leads to immediate and exact estimation of vector $X(\theta) = B(\theta).X \in \{0, 1\}^b$.

Characterisations of SOP can be obtained with an enumeration of the partitions for P_O or equivalently with the columns of the observable part W_O of incidence matrix (Lefebvre 2006, Lefebvre et al., 2007). For any marking variation ΔM_O let us define the disjoint partition of set P_O as $PA(\Delta M_O) = (P^+(\Delta M_O), P^-(\Delta M_O), P^0(\Delta M_O))$ with $P^+(\Delta M_O) = \{P_i\} \subset P_O$ such that $\Delta m_i > 0$, $P^-(\Delta M_O) = \{P_i\} \subset P_O$ such that $\Delta m_i < 0$ and $P^0(\Delta M_O) = \{P_i\} \subset P_O$ such that $\Delta m_i = 0$. Let us also consider the set of transitions $E(PA(\Delta M_O)) \subset T$:

$$E(PA(\Delta M_O)) = \left(\bigcap_{P_i \in P^+(\Delta M_O)} {}^\circ P_i \right) \cap \left(\bigcap_{P_i \in P^+(\Delta M_O)} P_i{}^\circ \right) \cap \left(\overline{\bigcup_{P_i \in P^0(\Delta M_O)} {}^\circ P_i \cup P_i{}^\circ} \right) \tag{7}$$

where $^{\circ}P_i$ stands for the set of P_i - upstream transitions and P_i° stands for the set of P_i - downstream transitions. The subset $P_O \subset P$ is a SOP for θ if and only if characterisation 1 or equivalently 2 is satisfied (Lefebvre et al., 2007):

1. For each subset $\theta_k \subset T$, $k = 1,\ldots,b$, there exist a list of r_k disjoint partitions $PA_O(i) = (P^+_O(i), P^-_O(i), P^0_O(i))$ of P_O, $i = 1,\ldots, r_k$, such that $P^+_O(i) \cup P^-_O(i) \neq \varnothing$ and :

$$\bigcup_{i=1,..,r_k} E(PA_O(i)) = \theta_k$$

2. For each subset $\theta_k \subset T$, $k = 1,\ldots,b$, and for any couple of transitions $T_\alpha \in \theta_k$, $T_\beta \notin \theta_k$ we have $w_O(\alpha) \neq 0$ and $w_O(\alpha) \neq w_O(\beta)$.

The preceding result leads to an algorithm of complexity q^2 that generates the exhaustive list $F(P_O)$ of groups of transitions θ_k with minimal cardinality for which P_O is a SOP. The reduction of the obtained list thanks to linear algebra can be obtained as a post processing (Lefebvre 2004).

Algorithm a

1. Initialise list F to be empty
2. While T is not empty do
3. Initialise subset θ_k to be empty
4. Select $T_j \in T$
5. Remove T_j from set T
6. If $w_O(j) \neq 0$, then
7. Add T_j to subset θ_k
8. For any $T_i \in T$, do
9. If $w_O(j) = w_O(i)$, then transition T_j is added to set θ_k and T_i is removed from set T
10. End for (step 8)
11. Add subset θ_k to the list F
12. End if (step 6)
13. End while (step 2)

A recursive algorithm based on a combinatory exploration of the PN subsets of places generates also the list $G(\theta_k)$ of all MSOP for θ_k. From a computational point of view, this non polynomial algorithm must be used with some precautions. But the complexity depends on the number of potential observable places, and not on the size of the whole PN. Thus, it is suitable even for large scale systems as long as the considered set of potential observable places remains small. In comparison with algorithms that partially expand the reachability graph, the complexity of our results does not depend on the size of that graph.

Let us consider again PN2 with $\theta_1 = \{T_1\}$. Applying the preceding characterisation (condition 1 or 2), it is easy to state that P_{O1} is not a SOP for θ_1, whereas P_{O2} is a SOP and also a MSOP for θ_1. Moreover, this characterization leads to the exhaustive list of MSOP for θ_1 : $G(\theta_1) = \{\{P_2\}, \{P_1, P_3\}\}$. It leads also to the exhaustive list of transitions for which P_{O1} is a SOP : $F(P_{O1}) = \{\{T_5\}, \{T_6\}, \{T_7\}, \{T_1, T_2\}, \{T_3, T_4\}\}$ and to the exhaustive list of transitions for which P_{O2} is a SOP : $F(P_{O2}) = \{\{T_1\}, \{T_2\}, \{T_7\}, \{T_4, T_5\}\}$. As a consequence, $\{P_2\}$ and $\{P_1, P_3\}$ are the two possible MSOP for single – step diagnosis.

4.3 Diagnosis with CR and DP

Causality relationships (CR) and directed paths (DP) in PN models (Lefebvre et al., 2005) can also be used for multi-steps diagnosis purposes. In that case, diagnosis is improved by considering that some transitions may be observable. For that purpose, the set T_N is divided into a set T_O of observable transitions and a set T_U of unobservable ones.

Let N and N' be two nodes (i.e. places or transitions) of PN model. A CR exists from N' to N if and only if the behaviour of the node N' could affect the variable attached to node N. The CR size (referred as CR – rank in the following) can be understood as the number of places in the shortest causality relationship from transition T_k to place P_i or transition T_j, and as the number of transitions in the shortest causality relationship from place P_k to place P_i or transition T_j. When no causality relationship exists, the CR - rank equals infinity. The CR – rank of PN nodes in range $I = [r_{min}, r_{max}] \cup \{\infty\}$ is characterised by the matrix CR(I) as given in (8) (Lefebvre et al., 2005):

$$CR(I) = \begin{pmatrix} CR_{PP}(I) & CR_{PT}(I) \\ CR_{TP}(I) & CR_{TT}(I) \end{pmatrix} \in I^{(n+q)x(n+q)} \tag{8}$$

with $CR_{PP}(I) = CR_{PP}(P_i, P_k, I) \in I^{n \times n}$, $CR_{PT}(I) = CR_{PT}(P_i, T_k, I) \in I^{n \times q}$, $CR_{TP}(I) = CR_{TP}(T_j, P_k, I) \in I^{q \times n}$, $CR_{TT}(I) = CR_{TT}(T_j, T_k, I) \in I^{q \times q}$.

Similarly, a DP exists from N' to N if and only if a token is able to move from N' to N. A DP between two nodes is also a CR but a CR is not necessary a DP. The DP - rank of PN nodes in range $I = [r_{min}, r_{max}] \cup \{\infty\}$ is characterised by a matrix $DP(I) \in I^{(n+q) \times (n+q)}$ similar to CR(I) (Lefebvre et al., 2005). From a computational point of view, the determination of the CR and DP matrices results from polynomial algorithms of complexity $(r_{max} - r_{min}).n.q$. The CR and DP ranks are defined according to the table 2.

M(A,r)	$A=W_{PR}+W_{PO}$	$A=W_{PR}$
$(A.(W_{PR})^T)^r$	$CR_{PP}(P_i, P_k, I)$	$DP_{PP}(P_i, P_k, I)$
$(A.(W_{PR})^T)^r.A$	$CR_{PT}(P_i, T_k, I)$	$DP_{PT}(P_i, T_k, I)$
$(W_{PR})^T.(A.(W_{PR})^T)^r$	$CR_{TP}(T_j, P_k, I)$	$DP_{TP}(T_j, P_k, I)$
$((W_{PR})^T.A)^r$	$CR_{TT}(T_j, T_k, I)$	$DP_{TT}(T_j, T_k, I)$

Table 2. CR and DP characterisation (Lefebvre et al., 2005)

In the next, the set I will be omitted as long as $I = [0, \min(n, q)] \cup \{\infty\}$ because CR and DP ranks cannot exceed the number of places or transitions.

In order to evaluate the potential of a set of observable nodes $P_O \cup T_O$ for diagnosis purpose, let us define the influence areas $I_{CR}(T_k)$ and $I_{DP}(T_k)$ of failure transition T_k, and dependence areas $D_{CR}(N)$ and $D_{DP}(N)$ of node N. The set $I_{CR}(T_k)$ of nodes that are CR - sensitive with respect to the transition T_k is called the CR - influence area of T_k. This area is a subnet of PN defined as $I_{CR}(T_k) = <P_{ICR}(T_k), T_{ICR}(T_k), Pre_{ICR}(T_k), Post_{ICR}(T_k)>$ where $P_{ICR}(T_k) \subset P$ is the set of places P_i such that $CR_{PT}(P_i, T_k) < \infty$. $T_{ICR}(T_k) \subset T$ is the set of transitions T_j such that $CR_{TT}(T_j, T_k) < \infty$, $Pre_{ICR}(T_k)$ and $Post_{ICR}(T_k)$ are the restrictions of the pre - incidence and post – incidence applications limited to the sets $P_{ICR}(T_k)$ and $T_{ICR}(T_k)$. The DP - influence area $I_{DP}(T_k)$ is defined in a similar way. The CR - dependence area $D_{CR}(N)$ of the node N is also a

subnet of PN defined as $D_{CR}(N) = <P_{DCR}(N), T_{DCR}(N), Pre_{DCR}(N), Post_{DCR}(N)>$ where $T_{DCR}(N)$ and $P_{DCR}(N)$ are the sets of transitions and places that are likely to influence the node N through a causality relationship. The DP - dependence area $D_{DP}(N)$ is defined in a similar way. The characterisation of the sets $P_{ICR}(T_k)$, $T_{ICR}(T_k)$, $P_{IDP}(T_k)$, $T_{IDP}(T_k)$, $T_{DCR}(P_i)$, $T_{DCR}(T_j)$ $T_{DDP}(P_i)$, and $T_{DDP}(T_j)$ is given in table 3 according to the position of finite entries in columns or rows of CR and DP matrices.

	CR		DP	
$P_{I..}(T_k)$	CR_{PT}	k^{th} column	DP_{PT}	k^{th} column
$T_{I..}(T_k)$	CR_{TT}	k^{th} column	DP_{TT}	k^{th} column
$T_{D..}(P_i)$	CR_{PT}	i^{th} row	DP_{PT}	i^{th} row
$T_{D..}(T_j)$	CR_{TT}	j^{th} row	DP_{TT}	j^{th} row

Table 3. Influence and dependence areas (Lefebvre et al., 2005)

The CR and DP investigation is helpful for delayed diagnosis of systems modelled by PN, in the sense that it provides in a systematic way the relationships between a fault transition and other nodes of PN.

1. Let $N \in P_O \cup T_O$. A necessary condition such that the observation of node N contributes to the diagnosis of F_k is $N \in I_{CR}(T_k)$ (Lefebvre et al., 2005).
2. Let $N \in P_O \cup T_O$. A sufficient condition to detect and isolate the firing of the fault transition T_k with the observation of node N is $N \in I_{DP}(T_k)$ and $T_{DDP(PN/Tk)}(N) = \varnothing$ if N is a place or $T_{DDP(PN/Tk)}(N) = \{N\}$ if N is a transition in PN/T_k (i.e. PN where the transition T_k has been removed) (Lefebvre et al., 2005).

If the preceding propositions cannot be applied, the nodes that have to be observed at first are the ones with the smaller dependence areas including fault transition T_k. This choice consists to select sensors in order to be sensitive with respect to the smaller set of events.

4.4 Application

PN can be used to model and monitor batch or chemical processes, like the system represented in figure 5a (Lefebvre et al., 2007). This system is composed of a tank R that can be filled and emptyed according to the flows Q_{source} provided by the source and Q_{demand} required by the distribution network. The system has three logical actuators: the input valves V_1 and V_2 and the output valve V_3 with two states {open = 1, closed = 0}. The continuous state variable h corresponds to the tank level and is defined according to $S.dh/dt = D - A.(2.g.h)^{1/2}$ with S the tank section, A the output pipe section and g the gravity acceleration.

The goal of the PN supervisor PN4 is to keep the level h below the treshold LSH+ and above the treshold LSH- in order to limit the pressure in distribution network. When LSH- is reached V_1 is opened during an appropriate time to fill the tank. Then V_1 is closed. Eventually V_2 is closed and V_3 is opened if LSHH is reached. Two logical level sensors are used to detect the tresholds LSH- and LSHH.

$$CR_{PT}(PN4)=\begin{array}{c} \begin{matrix} T_1 & T_2 & T_3 & T_4 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \end{array} \begin{matrix} P_1 \\ P_2 \\ P_3 \end{matrix}, \qquad DP_{PT}(PN4/T_1)=\begin{array}{c} \begin{matrix} T_2 & T_3 & T_4 \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 \\ \infty & \infty & \infty \\ \infty & 0 & \infty \end{pmatrix} \end{array} \begin{matrix} P_1 \\ P_2 \\ P_3 \end{matrix} \tag{9}$$

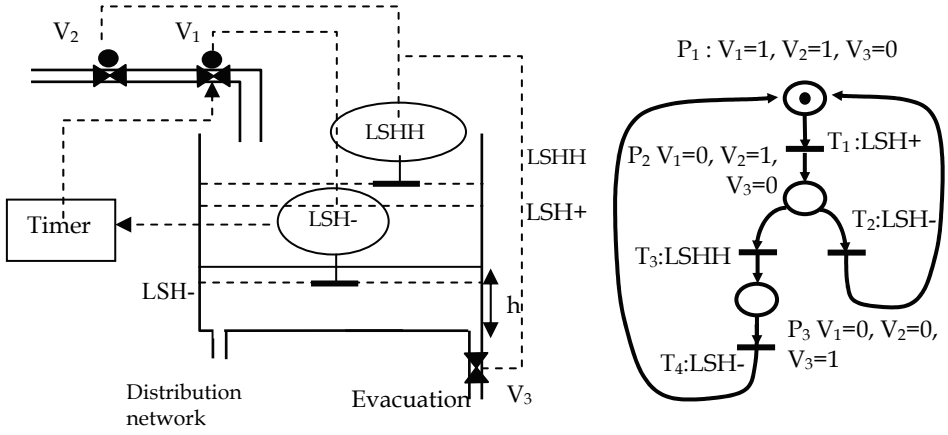


Fig. 5. Tank system a) Sens a) l actuators; b) PN4 model of the controller b)

The set of observable places is assumed to be defined as $P_O = \{P_3\}$. A single fault is considered when the treshold LSH+ is exceeded. The MSOP for LSH+, included in P, are given by $G(\{LSH+\}) = \{\{P_1\}, \{P_2\}\}$. The resulting MSOP are not suitable because no sensor is used to detect the threshold LSH+. Matrix CR_{PT} shows that $P_{ICR}(T_1) = \{P_1, P_2, P_3\}$: the observation of each place contributes to the diagnosability of PN4. Matrix DP_{PT} in $PN4/T_1$, shows that $T_{DDP(PN4/T1)}(P_2) = \varnothing$ (but P_2 is not observable) and $T_{DDP(PN4/T1)}(P_3) = \{T_3\}$. As a conclusion, the observation of P_3 can be used as a two-steps diagnoser to detect a fault of sensor LSH+.

5. Diagnosis based on algebraic approaches

5.1 Diagnosis based on coding theory

Event sequences estimation is an important issue for fault diagnosis of DES, so far as fault events cannot be directly measured. This section is about event sequences estimation with PN models. Events are assumed to be represented with transitions and firing sequences are estimated from measurements of the marking variation. Estimation with and without measurement errors can be discussed in n – dimensional vector space over alphabet Z_3 (Lefebvre, 2006; Lefebvre, 2007). The basic idea to correct measurement errors by projecting measurements in orthogonal subspace of Vect(W) where Vect(W) stands for the subspace generated by the columns of W. This method is inspired from linear coding theory (Van Lint, 1999) and extend the results presented for continuous PN in (Lefebvre et al., 2001).

Measurement $\Delta\hat{M}$ of marking variation $\Delta M \in (Z_3)^n$ may be affected by an additive error vector $E \in (Z_3)^n$: $\Delta\hat{M} = \Delta M + E$. The error vector will be characterized according to the Hamming distance d(W) of the considered PN that is defined with the Hamming distance of the columns of incidence matrix :

$$d(W) = \min\{\min\{d(w_i, w_j), i \neq j\}, \min\{d_0(w_i)\}\} \quad (10)$$

where $d(w_i, w_j)$ stands for the Hamming distance between columns w_i and w_j of matrix W and $d_0(w_i) = d(w_i, 0)$ stands for the weight of vector w_i.

It is assumed that error vector E verifies the following conditions:

1. $Pr(d_0(E) = 0) > Pr(d_0(E) = 1) > ... > Pr(d_0(E) = n)$ where $Pr(d_0(E) = i)$ is the probability that weight of E equals i;
2. An error in position i does not influence other positions;
3. A symbol in error can be each of the remaining symbols with equal probability.

A short estimation algorithm easy to use and to implement when state measurement is complete (i.e. all entries of $\Delta\hat{M}$ are measured), and error free (i.e. measurement equals actual marking variation ΔM), is based on the comparison of the measurement with respect to columns of W and zero vector (this corresponds to the condition of event-detectability in case that all places are observable). When this measurement equals a single column of W, the algorithm decides that the corresponding transition fired. When it equals the zero vector, the algorithm decides that no transition fired.

When measurement is perturbed by non zero error E, two problems must be mentioned:

1. A miss estimation may occur when $\Delta\hat{M}$ is non zero and different from any columns of W. The estimation algorithm is not able to decide if a transition fired or not and which transition fired. As consequence the algorithm does not give any decision.
2. A wrong estimation may occur when $\Delta\hat{M}$ does not equal actual marking variation ΔM but equals zero vector or another column of W. The estimation algorithm decides if a transition fired or not and which transition fired, but the decision is wrong due to the measurement error.

To overcome these difficulties and to improve estimation, diagnosis can be reformulated as a linear problem in $((Z_3)^n, +, *)$, with the Smith transformation of W, where "+" and "*" stand for the sum and product endowed over Z_3. The Smith transformation results from elementary operations (i.e. row or column permutations, linear combinations and external products), summed up in matrices $P \in (Z_3)^{n \times n}$ and $Q \in (Z_3)^{q \times q}$ such that:

$$P * W * Q = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \quad (11)$$

I_r is the identity matrix of dimension r x r, and r is the rank of matrix W. The Smith transformation leads to reduced incidence matrix W' defined as in equation (12) :

$$W' = (I_r\ 0) * Q^{-1} = (I_r\ 0) * P * W = F * W \in (\mathbf{Z_3})^{r \times q} \quad (12)$$

Necessary and sufficient conditions for firing sequences estimation can be stated when measurement is error free and basic assumption in section 2.2 is satisfied: columns of incidence matrix W' defined by equation (12) are distinct and non zero (Lefebvre, 2006).
In case of measurement errors that satisfy assumptions 1 to 3, two sets of sufficient conditions for firing sequences estimation have been also proposed (Lefebvre, 2007):

1. Columns of incidence matrix W are distinct, non zero and errors E that disturb satisfy $d_0(E) \leq (d(W) - 1) / 2$ (i.e. the number of disturbed entries of measurement is no larger than $(d(W) - 1) / 2$).
2. Columns of incidence matrix W' defined by equation (12) are distinct and non zero, and considered errors E belongs to distinct cosets different from C(0). The coset C(u) of u is defined as $C(u) = \{x \in (Z_3)^n$ such that $x = u + y$ with $y \in Vect(W)\}$, for any vector $u \in (Z_3)^n$.

Moreover, the use of the Smith transformation of incidence matrix is also helpful to define the parity check matrix $H^T = (0\ I_{n-r}) * P \in (Z_3)^{(n-r) \times n}$, and to work out the syndrome of marking variation measurements $S(\Delta\hat{M}) = H^T * \Delta\hat{M}$ and to compare it with the syndrome of marking variation errors $S(E) = H^T * E$. As a consequence the method leads to a less complex and more efficient diagnosis algorithm for DES modeled with PN (algorithm c) in comparison with usual method based on Hamming distance (algorithm b).

Algorithm b

1. For each time k, measure $\hat{M}(k)$ the current state of DES
2. Compute $\Delta\hat{M}(k) = \hat{M}(k) - \hat{M}(k-1)$
3. Compute weight $d_0(\Delta\hat{M}(k))$. If $d_0(\Delta\hat{M}(k)) \leq (d(W) - 1) / 2$, then no event occurs between two consecutive state measurements. Go to step 6.
4. Compute Hamming distance $d(\Delta\hat{M}(k), w_j)$ for each column w_j of W. If $d(\Delta\hat{M}(k), w_j) \leq (d(W) - 1) / 2$ then T_j fired. Go to step 6.
5. If for all $j = 1,...,q$, $d(\Delta\hat{M}(k), w_j) > (d(W) - 1) / 2$ then measurement is too much disturbed by errors (i.e. $d_0(E) > (d(W) - 1) / 2$) and no decision is provided (i.e. a miss estimation occurs).
6. Wait until time k + 1. Go to step 1.

Algorithm c

1. For each time k, measure $\hat{M}(k)$ the current state of DES
2. Compute $\Delta\hat{M}(k) = \hat{M}(k) - \hat{M}(k-1)$
3. Compute $H^T * \Delta\hat{M}(k)$. If $H^T * \Delta\hat{M}(k) = 0$ then measurement is not disturbed by errors: $\Delta M(k) = \Delta\hat{M}(k)$. Go to step 5.
4. If syndrome $H^T * \Delta\hat{M}(k) \neq 0$, compute coset leader E(k) and $\Delta M(k) = \Delta\hat{M}(k) - E(k)$. Go to step 5.
5. Compute $\Delta M'(k) = F * \Delta M(k)$.
6. Compare $\Delta M'(k)$ with zero vector. If $\Delta M'(k) = 0$ then no event occurs between 2 consecutive state measurements. Go to step 8.
7. Compare $\Delta M'(k)$ with columns of matrix W'. If $\Delta M'(k) = w'_j$ then T_j fired. Go to step 8.
8. Wait until time k + 1. Go to step 1.

The correction capacity (i.e. number of error vectors that are corrected) of algorithm b is given by equation (13):

$$\sum_{i=1}^{(d(W)-1)/2} 2^i.\left(\frac{n!}{i!(n-i)!}\right) \tag{13}$$

and its complexity results from 2n.(q+1) scalar comparisons or operations whereas correction capacity of algorithm c equals $3^{n-r} - 1$, and its complexity results from $r.(2n+q)+(n-r).(2n-1+3^{n-r})$ scalar comparisons or operations (Lefebvre, 2007).
As a conclusion, one can notice that algorithm c is more efficient for PN with small rank in comparison with the number of places, and that it is of particular interest for PN with few transitions in comparison with the number of places. Another conclusion is to prefer algorithm c for PN with a small Hamming distance. This result is not surprising as long as the correction capacity of algorithm a is directly related to the value of Hamming distance. For PN with small Hamming distance, the number of biased markings that belong to a single sphere is also small.

5.2 Redundant Petri nets embedding

This method incorporates redundancy into Petri nets and uses algebraic decoding techniques as the Berlekamp – Massey decoding (Berlekamp, 1984) to detect and identify faults (Li et al., 2004; Wu et al., 2005). The marking of the original PN is embedded into a redundant one and the diagnosis of faults is performed by mean of linear parity checks. The algorithm operates in the integer finite field of order p, referred as (Z^+_p) with p a prime integer large enough. This approach has a complexity of $m^2.(n+q)$ (Wu et al., 2002) improved to complexity m.(n+q) (Wu et al., 2005) where 2.m represent the number of places that are added to the original PN.
In comparison with the method in section 5.1, two kinds of faults are considered : (1) place faults are associated with conditions that cause the corruption of the number of tokens in some places of the PN. Place faults are measured, with the Hamming distance metric, in terms of the number of faulty places independent of the number of erroneous tokens in each faulty place ; (2) transition failures are associated with preconditions that prevent tokens from being removed from the input places in some transitions (even though tokens are deposited at the corresponding output places) or postconditions that prevent tokens from being deposited at the output places in some transitions (even though tokens are removed from the corresponding input places). Errors "-1" and "+1" are used respectively and transitions faults are measured with the Lee distance metric (Berlekamp, 1984). By adding 2.m places in the redundant PN, Wu et al. proves that the method allows the simultaneous identification of m place faults and 2.m – 1 transition failures.
It is assumed that the firing of the transitions in the redundant PN are not directly observable whereas the marking is periodically observed, and that the diagnosis is performed over a time interval of N sampling periods (N is eventually chosen equal to 1). It is also assumed that

1. A particular transition does not suffer both a precondition and postcondition during interval [1, N] (otherwise, their effect will be cancelled);
2. The erroneous number of tokens in each place is also bounded within interval [-(p - 1/2), (p - 1/2)];
3. Parameter p is a prime integer that satisfy p > max(n + 2.m, q).

The key idea of the method is to design two matrices $C \in (Z^+_p)^{2m \times n}$ and $D \in (Z^+)^{2m \times n}$ that define the state of the embedded PN such that equation (14) is satisfied :

$$M(k+1) = M(k) + \begin{pmatrix} W_{PO} \\ C.W_{PO} - D \end{pmatrix}.X(k) - \begin{pmatrix} W_{PR} \\ C.W_{PR} - D \end{pmatrix}.X(k) \tag{14}$$

Defining the parity check matrix as in $P = (-C\ I_{2m})$, the syndrome of marking M(k) is given by $S(k) = P.M(k)$.

Let us define $E_T^+(N) \in (Z^+)^q$ as the vector of postcondition faults, $E_T^-(N) \in (Z^+)^q$ as the vector of precondition faults and $E_P(N) \in (Z^+)^{2m+n}$ as the place faults vector during time interval [1, N]. As a consequence the faulty marking is defined as :

$$\hat{M}(N) = M(N) - \begin{pmatrix} W_{PO} \\ C.W_{PO} - D \end{pmatrix}.E_T^+(N) + \begin{pmatrix} W_{PR} \\ C.W_{PR} - D \end{pmatrix}.E_T^-(N) + E_P(N) \tag{15}$$

The identification of both transition failures and place faults based on the syndrome *S(N)* is completely determined by matrices *D* and *C* :

$$S(N) = D.(E_T^+(N) - E_T^-(N)) + P.E_P(N) \tag{16}$$

On one hand, Wu et al. propose to define matrix *D* as in equation (17) :

$$D_{2m} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_q \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \cdots & \alpha_q^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{2m-1} & \alpha_2^{2m-1} & \alpha_3^{2m-1} & \cdots & \alpha_q^{2m-1} \end{pmatrix} \in \left(Z^+\right)^{2m \times q} \tag{17}$$

where α_i are q distinct non zero elements in Z^+_p. In case m = 1, the determination of matrix D is given according to :

$$D_2 = \begin{pmatrix} 1 & 2 & 3 & \cdots & q \\ 1 & 2^2 & 3^2 & \cdots & q^2 \end{pmatrix} \bmod p \in \left(Z^+\right)^{2 \times q}$$

On the other hand, they propose to define matrix C such that equation (18) is satisfied (operations are defined in Z^+_p) :

$$\Phi .(-C\ I_{2m}) = H_{2m} \tag{18}$$

with :

$$H_{2m} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n+2m} \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \cdots & \alpha_{n+2m}^2 \\ \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \cdots & \alpha_{n+2m}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{2m} & \alpha_2^{2m} & \alpha_3^{2m} & \cdots & \alpha_{n+2m}^{2m} \end{pmatrix} \in \left(Z^+\right)^{2m \times (n+2m)}$$

(19)

$$\Phi = \begin{pmatrix} \alpha_{n+1} & \alpha_{n+2} & \alpha_{n+3} & \cdots & \alpha_{n+2m} \\ \alpha_{n+1}^2 & \alpha_{n+2}^2 & \alpha_{n+3}^2 & \cdots & \alpha_{n+2m}^2 \\ \alpha_{n+1}^3 & \alpha_{n+2}^3 & \alpha_{n+3}^3 & \cdots & \alpha_{n+2m}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n+1}^{2m} & \alpha_{n+2}^{2m} & \alpha_{n+3}^{2m} & \cdots & \alpha_{n+2m}^{2m} \end{pmatrix} \in \left(Z^+\right)^{2m \times 2m}$$

In order to identify simultaneously place and fault transition Wu et al. define :

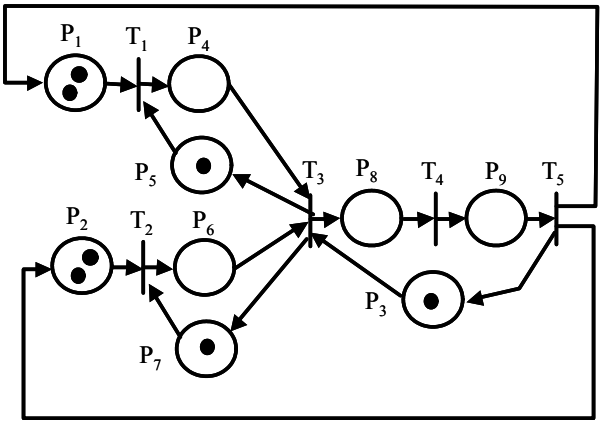
$$\begin{aligned} D^* &= -p.D \\ C^* &= p.1 - C \\ P^* &= (C - p.1\ I_{2m}) \end{aligned} \tag{20}$$

where 1 is a 2m x n matrix with all entries being 1. The syndrome S(N) defined as S(N) = $P^*.M(N)$ is used to identify first m or less place faults by means of the Berlekamp – Massey algorithm and then 2m – 1 transitions by computing the modified syndrome :

$$S_T(N) = (S(N) - P^*.E_p(N)) \ / \ p = D.(E_T^+(N) - E_T^-(N)) \tag{21}$$

5.3 Applications

Algebraic methods have been used for the diagnosis of manufacturing and robotic systems (Lefebvre, 2007, Wu et al., 2005) and for large scale power networks like the IEEE 118-bus power system (Ren et al., 2006).



$$W = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Fig. 6. PN5 model of a manufacturing system

In order to illustrate algebraic methods, let us consider PN5 in figure 6, that is a simplified model of a manufacturing workshop (Silva et al., 2004). The final product is composed of two different parts that are processed in two separate machines modelled by transitions T_1 and T_2, and stored in buffers P_4 and P_6, respectively. Then, they are

assembled by the machine T_3, and processed in T_4 and T_5. During the processing, several tools are needed, modelled by places P_3, P_5 and P_7.
PN5 has n = 9 places, q = 5 transitions, is of rank r = 4 and incidence matrix W has a Hamming distance d = 2. Matrices F and H^T, worked out as in section 5.1, are given according to equation (22):

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad H^T = \begin{pmatrix} -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{22}$$

Syndromes	Errors of weight 1	Syndromes	Errors of weight 1
$(-1\ 0\ 0\ 1\ 0)^T$	$(1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(1\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0)^T$
$(1\ 0\ 0\ -1\ 0)^T$	$(-1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(-1\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ -1\ 0\ 0\ 0\ 0)^T$
$(0\ 1\ -1\ 0\ 0)^T$	$(0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ 1\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0)^T$
$(0\ -1\ 1\ 0\ 0)^T$	$(0\ -1\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ -1\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 0\ -1\ 0\ 0\ 0)^T$
$(1\ -1\ 1\ -1\ 1)^T$	$(0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ 1\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0)^T$
$(-1\ 1\ -1\ 1\ -1)^T$	$(0\ 0\ -1\ 0\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ -1\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 0\ 0\ -1\ 0\ 0)^T$
$(0\ 0\ 0\ 1\ 0)^T$	$(0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ 1)^T$	$(0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0)^T$ $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1)^T$
$(0\ 0\ 0\ -1\ 0)^T$	$(0\ 0\ 0\ -1\ 0\ 0\ 0\ 0\ 0)^T$	$(0\ 0\ 0\ 0\ -1)^T$	$(0\ 0\ 0\ 0\ 0\ 0\ 0\ -1\ 0)^T$ $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -1)^T$

Table 4. Correspondence between syndromes and coset leaders for PN5

PN5 has 243 cosets and each coset has 81 vectors. The table 4 gives the relationships between syndromes and coset leaders. Let us notice that the two last syndromes correspond to two different coset leaders. As a consequence not all errors of weight 1 will be corrected by algorithms b and c (errors $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0)^T$ and $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1)^T$ cannot be separated as errors $(0\ 0\ 0\ 0\ 0\ 0\ 0\ -1\ 0)^T$ and $(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -1)^T$).
Analysis of performance and numerous simulations show that the miss estimation rate for algorithm c is quite better in comparison with algorithm b (Lefebvre 2007), but the wrong estimation rate for c increases in comparison with b. Let us mention that whatever the algorithm used, the presence of miss estimation depends strongly on the Hamming distance of W.
Applying the method developed in 5.2 in order to identify 1 place fault and 1 transition failure (m = 1) with p = 13, the matrices D and D^*, that lead to transition failure diagnosis are given according to equation (23):

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 3 & 12 \end{pmatrix}, \quad D^* = -13 \text{ x } D = \begin{pmatrix} -13 & -26 & -39 & -52 & -65 \\ -13 & -52 & -117 & -39 & -156 \end{pmatrix} \tag{23}$$

On the other hand, the matrices H, C and C*, that lead to place fault diagnosis are given according to equations (24) and (25):

$$H = \begin{pmatrix} 5 & 10 \\ 12 & 9 \end{pmatrix} \begin{pmatrix} 4 & 10 & 2 & 10 & 5 & 2 & 11 & 11 & 4 & 1 & 0 \\ 2 & 3 & 2 & 1 & 3 & 10 & 10 & 6 & 8 & 0 & 1 \end{pmatrix} \tag{24}$$

$$C = \begin{pmatrix} 4 & 10 & 2 & 10 & 5 & 2 & 11 & 11 & 4 \\ 2 & 3 & 2 & 1 & 3 & 10 & 10 & 6 & 8 \end{pmatrix}$$

$$C^* = \begin{pmatrix} 9 & 3 & 11 & 3 & 8 & 11 & 2 & 2 & 9 \\ 11 & 10 & 11 & 12 & 10 & 3 & 3 & 7 & 5 \end{pmatrix} \tag{25}$$

The parity check matrix is defined according to (26) :

$$P^* = \begin{pmatrix} -9 & -3 & -11 & -3 & -8 & -11 & -2 & -2 & \_9 & 1 & 0 \\ -11 & -10 & -11 & -12 & -10 & -3 & -3 & -7 & -5 & 0 & 1 \end{pmatrix} \tag{26}$$

As a consequence 2 places are added in the PN model of figure 6 for diagnosis purposes. These new places are defined according to equation (27):

$$C^*.W_{PO} - D^* = \begin{pmatrix} 16 & 37 & 51 & 61 & 88 \\ 25 & 55 & 137 & 44 & 188 \end{pmatrix}, \quad C^*.W_{PR} - D^* = \begin{pmatrix} 30 & 31 & 64 & 54 & 74 \\ 34 & 65 & 143 & 46 & 161 \end{pmatrix} \tag{27}$$

and the initial marking of embedded PN is given according to (28):

$$M_I^* = \begin{pmatrix} I_9 \\ C^* \end{pmatrix} M_I = \begin{pmatrix} 2 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 45 & 66 \end{pmatrix}^T \tag{28}$$

The use of embedded PN defined with equations (27) and (28) is useful to detect at first place faults and then transitions failures according to the comparison of syndrome S(N) = P*.M(N) with the columns of matrices H given by equation (24) and with the columns of matrix D given by (23).

6. Conclusions

The investigation of diagnosis methods for discrete event systems shows that Petri nets is efficient not only to model the considered systems but also to support the diagnosis methods. Several approaches can be used in order to check diagnosability, to select sensors and to work out diagnosers. The table 5 sums up the main characteristics of these method.

As a conclusion it is important to notice the great effort, observed this last years to develop and improve diagnosis methods for DES. The strong connection with observation properties in automata and the use of advances in computer science like the coding theory have played an important role in that development. Now, the challenges are, from our

point of view, to continue this investigation, by combining the different methods together and also to take advantages from many important contributions that have been proposed for continuous systems. To build a bridge from continuous variable systems to DES theories remains one of the most promising issues for the next years.

	State based approach (section 3)	Event detectability and SOP (sections 4.1 and 4.2)	CR and DP investigation (section 4.3)	Diagnosis in Z_3 (section 5.1)	Diagnosis in Z^+_p (section 5.2)
Diagnosability checking	Yes	Yes	Yes	No	No
Sensor selection	No	Yes	Yes	No	No
Immediate / delayed diagnosis	Immediate and delayed diagnosis at most in k steps	Immediate diagnosis	Delayed diagnosis at least in k steps	Immediate diagnosis	Immediate and delayed diagnosis at most in N steps
Marking measurements	Partial and unbiased	Partial and unbiased	X	Partial and biased	Complete and biased
Partial firing sequence observation	Yes	No	X	No	No
Complexity : Polynomial or Non polynomial	NP	P / NP	P	P	P

Table 5. Main features of several diagnosis methods for DES

7. References

Alcaraz-Mejia M., Lopez-Mellado E., Ramirez-Trevino A., Rivera-Rangel I. (2003). Petri net based fault diagnosis of DES, *Proc. IEEE-SMC03*, pp. 4730-4735, Washington, USA.

Aramburo-Lizarraga, J.; Lopez-Mellado, E.; Ramirez-Trevino, A. (2005). Distributed fault diagnosis using Petri net reduced models**,** *Proc. IEEE-SMC05*, vol. 1, pp. 702-705.

Askin R.G., Standridge C. R. (1993). *Modeling and analysis of Petri nets*, John Wiley and sons Inc.

Berlekamp R.E. (1984). *Algebraic coding theory*, Laguna Hills, CA, Aegean Park.

Benveniste A., Fabre F., Jard C., Haar S. (2003). Diagnosis of asynchronous discrete event systems, a net unfolding approach, *Trans. IEEE –TAC*, vol. 48, no.5.

Blanke M. (1996). Consistent design of dependable control systems, *Control Engineering Practice*, vol. 4, no. 9, pp. 1305 – 1312.

Blanke M., Kinnaert M., Lunze J., Staroswiecki M. (2003). *Diagnosis and fault tolerant control*, Springer Verlag, New York.

Cassandras C.G., Lafortune S. (1999). *Introduction to discrete event systems,* Kluwer Academic Pub.

Chung S.L, Wu C.C., Jeng M. (2003). Failure diagnosis: a case study on modeling and analysis by Petri nets, *Proc. IEEE-SMC03*, pp. 2727-2732, Washington, USA.

Cordier M.O., Dague P., Lévy F., Dumas M., Montmain J. Staroswiecky M., Travé-Massuyès L. (2000). AI and automatic control approaches of model-based diagnosis : links and underlying hypothesis, *Proc. IFAC Symposium on fault detection, Supervision and Safety for Technical Processes*, pp. 274 – 279, Budapest, Hungary.

David R., Alla H. (1992). *Petri nets and grafcet – tools for modelling discrete events systems*, Prentice Hall, London.

Gertler J. (1998). *Fault detection and diagnosis in engineering systems*, Marcel Dekker, New York.

Giua A., Seatzy C. (2002). Observability of place / transition nets, *Trans. IEEE – TAC*, vol. 47, no. 9, pp. 1424 – 1437.

Ichikawa, A., Hiraishi, K. (1988). Analysis and Control of Discrete Event Systems Represented by Petri Nets, *Proc. IIASA Conf.*, pages 115-134. Springer-Verlag, Berlin, West Germany.

Lefebvre D., El Moudni A. (2001). Firing and enabling sequences estimation for timed Petri nets, *Trans. IEEE - SMCA*, vol. 31, no.3, pp. 153- 162.

Lefebvre D., Delherm C. (2005). Diagnosis with causality relationships and directed paths in Petri net models, *Proc. IFAC WC05*, Prague, Czech Republic.

Lefebvre D. (2004). About estimation problems with Petri net models for fault detection and isolation with discrete event and hybrid systems, *Proc. SAUM04*, Invited lecture, pp. 42 – 51, Beograd, Serbia and Montenegro.

Lefebvre D., Delherm C., Leclercq E., Druaux F. (2006). Some contributions with Petri nets for the modelling, analysis and control of HDS, *Proc. ICHSA*, Invited lecture, Lafayette, Louisiana, USA.

Lefebvre D. (2006a). Sensoring and diagnosis of DES with Petri net models, *Proc. IFAC Safeprocess 2006*, invited session "Model based fault analysis during a system's entire life cycle", pp. 1213 - 1218, Beijing, China.

Lefebvre D. (2006b) Firing sequences estimation for ordinary Petri nets, *Proc. Workshop IAR - ACD*, Nancy, France.

Lefebvre D., Delherm C. (2007). Fault detection and isolation of discrete event systems with Petri net models, *Trans. IEEE – TASE*, vol. 4, no. 1, pp. 114 – 118.

Lefebvre D. (2007). Firing sequences estimation in vector space over Z_3 for ordinary Petri nets, *Trans. IEEE – SMCA*, under review.

Li L., Hadjicostis C. N. Sreenivas R. S. (2004). Fault Detection and Identification in Petri Net Controllers, *Proc. IEEE-CDC04*, pp. 5248 – 5253, Atlantis, Paradise Island, Bahamas.

Patton R.J., Frank M., Clark R.N. (Eds), (1989). *Fault diagnosis in dynamic systems theory and applications*, Prentice Hall, New York.

Patton R.J., Frank M., Clark R.N. (Eds), (1999). Iss*ues of fault diagnosis for dynamical systems*, Springer Verlag, London.

Ramirez-Trevino A., Ruiz-Bletran E., Rivera-Rangel I., Lopez-Mellado E. (2004). Diagnosability of discrete event systems. A Petri net based approach, *Proc. IEEE-ICRA*, pp. 541 – 546.

Ramirez-Trevino A., Ruiz-Bletran E., Rivera-Rangel I., Lopez-Mellado E. (2007). Online Fault Diagnosis of Discrete Event Systems. A Petri Net-Based Approach, *Trans. IEEE – TASE*, vol. 4, no. 1, pp. 31-39.

Rausand M., Hoyland A. (2004). *System reliability theory : models, statistical methods, and applications*, Wiley, Hoboken, New Jersey.

Ren H., Mi Z. (2006). Power system fault diagnosis modeling techniques based on encoded Petri nets, *Proc. IEEE Power Engineering Society General Meeting*.

Sampath M., Sengupta R., Lafortune S., Sinnamohideen K., Teneketzis D. (1995). Diagnosibility of discrete event systems, *Trans. IEEE-TAC*, vol. 40, no.9, pp. 1555-1575.

Silva M., Recalde L. (2004). On fluidification of Petri Nets: from discrete to hybrid and continuous models, *Annual Reviews in Control,* vol. 28, no. 2, pp. 253-266.

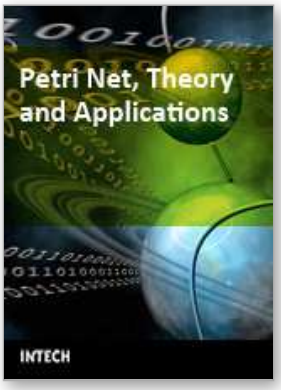
Ushio T., Onishi I., Okuda K., (1998). Fault detection based on Petri net models with faulty behaviours, *Proc. IEEE – SMC98*, pp 113-118.

Van Lint J.H. (1999). *Introduction to Coding Theory*, Graduate Texts in Mathematics, vol. 86, Springer Verlag.

Wen Y.L, Li C.H., Jeng M. (2005). A polynomial algorithm for checking diagnosability of Petri nets, *Proc. IEEE-SMC05*, pp. 2542-2547, vol. 3.

Wu Y., Hadjicostis N. (2002). Non-concurrent fault identification in discrete event systems using encoded Petri net states, *Proc. IEEE – CDC02*, vol. 4, pp4018-4023.

Wu Y., Hadjicostis N. (2005). Algebraic approaches for fault identification in discrete event systems, *Trans. IEEE - TAC*, vol. 50, no. 12, pp. 2048 – 2053.



Petri Net, Theory and Applications

Edited by Vedran Kordic

ISBN 978-3-902613-12-7
Hard cover, 534 pages
Publisher I-Tech Education and Publishing
Published online 01, February, 2008
Published in print edition February, 2008

Although many other models of concurrent and distributed systems have been de- veloped since the introduction in 1964 Petri nets are still an essential model for concurrent systems with respect to both the theory and the applications. The main attraction of Petri nets is the way in which the basic aspects of concurrent systems are captured both conceptually and mathematically. The intuitively appealing graphical notation makes Petri nets the model of choice in many applications. The natural way in which Petri nets allow one to formally capture many of the basic notions and issues of concurrent systems has contributed greatly to the development of a rich theory of concurrent systems based on Petri nets. This book brings together reputable researchers from all over the world in order to provide a comprehensive coverage of advanced and modern topics not yet reflected by other books. The book consists of 23 chapters written by 53 authors from 12 different countries.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Dimitri Lefebvre (2008). Diagnosis of Discrete Event Systems with Petri Nets, Petri Net, Theory and Applications, Vedran Kordic (Ed.), ISBN: 978-3-902613-12-7, InTech, Available from: http://www.intechopen.com/books/petri_net_theory_and_applications/diagnosis_of_discrete_event_systems_with_petri_nets

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821