REGULAR ARTICLE

Diagnostic Based on Estimation Using Linear Programming for Partially Observable Petri Nets with Indistinguishable Events

Amira. Chouchane [a] Philippe. Declerck [b] Atef. Khedher [c] and Anas. Kamoun [a]

[a]National engineering School of Sfax, Tunisia; [b]University of Angers, Laris, Istia, France; [c]National engineering School of Tunis, Tunisia

ARTICLE HISTORY
Compiled November 27, 2018

ABSTRACT
In this paper, we design a diagnostic technique for a partially observed labeled Petri net where the faults of the system are modeled by unobservable transitions. The fault detection and isolation uses an on-line count vector estimation associated with the firing of unobservable transitions exploiting the observation of firing occurrences of some observable transitions. The support of the approach is an algebraic description of the process under the form of a polyhedron developed on a receding horizon. We show that a diagnostic can be made despite that different transitions can share the same label and that the unobservable part of the Petri net can contain circuits.

KEYWORDS
Petri net; count vector estimation; algebraic description; diagnostic

1. Introduction

Fault diagnosis is the process of Fault Detection and Isolation (FDI) which can also include fault identification. Fault detection is a binary decision that decides whether the system is in normal or abnormal operation. Fault isolation consists in identifying the system component (plant equipment as a sensor or actuator, component malfunctioning, etc.) responsible for the occurrence of this fault. Fault identification consists in determining the amplitude and possible evolution of the faults over time.

This paper focuses on the diagnosis of a process whose state evolves at the occurrence of events. This process can be modeled as a discrete event system (DES). This class of models presents numerous applications as transport networks, computer systems, multimedia systems, food industry, and manufacturing systems. In the literature, we find two graphical and mathematical tools for modeling the DES which are the Finite State Automaton (FSA) and the Petri Net (PN). In this paper, we choose the PN model that can model the phenomena of synchronism, assembly and sharing of resources, thanks to its structure. For various economic and/or technical reasons, the presence of a sensor for each system variable is not always possible. As a result, the PN may contain transitions that model unobservable events in the system. Unobservable events can also model faults, disturbances, or noises that can affect the system. The relevant transitions are named unobservable transitions and the PN is then called a

choucheneamira41@yahoo.fr

partially observed PN.

In the litterature, the most known methods for PN diagnosis has been based on the faulty model for fault diagnosis (Cabasino et al., 2013; Wang et al., 2015). In fact, methods working with fault-model are well adapted for fault isolation. The techniques used for the diagnostis depend on the knowledge available on the system as well as on the detectable faults affecting the system (Wu et al., 2005), (Benveniste et al., 2003). For fault diagnosis, we can find event-based, state-based, and mixed-based faults. The event-based faults model the system faults into a set of transitions, and the occurrence of certain faults is equivalent to firing the associated transitions. The detection and isolation of faults are carried out based only on observed events. These event-based models have the advantage of detecting intermittent faults (Genc et al., 2007)(Garcia et al., 2008)(Ramirez et al., 2012). The state-based faults consider that the occurrence of a fault is equivalent to the change in the state of the PN deviating from its nominal behavior, which is expressed by losses or duplications of tokens. The disadvantage of the state-based faults' modulation is that it can not detect intermittent faults that are short events leading to unstable states (Wu et al., 2005)(Benveniste et al., 2003). The mixed-based faults modulation is a combination of the occurrence of fault events and the attainability of fault states (Wu et al., 2005).

In this paper, we are interested in the diagnosis of a partially observed Labeled PN (LPN). Assuming that the faults of the system are modeled by some unobservable transitions (event-based faults), we focus on the optimistic and pessimistic cases of the occurrences of faults, which allow the interpretation of fault detection and isolation. Our diagnosis approach is based on an observer that estimates the events that cannot be directly observed by an outside observer. Exploiting the observation of events available from the PN, the observer ideally makes an estimation of the events relevant to unobservable firing sequences that enable a sequence of observed transitions from a given initial marking. More practically, as the image of the sequence is a firing count vector that counts the number of each unobservable transition firing, an objective is to estimate the count vectors coherent with the observed events. A fault will be detected if the observer notices that all the possible unobservable firing sequences include a fault transition at least. Precisely, the count vector of fault transitions is always different from the null vector. Symmetrically, no fault will appear if the observer cannot generate a possible unobservable firing sequence including a fault transition.

The first work on the diagnosis of PNs is based on the control of the changes of the tokens in P-invariants of a generalized PN (Portinale, 1995). A fault is detected when a P-invariant do not keep their number of tokens. In (Ramirez et al., 2012), the difference of the marking of the actual behavior model and the estimated marking, called residue, provided enough information for the immediate isolation of faults. In (Cabasino et al., 2013; Ran et al., 2017), a diagnosis approach of partially observed LPN was based on the notion of basis markings which was a reduced set of actual markings coherent to an observed sequence. The faults were modeled by unobservable transitions and might also been modeled by undistinguishable observable transitions. If the LPN was bounded, the diagnosis approach was based on the Basis Reachability Graph (BRG) which can be computed off-line. An Extended BRG (EBRG) was constructed in (Ran et al., 2017) which is a basis marking computed by assuming that all the system faults are observable. The EBRG has significantly fewer states than the reachability graph in most cases, but it still exponential with respect to the number of nodes. Wang et al. (Wang et al., 2015) used a Fault Diagnosis Graph (FDG) for fault diagnosis of partially observed Time PN (TPN) were the faults are modeled by unobservable transitions. The FDG was incrementally computed using the State Class

Graph (SCG). Starting from the SCG, Basile et al. (Basile et al., 2015) determine a new graph called the Modified SCG (MSCG), that visualize in a compact form the main information on all the possible evolutions of a given labeled TPN system initialized at a given marking. Given a timed observation, i.e., a sequence of observed labels with the corresponding time instants of observation, the fault diagnosis is based on the exploration of the MSCG and on the solution of some ILP problems to determine which states are consistent with the observation. If all the consistent state that is reachable at $t = 0$ via sequences containing some fault transitions, then a fault is detected. Lefebvre (Lefebvre, 2014) used linear matrix inequalities to compute the firing sequences consistent with each elementary observation sequence for a partially observed PN diagnosis. A forward-backward algorithm was suggested, which analysed the sub-sequences of bounded lengths. A fault is detected if these estimated sub-sequences include fault transition. Sufficient conditions for detection and isolation were formulated as ILP problems. In (Dotoli et al., 2009), the diagnosis approach of partially observed LPN was based on the resolution of some ILP problems for each observed transition of the observed firing sequence. The initial marking update after each considered transition of the observed sequence was avoided. The system faults were modeled by some unobservable transitions. A fault is detected if it is includes in a subset of unobservable sequences coherent with the observation verifying an ILP problem. In (Basile et al., 2009), authors proposed an approach of fault diagnosis of a partially observed PN requiring an on-line computation of the set of possible fault events explaining the last observed event. The on-line computation consisted in solving the ILP problems formulated on a net structure and based on g-markings. The resolution time of an ILP problem is in general case NP-hard. Therefore, the diagnosis procedure is time consuming. To reduce the temporal complexity, the fluidization technique was explored for fault diagnosis (Mahulea et al., 2012).The main idea of the fluidization of PNs is the relaxation of the transitions firings allowing them to fire in positive real amounts. The fluidization of a partially observed purely logic PN produces an untimed Continuous PN (CPN) with a set of markings that are consistent with a given observation which is convex. Using this convexity property, the diagnosis states are computed solving LP problems for the untimed CPNs. For fault detection, two LP problems were defined for each observed transition of the observed sequence and for each fault transition under the assumption that the unobservable subnet was acyclic. Moreover, the update of the initial marking after each observed transition of the observation was necessary. Therefore, this diagnostic approach is time and space consuming. In (Hadjicostis et al., 1999), fault diagnosis is based on incorporating structured redundancy into a PN. Monitors were constructed, which operated concurrently with the original system and allowed detecting and identifying different types of faults by performing consistency checks between the state of the original PN and that of the monitor state. In (Yang et al., 2009), the authors were interested in diagnosing the faults of controller that was modeled by PN with uncontrollable and unobservable transitions. The inadmissible constraints with uncontrollable and unobservable transitions were transformed into admissible conditions in this method. Separate redundant PN controllers utilizing additional places, connections and tokens to impose invariant conditions enabled the systematic detection and identification of faults via the Hamming code. In (Wu et al., 2005), the authors devised an algebraic approach for partially observed PN diagnosis. Two types of faults to be diagnosed based on the observed marking at the end of a period were the transition faults and the place faults. For this purpose, a number of places were added to the PN model. The resulting model was a redundant PN that included $2q$ additional places and allowed the identification of $2q-1$ transition faults and

q place faults. The incorporated redundancy permitted fault diagnosis using algebraic decoding techniques. In terms of coding theory, the transition faults were measured in the form of the "Lee metric distance" and the place faults were measured in the form of the "Hamming metric distance". However, the computational Techniques in algebraic coding theory are both time and space consuming.

In this paper, the fault detection is based on the resolution of at most two LP problems for an observed sequence w. We assume that the incidence matrices and the initial marking (denoted M^{init} in what follows) are known. The occurrences of observable transitions are non-simultaneous. We also assume the feasibility of the system and the presence of observations during the application of the estimation procedure. The same label can be associated with more than one observable transition and circuits can exist in the unobservable subnet.

The procedure of estimation is usually based on the sequential treatment of each transition produced by the observed sequence in an on-line procedure. The firing of a found unobservable sequence and the observed transition generates a set of new current markings used in the following step. The search procedure is repeated at the occurrence of each event. However, this procedure presents some drawbacks. As this technique needs to consider any starting marking in the estimation at each iteration, this procedure is time-consuming and the boundedness of spaces is problematic (Ru et al., 2009). Another difficulty is to extend the obtained results (as an optimality) for one observed transition to a sequence of observed transitions even if some results exist for minimal solutions (Li et al., 2011). Therefore, we generalize this procedure by introducing in this paper an approach based on a new description of the problem, not for a unique observed transition but for a set of observed transitions which are possibly indistinguishable. Thus, the initial marking update after each considered transition of w is avoided. The estimation problem is algebraically described under the form of a polyhedron $A.x \leq b$ over $\mathbb{Z}$ with an unknown x which is developed on a receding horizon. This construction implies that the different possible current markings are not computed but are algebraically expressed in the polyhedron. To the best of our knowledge, the case of indistinguishable observable transitions has been considered by only few authors and the proposed procedure has been the only one that could consider the indistinguishable observable transitions under an algebraic point of view, which can express it clearly and allows its reuse in other fields. The hypothesis of acyclicity is not taken in this article, contrary to many papers in this topic, except (Cabasino et al., 2013; Dotoli et al., 2009).

The paper is organized as follows. In section 2, we present the preliminary notions. In section 3, we build a polyhedron defined on a receding horizon, which describes the estimation problem under an algebraic point of view. In section 4, we introduce two criteria which permit the fault detection and isolation. The proposed approach is illustrated in section 5 by a pedagogical example containing a circuit and two transitions sharing the same label. In section 6, we consider the complexity. In section 7, a numerical comparison with the discrete approach based on "basis markings" is performed. In section 8, we end with a conclusion and some perspectives.

2. Preliminary

The notation $|Z|$ is the cardinality of set Z, and the notation A^Tcorresponds to the transpose of the matrix A. The upper integer part and the lower integer part of $\alpha \in \mathbb{R}$ denoted $\lceil \alpha \rceil$ and $\lfloor \alpha \rfloor$, respectively. A Place/Transition (P/TR) net is the structure $N =$

(P, TR, W^+, W^-), where P is a set of $|P|$ places and TR is a set of $|TR|$ transitions. The matrices W^+ and W^- are respectively the $|P| \times |TR|$ post and pre-incidence matrices over $\mathbb{N}$, where each row $l \in \{1, ..., |P|\}$ specifies the weight of the incoming and outgoing arcs of the place $p_l \in P$. The incidence matrix is $W = W^+ - W^-$. The pre-set and post-set of the node $z \in P \bigcup TR$ are denoted by ${}^\bullet z$ and $z^\bullet$, respectively.

A labeling function $L : TR \to AL \cup \{\varepsilon\}$ assigns to each transition $x_i \in TR$ either a symbol from a given alphabet AL or the empty string ε. Without loss of generality, the mapping L is assumed to be surjective. In a partially observed LPN, we assume that the set of transitions TR can be partitioned as $TR = TR_{obs} \bigcup TR_{un}$, where the set TR_{obs} (respectively TR_{un}) is the set of observable transitions (respectively unobservable transitions) associated with a label of AL (respectively the empty string ε).

The unobservable induced subnet of the Petri net N is defined as the new net $N_{un} = (P, TR_{un}, W^+_{un}, W^-_{un})$ where W^+_{un} and W^-_{un} are the restrictions of W^+ and W^- to $P \times TR_{un}$. Therefore, $W_{un} = W^+_{un} - W^-_{un}$. The observed subnet of N is defined as the new net $N_{obs} = (P, TR_{obs}, W^+_{obs}, W^-_{obs})$ where W^+_{obs} and W^-_{obs} are the restrictions of W^+ and W^- to $P \times TR_{obs}$. A reorganization of the columns with regards to TR_{un} and TR_{obs} yields $W = \begin{pmatrix} W_{un} & W_{obs} \end{pmatrix}$. The notation $(x_{un})_i$ expresses an unobservable transition, belonging to TR_{un}, while an observable transition belonging to TR_{obs} is denoted $(x_{obs})_i$.

The notation Θ^* represents the set of firing sequences, denoted σ, consisting of transitions of the set $\Theta \subset TR$. The vector $\overline{\sigma}$ (respectively $\overline{x_{un}}$) of dimension $|TR|$ (respectively $|TR_{un}|$) expresses the firing vector or count vector of the sequence $\sigma \in TR^*$(respectively $x_{un} \in TR^*_{un}$), where the i-th component $\overline{\sigma}_i$ (respectively $(\overline{x_{un}})_i$) is the firing number of the i-th transition of TR, which is fired $\overline{\sigma}_i$ times in the sequence σ (respectively $(x_{un})_i \in TR_{un}$ which is fired $(\overline{x_{un}})_i$ times in the sequence x_{un}). The same notation is taken for $\overline{x_{obs}}$ of dimension $|TR_{obs}|$. The reorganization of the components of $\overline{\sigma}$ yields $\overline{\sigma} = \begin{pmatrix} \overline{x_{un}}^T & \overline{x_{obs}}^T \end{pmatrix}^T$.

The marking of the set of places P is a vector $M \in \mathbb{N}^{|P|}$ that assigns to each place $p_i \in P$ a non-negative integer number of tokens M_i, represented by black dots. The i-th component M_i is also written as $M(p_i)$. The marking M reached from the initial marking M^{init} (which replaces the usual notation M_0) by firing the sequence σ can be calculated by the fundamental relation: $M = M^{init} + W.\overline{\sigma}$. The transition $x_i \in TR$ which can be $(x_{un})_i \in TR_{un}$ or $(x_{obs})_i \in TR_{obs}$ is enabled at M if $M \geq W^-(., x_i)$ and may be fired yielding the marking $M' = M + W(., x_i)$. We write $M[\sigma \succ$ to denote that the sequence of transitions σ is enabled at M, and we write $M[\sigma \succ M'$ to denote that the firing of σ yields M'.

We take $n = |TR_{un}|$, $n' = |TR_{obs}|$, $n'' = |AL|$ and $m = |P|$.

3. Relaxed problem on a horizon

The problem considered in this part is as follows. Let us consider a LPN where the incidence matrix W and the initial marking M^{init} are known. Given a sequence of labels of AL emitted by the firing of the observable transitions of TR_{obs}, which are generated by the activity of the LPN, we want to algebraically describe the space of the count vectors that are coherent with the observations.

3.1. Polyhedron of LPN

We present here a linear algebraic approach based on the fundamental equation of marking and the conditions of firing transitions in a LPN. Assuming that the occurrences of observable transitions and the production of a relevant label are non-simultaneous, we associate an iteration $< i >$ to each occurrence of an observable transition. The k observations defines the horizon $\{1, \ldots, k\}$ where 1 and k corresponds the first and last observations, respectively.

The firing of the observable transition $(x_{obs})_j$ for $j \in \{1, \ldots, |TR_{obs}|\}$ and the relevant count vector for each iteration $i \in \{1, ..., k\}$ are respectively denoted $(x_{obs})_j^{<i>}$ and $(\overline{x_{obs}})_j^{<i>}$. The same notations are taken for the count vector of the unobservable transitions $(x_{un})_j$ for $j \in \{1, \ldots, |TR_{un}|\}$, and the count vector of the observed labels is defined as follows: The vector $\overline{y}^{<i>}$ of dimension $|AL|$ describes the count vector of the observed labels for iteration $< i >$ where the component associated to a unique label $a \in AL$ is the number of appearances of this label generated by the observable transitions of the LPN expressed by the firings of the transition $(x_{obs})_j$. Thus, the following notations are relevant to the horizon $\{1, \ldots, k\}$:

$\overline{x_{un}} = \left(\ (\overline{x_{un}}^{<1>})^T \quad (\overline{x_{un}}^{<2>})^T \quad (\overline{x_{un}}^{<3>})^T \quad \ldots \quad (\overline{x_{un}}^{<k>})^T \ \right)^T$,

$\overline{x_{obs}} = \left(\ (\overline{x_{obs}}^{<1>})^T \quad (\overline{x_{obs}}^{<2>})^T \quad (\overline{x_{obs}}^{<3>})^T \quad \ldots \quad (\overline{x_{obs}}^{<k>})^T \ \right)^T$

and $\overline{y} = \left(\ (\overline{y}^{<1>})^T \quad (\overline{y}^{<2>})^T \quad (\overline{y}^{<3>})^T \quad \ldots \quad (\overline{y}^{<k>})^T \ \right)^T$.

The dimensions of $\overline{x_{un}}$, $\overline{x_{obs}}$ and $\overline{y}$ are $k.n$, $k.n'$ and $k.n''$, respectively. We have $(\overline{x_{obs}})_j^{<i>}$ and $(\overline{y})_j^{<i>} \in \{0, 1\}$ as we consider a unique firing of an observable transition for each step $< i >$ by assumption. We take later $M^{<1>} = M^{init}$.

Definition 3.1. When the estimate of the firing count vector $\overline{x_{un}}$ associated to the unobservable transitions corresponds to a sequence that can be followed by the LPN, this count vector will be named "*explanation vector*". The sets of all possible explanation vectors for the starting marking $M^{<1>}$ and the observation $\overline{y}$ are denoted $E(\overline{y})$.

The set of explanation vectors exists as we assume that the LPN is live on the horizon $\{1, \ldots, k\}$. Let us consider an iteration $< i >$. The algebraic formulation of a possible explanation vector is made using the above reasoning for a given marking $M^{<i>}$ and the unique observation $(x_{obs})_j^{<i>}$ with an iteration $< i >$. As $M^{<i>}[x_{un}^{<i>} \succ M'$, the marking M' satisfies the following equation:

$$M' = M^{<i>} + W_{un} \cdot \overline{x_{un}}^{<i>} \tag{1}$$

In addition, the transition $(x_{obs})_j^{<i>}$ is enabled for M'. As $M'[(x_{obs})_j^{<i>} \succ$, we can write the following inequality:

$$M' \geq W_{obs}^{-}.\overline{x_{obs}}^{<i>} \tag{2}$$

where $\overline{x_{obs}}^{<i>} = 0$, except a unique entry $(\overline{x_{obs}})_j^{<i>}$. By replacing M' by its expression (1), we obtain:

$$-W_{un}.\overline{x_{un}}^{<i>} + W_{obs}^{-}.\overline{x_{obs}}^{<i>} \leq M^{<i>} \tag{3}$$

for $i \geq 1$, with the constraint of non-negativity $\overline{x_{un}}^{<i>}, \overline{x_{obs}}^{<i>} \geq 0$.

Let us develop the relations on the horizon $\{2, \ldots, k\}$. The firing of a found unobservable sequence and of the observed transition generates a new current marking used in the following step. Formally, $M'[(x_{obs})_j^{<i>} \succ M^{<i+1>}$. Hence, $M^{<i>} = M^{<i-1>} + W_{un}.\overline{x_{un}}^{<i-1>} + W_{obs}.\overline{x_{obs}}^{<i-1>}$ and more generally, as

$$M^{<i>} = M^{<1>} + W_{un}.\sum_{j=1}^{i-1} \overline{x_{un}}^{<j>} + W_{obs}.\sum_{j=1}^{i-1} \overline{x_{obs}}^{<j>}$$

for $i \geq 2$, we obtain:

$$-W_{un}.\sum_{j=1}^{i} \overline{x_{un}}^{<j>} - W_{obs}.\sum_{j=1}^{i-1} \overline{x_{obs}}^{<j>} + W_{obs}^{-}.\overline{x_{obs}}^{<i>} \leq M^{<1>} \tag{4}$$

The consideration of (3) for $i = 1$ and (4) for $i \geq 2$ leads to the following system:

$$A.\overline{x_{un}} + B.\overline{x_{obs}} \leq b \text{ with } \overline{x_{un}}, \overline{x_{obs}} \geq 0 \tag{5}$$

where $A = \begin{pmatrix} -W_{un} & 0 & 0 & & 0 \\ -W_{un} & -W_{un} & 0 & & 0 \\ -W_{un} & -W_{un} & -W_{un} & & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -W_{un} & -W_{un} & -W_{un} & \ldots & -W_{un} \end{pmatrix}$,

$B = \begin{pmatrix} W_{obs}^{-} & 0 & 0 & \ldots & 0 \\ -W_{obs} & W_{obs}^{-} & 0 & \ldots & 0 \\ -W_{obs} & -W_{obs} & W_{obs}^{-} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -W_{obs} & -W_{obs} & -W_{obs} & \ldots & W_{obs}^{-} \end{pmatrix}$ and $b = \begin{pmatrix} M^{<1>} \\ M^{<1>} \\ M^{<1>} \\ \vdots \\ M^{<1>} \end{pmatrix}$.

The dimension of the matrices A and B and the vector b are $(k.m \times k.n)$, $(k.m \times k.n')$ and $(k.m \times 1)$, respectively. The matrices A and B depend on the structure of the LPN, whereas the vector b depends on the initial marking.

Remark:

Note that A, B, b, $\overline{x_{un}}$ and $\overline{x_{obs}}$ depend on the horizon $\{1, \ldots, k\}$. If the horizon is reduced to $\{1, 1\}$, we obtain the well-known inequality (3).

3.2. Indistinguishable events

The first non-determinism is produced by the fact that the label ε is associated with all the unobservable transitions of TR_{un}. The second non-determinism is as follows. Let us consider the case where the mapping L is not injective. The same label of the alphabet AL can be associated with more than one observable transition. Consequently, several observable transitions may share the same label $a \in AL$, which leads to an ambiguity in the data as we cannot deduce the effective transition. Accordingly, the emission of the label is said to be ambiguous as it does not permit distinguishing associated transitions. Let $\Omega_a \subset TR_{obs}$ be a set of observable transitions for the label a. Formally, $\Omega_a = \{(x_{obs})_j \in TR_{obs} \mid L((x_{obs})_j) = a\}$.

We can deduce a connection between $(\overline{x_{obs}})_j^{<i>}$ for $(x_{obs})_j \in \Omega_a$ and $(\overline{y_a})^{<i>}$ for each observation, which is $\sum_{(x_{obs})_j \in \Omega_a} (\overline{x_{obs}})_j^{<i>} = (\overline{y_a})^{<i>}$ with $(\overline{x_{obs}})_j^{<i>}$, $(\overline{y_a})^{<1>} \in \{0,1\}$.

As we assume that there is a unique firing of a transition $(x_{obs})_j$ which occurs at each iteration, the case $(\overline{x_{obs}})_j^{<i>} = 1$ implies $(\overline{y_a})^{<i>} = 1$. Conversely, $(\overline{y_a})^{<i>} = 1$ implies the existence of a transition $(x_{obs})_j^{<i>}$. Therefore, after building the relevant matrix $C \geq 0$ with the appropriate components 0 or 1, we can write: $C.\overline{x_{obs}}^{<i>} = \overline{y}^{<i>}$ for $i \in \{1, \ldots, k\}$, where the dimensions of C are $(n''\text{x}n')$, and we obtain:

$$\boldsymbol{C}.\overline{x_{obs}} = \overline{y} \tag{6}$$

with $\boldsymbol{C} = \begin{pmatrix} C & 0 & 0 & \ldots & 0 \\ 0 & C & 0 & \ldots & 0 \\ 0 & 0 & C & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & C \end{pmatrix}$.

The dimension of $\boldsymbol{C}$ is $(k.n''\text{x}k.n')$. Accordingly, $\boldsymbol{C}.\overline{x_{obs}} = \overline{y}$ is equivalent to $\boldsymbol{C}.\overline{x_{obs}} \leq \overline{y}$ and $-\boldsymbol{C}.\overline{x_{obs}} \leq -\overline{y}$ and the relaxation of $(\overline{x_{obs}})_j^{<i>} \in \{0,1\}$ leads to $0 \leq \overline{x_{obs}} \leq u$, where u is the unitary vector of the dimension of $\overline{x_{obs}}$.

3.3. Final polyhedron

Finally, the final system used in this paper is as follows:

$$\boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0 \tag{7}$$

with $\boldsymbol{A} = \begin{pmatrix} I & 0 \\ \boldsymbol{C} & 0 \\ -\boldsymbol{C} & 0 \\ B & A \end{pmatrix}$, $\overline{x} = \begin{pmatrix} \overline{x_{obs}} \\ \overline{x_{un}} \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} u \\ \overline{y} \\ -\overline{y} \\ b \end{pmatrix}$.

The dimensions of $\boldsymbol{A}$, $\overline{x}$ and $\boldsymbol{b}$ are $(k.(m+n'+2.n'')\text{x}k.(n+n'))$, $(k.(n+n')\text{x}1)$ and $(k.(m+n'+2.n'')\text{x}1)$, respectively. Note that if $n' = n''$, $C{=}I_{n''\text{x}n''}$, then $\boldsymbol{C}{=}I_{k.n''\text{x}k.n''}$, $\overline{x_{obs}} = \overline{y}$ and the system is reduced to $A.\overline{x_{un}} \leq b - B.\overline{y}$ with $\overline{x_{un}} \geq 0$.

Remark

Generally, the integer solution of matrix inequality (7) describes a set that includes the set of explanation vectors. Indeed, the existence of non-negative integer solutions of fundamental inequality (7) is not sufficient to determine the true explanation vectors that express valid firing sequences (Kostin, 2003). For the PN with acyclic TR_{un}-induced subnets, the solution set of matrix inequality (7) in $\mathbb{Z}^{k.(n+n')}$ is coincident with the set of explanation vectors.

We denote $S^{\mathbb{R}}(\overline{y}) = \{\overline{x} \in \mathbb{R}^{\text{k.(n+n')}} | \boldsymbol{A}.\overline{x} \leq \boldsymbol{b},\ \overline{x} \geq 0\}$ the space of admissible solutions over $\mathbb{R}$ and we denote $S^{\mathbb{Z}}(\overline{y}) = S^{\mathbb{R}}(\overline{y}) \cap \mathbb{Z}^{k.(n+n')}$ the space of admissible solutions over $\mathbb{Z}$. Accordingly, the solutions of the obtained algebraic model (7) over $\mathbb{R}$ or $\mathbb{N}$ always includes the set of explanation vectors $E(\overline{y})$. Therefore, the solutions of $S^{\mathbb{R}}(\overline{y})$ and $S^{\mathbb{N}}(\overline{y})$ are named *candidate solutions* or *candidate vectors over* $\mathbb{R}$ *and* $\mathbb{N}$, respectively. This inclusion can possibly be strict as the firing conditions of the unobservable transitions are neglected in this part. As a result, we have:

$$S^{\mathbb{R}}(\overline{y}) \supseteq S^{\mathbb{N}}(\overline{y}) \supseteq E(\overline{y}) \tag{8}$$

which defines the context of this paper. Considering $\overline{x_{un}}, \overline{x_{obs}}$ in the integers, this system can be solved by standard solvers of integer linear programming if we add a criterion. However, this resolution is exponential in the worst case, and we can prefer the relaxation of system (7) where efficient algorithms of linear programming, such as the Simplex algorithm, the Karmarkar's algorithm and the Khashiyan's algorithm, can be applied.

Analyzing the relaxation of (7), the objective is now to determine the optimal candidate solution in $S^{\mathbb{R}}$ for a linear criterion dedicated to the diagnostic problem. We show that a diagnostic can be made despite the relaxation over $\mathbb{R}$ by determining a lower bound and an upper one (if it is possible) for fault occurrences.

4. Diagnosis of a partially observable LPN

We assume that the faults occurring in the process are modeled by unobservable transitions and the notation TR_f represents the relevant set. The set of unobservable transitions describing a normal behavior is denoted TR_n . Therefore, $TR_f \subset TR_{un} =$ and

$$TR_{un} = TR_n \cup TR_f \tag{9}$$

with $TR = TR_{obs} \cup TR_{un}$.

4.1. Fault detection

The process contains a fault if there is at least the firing of one fault transition for the observed word w. This means that the firing sum of the fault transitions is greater than or equal to 1 for the k observed transitions of w. Let us define the relevant criterion $\boldsymbol{c}_{\text{det}}.\overline{x}$ where the row-vector $\boldsymbol{c}_{\text{det}}$ is the concatenation of the submatrix of $k.n'$ zeros and the submatrix $\begin{pmatrix} c^{<1>} & c^{<2>} & c^{<3>} & \dots & c^{<k>} \end{pmatrix}$ relevant to TR_{obs} and TR_{un}, respectively. The components $(c^{<i>})_j$ for any fault transition $j \in TR_f \subset TR_{un}$ are equal to 1 for any $i = 1, ..., k$, while the other ones are null. As a result, $c^{<1>} = c^{<2>} = ... = c^{<k>}$. Thereby, we can consider the following J^-_{det} criterion:

$$\begin{cases} J^-_{\text{det}} = \min(\boldsymbol{c}_{\text{det}}.\overline{x}) \\ s.t. \quad \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0 \end{cases} \tag{10}$$

By solving the optimization problem (10) in $\mathbb{Z}$, the computed criterion J^-_{det} is a lower bound of the number of detected faults. These detected faults can be the repetition of the firing of the same fault transition. Symmetrically, the fault can be transient. Particularly, we have:

- If $\min\limits_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq 1$, then at least a fault is detected on the horizon.

Symmetrically, we can define another J^+_{det} criterion giving the maximum number by replacing the expression "min" by "max" above.

$$\begin{cases} J^{+}_{\text{det}} = \max(\boldsymbol{c}_{\text{det}}.\overline{x}) \\ s.t. \quad \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0 \end{cases} \tag{11}$$

A simpler way to express the results and to apply the solvers of linear programming is to keep the operator *min* but to replace $\boldsymbol{c}_{\text{det}}$ by $-\boldsymbol{c}_{\text{det}}$ in the expressions. Thus, the results are similar after replacing "majorant" by "minorant" and "upper bound" by "lower bound". The interpretation is that the maximum number of detected faults cannot be greater than the obtained value $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ of J^{+}_{det} solving (11) in $\mathbb{Z}$. Particularly, we have:

- If $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) = 0$, then no fault is detected.

Note that the aforementioned interpretations based on sufficient conditions do not need an additional assumption as the acyclicity of the unobservable induced subnet.

Proposition 4.1. *For an LPN with a cyclic unobservable induced subnet, if an obtained vector is not an explanation vector, then the possible explanation vectors can only give the same value or a greater value than* $\min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ *and the same value or a lower value than* $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$.

Proof. (1) We consider the following two cost functions:

$$f^{-} = \{\min(c_{det}.\overline{x}) \mid A.\overline{x} \leq b,\ \overline{x} \geq 0\ ,\ \overline{x} \in \mathbb{Z}^{k.(n+n')}\}$$
$$f^{-}_{cyc} = \{\min(c_{det}.\overline{x}) \mid A.\overline{x} \leq b,\ \overline{x} \geq 0\ ,\ \overline{x} \in \mathbb{Z}^{k.(n+n')} \text{ and } \exists \sigma \in TR^{*}_{un} \text{ such that} \Pi(\sigma) = \overline{x}\}$$

$\forall \overline{x}$ such that $A.\overline{x} \leq b,\ \overline{x} \geq 0\ ,\ \overline{x} \in \mathbb{Z}^{k.(n+n')}$, we have $c_{det}.\overline{x} \geq f^{-}$ and in particular $f^{-}_{cyc} \geq f^{-}$. The case f^{-}_{cyc} different from f^{-} is possible as we have the condition $\Pi(\sigma) = \overline{x}$.

(2) We consider the next two cost functions:

$$f^{+} = \{\max(c_{det}.\overline{x}) \mid A.\overline{x} \leq b,\ \overline{x} \geq 0\ ,\ \overline{x} \in \mathbb{Z}^{k.(n+n')}\}$$
$$f^{+}_{cyc} = \{\max(c_{det}.\overline{x}) \mid A.\overline{x} \leq b,\ \overline{x} \geq 0\ ,\ \overline{x} \in \mathbb{Z}^{k.(n+n')} \text{ and } \exists \sigma \in TR^{*}_{un} \text{ tel que} \Pi(\sigma) = \overline{x}\}$$

$\forall \overline{x}$ such that $A.\overline{x} \leq b,\ \overline{x} \geq 0,\ \overline{x} \in \mathbb{Z}^{k.(n+n')}$, we have $c_{det}.\overline{x} \leq f^{+}$, particularly $f^{+}_{cyc} \leq f^{+}$.

□

Naturally, the on-line interpretations of the above procedure tends to be more accurate if the computed count vector always describes a possible sequence (we can take the assumption of the acyclicity of the unobservable induced subnet), as it leads to the reduction in the space $S^{\mathbb{Z}}(\overline{y}) = E(\overline{y})$. This remark also holds for the next part.

Remark. If $\min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) = 0$ and $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq 1$, then we cannot conclude on the existence of a fault. Nevertheless, we can always say that the number of detected faults is between $\min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ and $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ under the liveness condition of the LPN.

The same reasoning holds if we relax the minimization and the maximization problems over $\mathbb{R}$. We can easily show that:

$$\begin{array}{ll} (1) & \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \leq \lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil \leq \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \\ (2) & \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq \lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor \geq \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \end{array} \tag{12}$$

In fact,

(1) Let us consider:

$$\begin{array}{l} f_{\mathbb{R}}^{-} = \{\min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \mid \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0, \overline{x} \in \mathbb{R}^{k.(n+n')}\} \\ f_{\mathbb{Z}}^{-} = \{\min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \mid \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0, \overline{x} \in \mathbb{Z}^{k.(n+n')}\} \end{array} \tag{13}$$

$\forall \overline{x} \in \mathbb{R}^{k.(n+n')}$ such that $\boldsymbol{A}.\overline{x} \leq \boldsymbol{b}$, $\overline{x} \geq 0$, we have $\boldsymbol{c}_{\text{det}}.\overline{x} \geq f_{\mathbb{R}}^{-}$, especially $\min_{\mathbb{Z}} \boldsymbol{c}_{\text{det}}.\overline{x} \geq f_{\mathbb{R}}^{-}$. Therefore, $\min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \leq \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$. We can also write that $\min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \leq \lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil \leq \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$

(2) Let us consider:

$$\begin{array}{l} f_{\mathbb{R}}^{+} = \{\max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \mid \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0, \overline{x} \in \mathbb{R}^{k.(n+n')}\} \\ f_{\mathbb{Z}}^{+} = \{\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \mid \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0, \overline{x} \in \mathbb{Z}^{k.(n+n')}\} \end{array} \tag{14}$$

$\forall \overline{x} \in \mathbb{R}^{k.(n+n')}$ such that $\boldsymbol{A}.\overline{x} \leq \boldsymbol{b}$, $\overline{x} \leq 0$, we have $\boldsymbol{c}_{\text{det}}.\overline{x} \leq f_{\mathbb{R}}^{+}$, in particular $\max_{\mathbb{Z}} \boldsymbol{c}_{\text{det}}.\overline{x} \leq f_{\mathbb{R}}^{+}$. Therefore, $\max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$. We can also write that $\max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq \lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor \geq \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$.

According to the two previous results, the decision of fault detection is taken by the following proposition:

Proposition 4.2. *Consider an LPN with $TR = TR_{obs} \bigcup TR_{un}$ descriped as (5). Let $TR_f \subset TR_{un}$ be the set of fault transitions. Therefore,*

- *If $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil \geq 1$, then at least a fault is detected on the horizon.*
- *If $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor = 0$, then no fault is detected.*
- *If $\min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) = 0$ and $\max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq 1$, then we cannot conclude on the existence of a fault.*

Proof. – We suppose that $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil \geq 1$. We have shown previously that $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil \leq \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$, so $\min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \geq 1$. Proposition 4.1 implies that a fault is detected.
– We suppose that $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor = 0$. We have demonstrated previously that $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor \geq \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$, so $\max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) = 0$. Consequently, we can deduce that no fault is detected, according to proposition 4.1. □

Naturally, the previous proposition holds for the two bounds $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil$ and $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor$ as well as for the integer solutions $\lceil \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rceil = \min_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ and

$\lfloor \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x}) \rfloor = \max_{\mathbb{Z}}(\boldsymbol{c}_{\text{det}}.\overline{x})$ provided by ILP. It is important to note that the accuracy of the results increases in this last case, as well as the execution time of the detection procedure. For simplicity, the relevant propositions are not presented.

Remark

If $\max(\boldsymbol{c}_{\text{det}}.\overline{x}) = +\infty$, the interpretation given by points 1 and 3 of the previous proposition can still be hold.

4.2. Fault isolation

In this section, we design a method for isolating a unique fault associated with a fault transition $(x_f)_j$. The criterion $\boldsymbol{c}_{loc,j}.\overline{x}$ enables the isolation, where the row-vector $\boldsymbol{c}_{loc,j}$ is the concatenation of a submatrix of $k.n'$ zeros relevant to TR_{obs} and the submatrix $= \begin{pmatrix} c^{<1>} & c^{<2>} & c^{<3>} & \dots & c^{<k>} \end{pmatrix}$ which is defined as follows. The relevant components $(c^{<i>})_j$ are equal to 1 for any $i = 1, ..., k$, while the other ones are null. Hence, $c^{<i>} = 0$ except $(c^{<i>})_j = 1$ for a given fault transition j and any $i = 1, ..., k$.

Following the same reasoning as the detection approach for fault isolation, we define two diagnostic indicators, $J^-_{loc}((\overline{x_f})_j)$ and $J^+_{loc}((\overline{x_f})_j)$, associated with the fault transition $(x_f)_j$ and verifying (7):

$$\begin{cases} J^-_{loc}((\overline{x_f})_j) = \min(\boldsymbol{c}_{loc,j}.\overline{x}) \\ s.t. \quad \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0 \end{cases} \quad \begin{cases} J^+_{loc}((\overline{x_f})_j) = \max(\boldsymbol{c}_{loc,j}.\overline{x}) \\ s.t. \quad \boldsymbol{A}.\overline{x} \leq \boldsymbol{b} \text{ with } \overline{x} \geq 0 \end{cases} \tag{15}$$

Fault $(x_f)_j$ will occur if it is fired at least once for the observed word w. This means that the firing sum of the fault transition $(x_f)_j$ is greater than or equal to 1 for the k observed transitions of w.

By solving the two optimizations problems (15) in $\mathbb{Z}$, we distinguish that the fault $(x_f)_j$ is detected on the horizon if $\min_{\mathbb{Z}}(\boldsymbol{c}_{loc,j}.\overline{x}) \geq 1$, and that $(x_f)_j$ surely does not occur if $\max_{\mathbb{Z}}(\boldsymbol{c}_{loc,j}.\overline{x}) = 0$.

In a similar way as for detection and by replacing $\boldsymbol{c}_{\text{det}}$ by $\boldsymbol{c}_{loc,j}$ for each fault transition $(x_f)_j$, we can demonstrate that the relaxation of the minimization and maximization problems over $\mathbb{R}$ give the following results:

$$\min_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \leq \lceil \min_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rceil \leq \min_{\mathbb{Z}}(\boldsymbol{c}_{loc,j}.\overline{x}) \tag{16}$$

$$\max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \geq \lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rfloor \geq \max_{\mathbb{Z}}(\boldsymbol{c}_{loc,j}.\overline{x}) \tag{17}$$

Therefore, we consider the following proposition for the fault isolation by problem relaxation:

Proposition 4.3. *Consider an LPN with $TR = TR_{obs} \bigcup TR_{un}$ descriped as (5). Let $(x_f)_j \in TR_f$ be a fault transition such that $TR_f \subset TR_{un}$. Therefore,*

- *If $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rceil \geq 1$ and $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rfloor$ is finite, then the fault $(x_f)_j$ presents an occurrence number between $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rceil$ and $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rfloor$.*
- *If $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rfloor = 0$, then the fault $(x_f)_j$ occurs.*

- *If* $\lceil \min_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rceil = 0$ *and* $\lfloor \max_{\mathbb{R}}(\boldsymbol{c}_{loc,j}.\overline{x}) \rfloor = +\infty$*, then the available pieces of information do not lead to a conclusion on the presence of the fault* $(x_f)_j$*.*

Proof. As aforementioned, the same results can be written by taking the integer bounds given in inequalities (16) and (17). □

5. Example with unobservable circuit and indistinguishable events

Let us consider the LPN of figure 1 containing an unobservable circuit and presenting the case of indistinguishable transitions. The chain going from transition x_3 to place p_2 describes the official behavior of the process, whereas the remaining part represents the unauthorized exploitation. Place p_2 represents the stock of parts in a manufacturing system and its marking is the number of parts. The firing of transition x_3 models the authorized input of a part and x_6 represents the legal input in the stock. However, an illegal use of the parts is expressed by the remaining subnet. The firing of transition x_5 models the illegal output of a part, while place p_1 represents its copy leading to the manufacturing of a new part but with a lower quality. Then transition x_4 describes that all the initial and new parts are stocked in place p_2, so the firings of x_4 and x_5 describe faulty situations. The supervisor system can observe the output of parts expressed by x_1 but cannot distinguish between the legal input and the unauthorized input of parts, and an image of these transitions is provided by transitions x_3 and x_4, respectively.

Let $TR_{un} = \{x_4, x_5, x_6\}$ and $TR_{obs} = \{x_1, x_2, x_3\}$. The alphabet AL is $\{a, b\}$. Label b is associated with the transition x_3 ($\Omega_b = \{x_3\}$), and the transitions of $\Omega_a = \{x_1, x_2\}$ share the same label a. Hence, $n = |TR_{un}| = 3$, $n' = |TR_{obs}| = 3$, $n'' = |AL| = 2$, $m = |P| = 5$ and $C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The set of fault transitions $TR_f \subset TR_{un}$ is $\{x_4, x_5\}$.

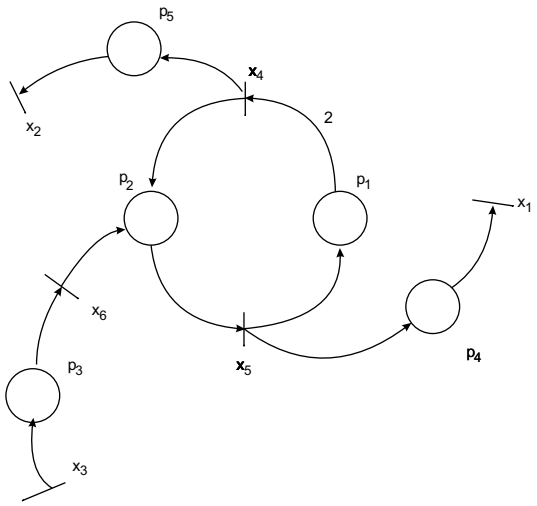

Figure 1. LPN of Example 1 with unobservable circuit and indistinguishable events

The incidence matrices W_{un} and W^-_{obs} relevant to TR_{un} and TR_{obs} are as follows:

$W_{un} = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, $W_{obs} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$ and $W^-_{obs} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

Simulation. The initial marking M is null. Remember that $M_i = M(p_i)$ is the starting marking of place p_i. The observations are defined by the sequence of label ba, which means that we have the firing of transition x_3 followed by the firing of x_1 or x_2. A realistic sequence yielding these observed transitions is $x_3x_6x_5x_1$ but x_6 and x_5 are not observed and x_1 is observed through the ambiguous label a. So, the relevant horizon is $\{1, 2\}$ and $h = 2$. We have $\overline{y} = \begin{pmatrix} (\overline{y}^{<1>})^T & (\overline{y}^{<2>})^T \end{pmatrix}^T$ with $(\overline{y}^{<1>})^T = \begin{pmatrix} 0 & 1 \end{pmatrix}$ and $(\overline{y}^{<2>})^T \begin{pmatrix} 1 & 0 \end{pmatrix}$ and the vectors to be estimated are $\overline{x_{un}} = \begin{pmatrix} (\overline{x_{un}}^{<1>})^T & (\overline{x_{un}}^{<2>})^T \end{pmatrix}^T$ and $\overline{x_{obs}} = \begin{pmatrix} (\overline{x_{obs}}^{<1>})^T & (\overline{x_{obs}}^{<2>})^T \end{pmatrix}^T$.

Detection

The row-vector $\boldsymbol{c}_{\text{det}}$ for the fault transitions x_4 and x_5 is the concatenation of a submatrix of $h.n' = 6$ zeros relevant to TR_{obs} and

$\begin{pmatrix} c^{<1>} & c^{<2>} \end{pmatrix}$ with $c^{<1>} = c^{<2>} = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$. So,

$\boldsymbol{c}_{\text{det}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$ and, the optimizations give $J^-_{\text{det}}(\{x_4, x_5\}) = 1$ and $J^+_{\text{det}}(\{x_4, x_5\}) = 3$ which show the existence of at least a fault and no more than 3 occurrences of faults (also given in the first row of the table below).

Isolation

The row-vector $\boldsymbol{c}_{loc}$ relevant to the fault transition x_4 is similar to $\boldsymbol{c}_{\text{det}}$ except that $c^{<1>} = c^{<2>} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$. The row-vector $\boldsymbol{c}_{loc}$ relevant to the fault transition x_5 is similar to $\boldsymbol{c}_{\text{det}}$ except that $c^{<1>} = c^{<2>} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$. The results are presented in the two last rows of the following table.

	$J^-()$	$J^+()$
$\{x_4, x_5\}$	1	3
$\{x_4\}$	0	1
$\{x_5\}$	0.666	2

The last row makes the isolation of the fault x_5 since the minimal criterion over the integers is 1. No conclusion can be made on the transition x_4 except that the occurrence of x_4 cannot be equal to or greater than 2.

Now, let us analyze the ambiguity of the labels by considering the isolation of the fault transition x_5. For the above C, the relaxed solutions over $\mathbb{R}$ are:

$\overline{x_{obs}} = \begin{pmatrix} 0 & 0 & 1 & 0.666 & 0.333 & 0 \end{pmatrix}$ and $\overline{x_{un}} = \begin{pmatrix} 0 & 0 & 0 & 0.333 & 0.666 & 0.333 \end{pmatrix}$ for $J^-_{loc}(x_5) = 0.666$ and

$\overline{x_{obs}} = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$ and $\overline{x_{un}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 & 1 \end{pmatrix}$ for $J^+_{loc}(x_5) = 2$. So, the relaxation over $\mathbb{R}$ leads to integer solutions for $J^+_{loc}(x_5) = 2$ but to a "fluidification" of the solutions at the iteration $< 2 >$ for $J^-_{loc}(x_5)$.

The sequence of label ba means that we have two possible cases which are now explored. We can remove the ambiguity of the label a if we modify the estimation problem such that $n'' = n' = 3$ and $C = I_{n' \text{x} n'}$. In other words, any pair of different observable transitions does not share the same label.

The case 1 (Respectively, case 2) which is the firing of transition x_3 followed by the firing of x_1 (Respectively, x_2) is built if we take $\overline{y} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$ (Respectively, $\overline{y} = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$). The results are given in the following table

($\overline{x_{obs}}$ is not given since $\overline{x_{obs}} = \overline{y}$)

	$J^-_{loc}()$	$\overline{x_{un}}$ for J^-	$J^+_{loc}()$	$\overline{x_{un}}$ for J^+
Case 1:	1	(0**0**0011)	2	(0**0**012**2**1)
Case 2:	2	(0**0**012**2**1)	2	(0**0**012**2**1)

The results are coherent as the non-ambiguous labels simplifying the problem leads to a restriction of the interval $[J^-_{loc}(x_5),\ J^+_{loc}(x_5)]$. The last solution $\overline{x_{un}}$ defined by Case 2 and J^+_{loc} corresponds to the second solution obtained for the ambiguous label a and J^+_{loc}.

6. Complexity study and comparisons

In this article, the fault detection needs two ILP problems for the observed sequence while the fault isolation is established by the resolution of at most $(2 * |TR_f|)$ ILP problems. An ILP problem is considered as an NP-hard problem and its resolution is done in an exponential time in the worst case (it can be of a polynomial complexity in particular cases). Among classical algorithms solving ILP problems, we can cite the cutting-plane method and the branch-and-bound method, where each iteration is based on a specific linear programming problem. In a similar way, the ILP problems of diagnosis can be solved more efficiently if a relaxation over $\mathbb{R}$ is made. We can use efficient standard algorithms of linear programming, which can be applied to relatively great full matrices (at least the size (100X100) can be treated with Scilab or Matlab and usual computers): The simplex algorithm is known to be efficient in practice as it has polynomial-time average-case complexity in some general cases. The modern algorithms of linear programming are polynomial in the worst case (the complexities of the ellipsoid algorithm of Khashiyan and the interior point algorithm of Karmarkar are respectively $O(n^4 \mathrm{x} L)$ and $O(n^{3.5} \mathrm{x} L)$ where n is the number of variables and L is the number of bits necessary in the storage of data.)

We now present a comparison between the proposed diagnosis approach and some approaches presented in the literature for the partially observed LPN diagnosis based on linear programming. Some work has been made to decrease the number of variables of ILP problems and to reduce the number of ILP problems to be solved. In (Cabasino et al., 2011), the authors designed a diagnosis approach based on the notion of a basis markings set $M_b(w)$ that was the set of possible reached markings by firing the sequences associated with the observation w and the possible shortest unobservable sequences coherent with w. The authors defined $(|TR_f| \times |M_b(w)|)$ ILP problems for the fault detection. The number of possible markings of $M_b(w)$ might be important, so the number of ILP problems to be solved for fault detection was considerable. Basile et al. (Basile et al., 2009) suggested an approach of fault diagnosis of a PN requiring an on-line computation of the set of possible fault events explaining the last observed event. The on-line computation consisted in solving the ILP problems formulated on a net structure and based on g-markings, which were net markings that might had negative components. The detection procedure was based on the resolution of at most $(2*|w|*|TR_f|)$ ILP problems. In (Dotoli et al., 2009), the authors proposed a diagnosis approach based on the resolution of at most $(|TR_f| + 1)$ ILP problems. The effectiveness of the approaches mentioned previously was indicated only in the particular case where the unobservable subnet of a bounded LPN was an acyclic "state machine". In fact, in this case the incidence matrix of the unobservable subnet is a totally unimodular matrix. The resolution in a polynomial time would be possible because the

solutions of the relaxed problem were integers. In this case, the proposed detection approaches were of a polynomial complexity, otherwise they were of exponential complexity.
The authors in (Mahulea et al., 2012) showed that the fluidization of a partially observed purely logic PN would produce an untimed Continuous PN (CPN) with a set of markings that were consistent with a given observation which was convex. Using this convexity property, the diagnosis states were computed solving LP problems for untimed CPNs. For fault detection, two LP problems were defined for each observed transition of the observed sequence w and for each fault transition $T_f^i \in TR_f$. Thus, for fault detection, $(2 * |w| * |TR_f|)$ LP problems had to be solved under the assumption that the $TR_{un}-$induced subnet had no spurious solutions. However, the update of the initial marking after each observed transition of the observation was necessary. Therefore, the complexity would increase with the number of observed transitions (Chouchane et al., 2018). This was why the authors in (Ru et al., 2009) took the assumption that the marking was periodically known or was periodically reset.
In this article, to avoid the procedure of estimation based on the sequential treatment of each transition of the observed sequence using loops, we generalize this procedure by introducing an approach based on a new description of the problem, not for one observed transition but for the set of transitions of the observed sequence, using matrix and vector calculations instead of loops. In fact, *"Matrix algebra is a general example of vectorisation. There loops are executed by highly tuned external libraries like BLAS. If you can figure out a way to use matrix algebra to solve your problem, you'll often get a very fast solution."*(cited in (Wickham, 2014)).

7. Numerical comparison with discrete approach based on basis markings

Let us check the numerical efficiency of our approach and compare with the discrete approach based on basis markings (Cabasino et al., 2010) on a case study presented in (Mahulea et al., 2012), which can be easily found on the web. The automated manufacturing system is shown in Figure 2 and the corresponding PN is depicted in Figure 3.

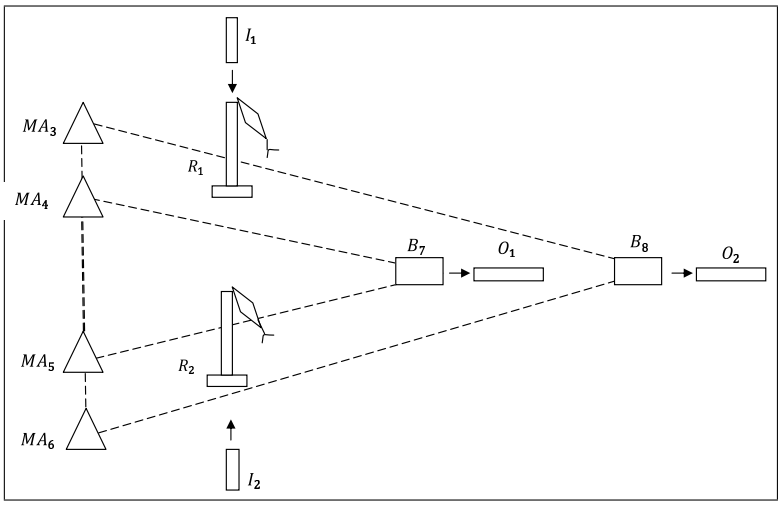

Figure 2. Manufacturing system

The plant represents a part of a large manufacturing system where MA_3, MA_4, MA_5 and MA_6 are four machines, R_1 and R_2 are two robots, B_7 and B_8 are two buffers, two inputs of parts to be processed are I_1 and I_2, and two outputs for the processed parts are O_1 and O_2. Two various types of a final product have been produced by two production lines. The two robots R_1 and R_2 modeled respectively by the transitions x_1 and x_2, take parts from two different buffers modeled by places p_1 and p_2, respectively. The four parts taken by robot R_1 are packed in couples and placed on two different conveyor belts modeled respectively by places p_3 and p_4, that follow two parallel lines at two different levels, where p_4 is located in the lowest level, and p_3 is located in the highest level.
Parts in the conveyor belts represented by place p_4 are processed by the machine modeled by transition x_7 and then put in a common buffer represented by place p_7. The bottom part of the net models similar operations.
Transitions x_3 and x_4 respectively model the output of parts from the conveyor belts modeled by p_3 and p_6 to a common buffer modeled by p_8. On the other hand, transition x_5 models the output of parts from the common buffer p_7. The system faults are modeled by the unobservable events x_9 and x_{10} consisting in some breakage in the highest conveyor belts at the beginning of the two main production lines. Nevertheless, the unobservable events x_7 and x_8 present non faulty events. To each part exiting p_7 and p_8 corresponds a new part entering p_1 and a new part entering p_2, and the process is cyclically repeated.

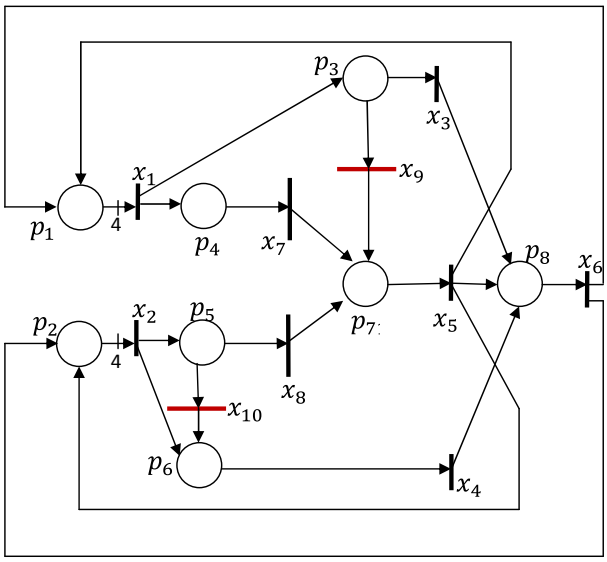

Figure 3. LPN model of the manufacturing system in Figure 2

The observable transitions of the set $TR_{obs} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ are labeled as follows: $l(x_1) = a$, $l(x_2) = b$, $l(x_3) = c$, $l(x_4) = d$, $l(x_5) = e$, $l(x_6) = a$. Let us consider the observation $w = aaaabbbbeeeeeeeeeeeeddddaaaa$ associated with the unique observed sequence $x_1x_1x_1x_1x_2x_2x_2x_2x_5x_5x_5x_5x_5x_5x_5x_5x_5x_5x_5x_4x_4x_4x_4x_1x_1x_1x_1$ from the initial marking $\begin{pmatrix} 80 & 80 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$. In Table 1, a comparative study is represented for the isolation of faults for three different approaches: an approach based on basis markings (Cabasino et al., 2010), an approach based on ILP problems resolution and our approach based on the relaxation of the ILP problem. Using Matlab, the computations are carried out on PC Intel with a clock of 1.80 Ghz,

as in (Mahulea et al., 2012). The three diagnosis states, namely N, U, and F, correspond respectively to No fault, Uncertain, and Fault state. We denote Δ_1 and Δ_2 the diagnosis state respectively of the fault transitions x_9 and x_{10}.

For the initial marking $\begin{pmatrix} 80 & 80 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$, the number of basis markings is huge, which considerably affects the execution time of the approach despite the limited number of transitions. In particular, considering one observed sequence of 28 transitions, the approach produces 34650 basis markings for just the last observed transition and takes 739 seconds (12mn). Contrariwise, the execution time of our approach is negligible (0.181 second). Moreover, the execution time of our approach does not depend on the initial marking, whereas for the discrete approach, a multiplication by 50 implies that the computation does not end after one day (Mahulea et al., 2012). Simulation also shows the effectiveness of our approach based on the resolution of non-integer LP problems compared to the approach based on the resolution of ILP problems.

Table 1. comparative table

Observation w	Basis markings approach		ILP based approach		LP based approach	
	(Δ_1, Δ_2)	*Execution time(s)*	(Δ_1, Δ_2)	*Execution time(s)*	(Δ_1, Δ_2)	*Execution time(s)*
$t_1t_1t_1t_1$	(U, N)	0.149	(U, N)	0.15	(U, N)	0.058
$t_1t_1t_1t_1t_2t_2t_2t_2$	(U, U)	5.79	(U, U)	0.15	(U, U)	0.052
$t_1t_1t_1t_1t_2t_2t_2t_5t_5t_5t_5$	(U, U)	18.10	(U, U)	0.155	(U, U)	0.0618
$t_1t_1t_1t_1t_2t_2t_2t_5t_5t_5t_5t_5t_5$	(U, U)	43.463	(U, U)	0.164	(U, U)	0.089
$t_1t_1t_1t_1t_2t_2t_2t_5t_5t_5t_5t_5t_5t_5t_5$	(F, N)	297.308	(F, N)	0.167	(F, N)	0.101
$t_1t_1t_1t_1t_2t_2t_2t_5t_5t_5t_5t_5t_5t_5t_5t_4t_4t_4$	(F, N)	227.509	(F, N)	0.168	(F, N)	0.144
$t_1t_1t_1t_1t_2t_2t_2t_5t_5t_5t_5t_5t_5t_5t_5t_5t_4t_4t_4t_1t_1t_1t_1$	(F, N)	739.524	(F, N)	0.182	(F, N)	0.181

8. Conclusion

In this paper, we have presented an estimation approach and a diagnostic technique of LPNs based on linear programming. We have considered the case where a label can be emitted by the firings of various observable transitions leading to non-determinism, and the induced unobservable LPN can contain circuits. The solution set of the obtained polyhedron includes the set of explanation vectors. The burdensome determination of any starting marking for each iteration is replaced by an algebraic development on a receding horizon. The technique is numerically efficient as the standard algorithms of linear programming can be used (remember that the Simplex algorithm is exponential in the worst case but is polynomial on the average under weak assumptions (Schrijver, 1987)). Note that the receding horizon cannot be increased infinitly as it is connected to the execution time, but in a future paper, we will generalize this study and propose an approach based on a sliding horizon which may solve this problem.

As illustrated by the example of section 7, the fault diagnosis can efficiently produce

results although the approach uses a relaxed model which neglects the firing conditions of the unobservable transitions or the integer characteristic of the count vectors. Certainly, the accuracy of the results can be improved (but with a loss of efficiency in terms of complexity) if solvers of integer linear programming are applied.

A possible perspective to extend this contribution is to introduce time, which can be made in various ways and frameworks. A first step can focus on the checking of the consistency of each obtained count vector. The technique proposed in (Declerck et al., 2017) is based on the addition of elementary time durations on the places. Knowing the consistent trajectories, the second step can improve the accuracy of the interpretations of the diagnostic in the untimed case if the probabilities of the consistent trajectories can be computed. For that purpose, a stochastic PN where a firing rate vector characterizes the transition firing delays, can be associated to an untimed PN (Ammour et al., 2018).

Another perspective is to insert this proposed approach in a general methodology where the resilience is the main criterion. It may be defined as follows: Resilience refers to the capability of systems to recover their functions after partial damage to turn a failure into a success (Zhang and van Luttervelt., 2011). An extension of this paper can enhance resilience since any action leading to the satisfaction of some roles of the system must be deduced from an estimation of its current state: Particularly, the diagnostic module provides fault states which are useful data, yielding decisions as a switch from a faulty subsystem to a redundant one or activating a plan for the emergency evacuation of victims when a part of the roads is damaged. Therefore, a natural perspective is the development of control systems exploiting faults states.

References

Ammour, R., Leclercq, E., Sanlaville, E., and Lefebvre, D. (2018). Faults prognosis using partially observed stochastic Petri-nets: an incremental approach. Discrete Event Dynamic Systems, 28(2), 247-267.

Basile, F., Chiacchio, P., and De Tommasi, G. (2009). An efficient approach for online diagnosis of discrete event systems. *IEEE Transactions on Automatic Control,* 54.4: 748-759.

Basile, F., Cabasino, M. P., and Seatzu, C. (2015). State estimation and fault diagnosis of labeled time petri net systems with unobservable transitions. *IEEE Transactions on Automatic Control,* 60.4: 997-1009.

Benveniste, A., Fabre, E., Haar, S., and Jard, C. (2003). Diagnosis of asynchronous discrete-event systems: a net unfolding approach. *IEEE Transactions on Automatic Control,* 48.5: 714-727.

Berlekamp, E. R. (1968). Algebraic coding theory. Negacyclic Codes, 207-217.

Cabasino, M. P., Giua, A., and Seatzu, C. (2010). Fault detection for discrete event systems using Petri nets with unobservable transitions. *Automatica,* 46(9), 1531-1539.

Cabasino, M. P., Giua, A., Pocci, M., and Seatzu, C. (2011). Discrete event diagnosis using labeled Petri nets. *An application to manufacturing systems. Control Engineering Practice,* 19.9: 989-1001.

Cabasino, M. P., Giua, A., and Seatzu, C. (2013). Diagnosis using labeled Petri nets with silent or undistinguishable fault events. *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* 43.2: 345-355.

Chouchane, A., Khedher, A., Nasri, O., and Kamoun, A. (2019). Diagnosis of Partially Observed Petri Net Based on Analytical Redundancy Relationships. *Asian Journal of Control.*

Declerck, P., Chouchane, A., and Bonhomme, P. (2017, April). A strategy for estimation in timed Petri nets. *In Control, Decision and Information Technologies (CoDIT), 2017 4th International Conference on* (pp. 0489-0494). IEEE.

Dotoli, M., Fanti, M. P., Mangini, A. M., and Ukovich, W. (2009). On-line fault detection in discrete event systems by Petri nets and integer linear programming. *Automatica,* 45(11), 2665-2672.

Garca, E., Rodrguez, L., Morant, F., Correcher, A., Quiles, E., and Blasco, R. (2008, June). Fault diagnosis with coloured petri nets using latent nestling method. *In Industrial Electronics,* 2008. ISIE 2008. IEEE International Symposium on (pp. 986-991). IEEE.

Genc, S., and Lafortune, S. (2007). Distributed diagnosis of place-bordered Petri nets. *IEEE Trans. Automation Science and Engineering,* 4(2), 206-219.

Hadjicostis, C. N., and Verghese, G. C. (1999). Monitoring discrete event systems using Petri net embeddings. *In International Conference on Application and Theory of Petri Nets* (pp. 188-207), Springer Berlin Heidelberg.

Kostin, A. E. (2003). Reachability analysis in T-invariant-less Petri nets. *IEEE Transactions on automatic control,* 48(6), 1019-1024.

Lefebvre, D. (2014). On-line fault diagnosis with partially observed Petri nets. *IEEE Transactions on Automatic Control,* 59.7: 1919-1924.

Li, L., and Hadjicostis, C. N. (2011). Least-cost transition firing sequence estimation in labeled Petri nets with unobservable transitions. *IEEE Transactions on Automation Science and Engineering,* 8(2), 394-403.

Mahulea, C., Seatzu, C., Cabasino, M. P., and Silva, M. (2012). Fault diagnosis of discrete-event systems using continuous Petri nets. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* 42.4: 970-984.

Portinale, L. (1995, October). Petri net reachability analysis meets model-based diagnostic problem solving. *In Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on* (Vol. 3, pp. 2712-2717). IEEE.

Ramirez-Trevino, A., Ruiz-Beltran, E., Aramburo-Lizarraga, J., and Lopez-Mellado, E. (2012). Structural diagnosability of DES and design of reduced Petri net diagnosers. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* 42(2), 416-429.

Ran, N., Wang, S., Su, H., and Wang, C. (2017). Fault Diagnosis for Discrete Event Systems Modeled By Bounded Petri Nets. *Asian Journal of Control.*

Ru, Y., and Hadjicostis, C. N. (2009). Bounds on the number of markings consistent with label observations in Petri nets. *IEEE Transactions on Automation Science and Engineering,* 6(2), 334-344

Ruiz-Beltran, E., Lopez-Mellado, E., and Ramirez-Trevino.A. (2006). Fault diagnosis based on Petri net reduced models," *In Electrical and Electronics Engineering,* 2006 3rd International Conference on (pp. 1-5). IEEE.

A. Schrijver (1987). Schrijver, A. (1998). Theory of linear and integer programming. *John Wiley and Sons.*

Wang, X., Mahulea, C., and Silva, M,. (2015). Diagnosis of Time Petri Nets Using Fault Diagnosis Graph. *IEEE Transactions on Automatic Control,* 60.9: 2321-2335.

Wickham, H. (2014). Advanced r. Chapman and Hall/CRC.

Wu, Y., and Hadjicostis, C. N. (2005). Algebraic approaches for fault identification in discrete-event systems. *IEEE Transactions on Automatic Control,* 50.12: 2048-2055.

Yang, X., and Chen, L. J. (2009). Design and fault diagnosis of Petri net controllers for Petri nets with uncontrollable and unobservable transitions. *Journal of Manufacturing Systems,* 28.1: 17-22.

Zhang, W.J., and van Luttervelt, C.A. (2011). Toward a resilient manufacturing system, *CIRP Annals,* Volume 60, Issue 1, Pages 469-472, ISSN 0007-8506.