On-line monitoring of large Petri Net models under partial observation

George Jiroveanu, René K. Boel and Behzad Bordbar

Abstract

This paper deals with the on-line monitoring of large systems modeled as Petri Nets under partial observation. The plant observation is given by a subset of transitions whose occurrence is (always) acknowledged by emitting a label received by the monitoring agent at the time of the occurrence. Other transitions not in this subset are silent (unobservable). Usually on-line applications require the computation of how the system has *evolved* from the last known (or estimated) marking(s) by enumerating the set of *all* the explanations of the observation received by the monitoring agent, i.e. the set of all allowable traces, such that the execution of these traces from the initial marking would generate the sequence of observed labels in the correct order. This can be accomplished by a forward search algorithm starting from the initial marking. However, the application of forward search techniques to large systems has several disadvantages. Firstly, the set of current allowable markings of the system can be large. Hence, its enumeration can be computationally demanding. Secondly, forward search techniques require knowing the exact initial marking, which can be a problem in case of systems with uncertain initial marking e.g. when only a lower bound on the initial marking is known. To alleviate these drawbacks, we propose a backward search method, which, starting from observation(s), enumerates a subset of explanations called the *set of minimal explanations*. The set of markings that are reached from the initial marking firing minimal explanations has the property that its unobservable reach (the markings obtained by firing legal, unobservable strings from any of its marking) is equal to the entire set of current estimated markings. Moreover, the faults are typically not predictable i.e. at every reachable marking there is at least one non-fault transition that is enabled. Making this assumption that the faults are not predictable allows us to conclude that the set of minimal explanations obtained via a reduced observer analysis detects the occurrence of all faults that must have happened for sure according to the complete set of explanations. Furthermore, the presented approach can deal with Petri Nets with an uncertain initial marking, which is a common situation in a distributed setting. In this case, local components modeled by Petri Nets and supervised by local agents interact unobservably by exchanging tokens via common places.

G. Jiroveanu is with Romanian Power Grid Company - Transelectrica SA, Brestei 5, Craiova 200581, DJ, Romania e-mail: george.jiroveanu@transelectrica.ro. R.K. Boel is with EESA - SYSTeMS Research Group, University of Ghent, Tehnologiepark 914, Zwijnaarde 9052, Belgium e-mail: rene.boel@ugent.be. B. Bordbar is with The School of Computer Science, University of Birmingham, Edgbaston, Birmingham B152TT, United Kingdom e-mail: b.bordbar@cs.bham.ac.uk.

I. Motivation and introduction

This paper deals with model based approaches to the centralized estimation of the states of large plants. We assume that the plant evolves over time satisfying constraints expressed by an abstract, discrete event dynamical systems modeled via Petri Nets (PNs). The set of transitions in the PN, which represents events of the physical plant, is partitioned into two disjoint subsets: observable and unobservable events. We assume that the occurrence of an observable event is always reported (a label is emitted) correctly to the supervisory agent whereas the occurrence of an unobservable events is never reported.

The plant monitoring at any time θ requires knowledge of the PN model of the plant, and knowledge of the ordered sequence of observable labels that have been recorded up to time θ. State observers combine this model information with the on-line plant observation in order to derive the set of possible current states the plant can be in, and the set of traces the plant model can have executed from the last known or estimated state(s) up to the current time θ.

The monitoring of any Discrete Event System (DES) under partial observation requires typically the implementation of an observer automaton [MBL00] that is used then for on-line applications like supervisory control. An observer automaton for a PN model is simply an automaton whose set of events is represented by the set of labels of the observable transitions of the PN model. The legal traces in the observer automaton are strings of labels that can be generated by the plant. A state of the observer-automaton stores all the markings of the PN model that can be reached from the initial marking of the PN model by firing traces that would generate the same observation as the corresponding sequence of events in the automaton.

When monitoring systems under partial observation the use of a classical (off-line derived) observer-automaton is hardly possible because of the high spatial complexity (e.g. exponential in the number of places for a DES modeled as an automaton [OW90]). Moreover any change in the plant structure requires the recalculation of the off-line observer-automaton.

A natural solution for the monitoring of PN models under partial observation is to construct on-line the branch of the off-line observer-automaton that corresponds to the received observation. This simply means that after each observation generated by the plant we calculate the set of markings the plant can be in. Thus a state of an on-line classical observer-automaton (`CO`) includes all the possible states (markings) the plant can be in after observing a string of labels. However the on-line construction of the branch of the `CO` that corresponds to the received observation may not be feasible when monitoring large PN because the set of estimated markings can be huge and the calculation of the set of markings that correspond to the current state of the on-line `CO` can be computationally prohibitive.

To overcome this limitation we propose the on-line construction of a reduced observer automaton (`RO`) that contains in a given state fewer markings than the on-line `CO`. The idea is simple, instead of computing all the possible markings the plant can be in after observing a string of labels, we compute a subset of possible current markings (a set of basis markings [GCS05]) such that the rest of the possible current markings that are not computed can be reached by firing legal strings of unobservable transitions starting from a marking that belongs to the subset of markings generated by `RO`.

We show that this subset of possible current markings can be obtained by generating minimal explanations of the received observation [JB04] via backwards induction starting form an observable transition whose occurrence would have emitted the latest observed event. Methods based on backward calculations for PNs were proposed in [LA94], [SJ94], [CKV95] for diagnosis purpose, in [NAH+98], [AIN00], [FRSB02], [DRvB04] for model checking, and in [GS02], [GCS05] for state estimation of a PN model with uncertain initial marking.

Backward search operates as follows. When the first observation (the label of some event occurrence) is received by the monitoring agent, it determines the minimal (partial) marking required by an observable transition (whose occurrence would have emitted the observed label) to fire. Unobservable transitions are recursively backfired (removing tokens from the output places and adding tokens to the input places of the backfired transition) until either we obtain backwards a marking that is smaller than or equal to the initial marking (in this case a minimal explanation is obtained) or a decision to stop the backward search is evaluated true. The set of minimal explanations enumerates the subset of markings that must be considered in the current state of the reduced observer `RO`.

The computational complexity of the algorithms to derive the on-line `CO` respectively the on-line `RO` can not be compared. The reason is twofold. First of all the two algorithms (the forward search respectively the backward search) explore different state spaces. Secondly the `RO` identifies states in its state space with only a small subset of the complete set of possible current markings that identify states in the state space of the `CO`. However the price paid for this simplification of the definition of individual states in `RO` is that the number of states required by `RO` may be larger than for `CO`. Nevertheless we demonstrate in this paper the following advantages of `RO`. Firstly the computational complexity of the backward search depends on the size of the largest sub-net of the PN model that contains only unobservable transitions and on the degree on non-determinism of the observation labeling function (i.e. the number of observable transitions that emit the same label when they are executed). Secondly the backward calculations can be made even though the initial marking is partially known (i.e. one only knows a lower bound on the marking of some places). This is a typical case in a distributed setting when components (modeled as PNs) interact via shared places [GL03],[FBHJ05] where the interactions are unobservable, i.e. tokens can unobservably exit from one component and unobservably enter another component [JB05].

A special observer-automaton is designed in [SSL$^+$95] in order to detect faults in a plant. Given the observation generated by the plant a diagnoser-automaton answers the question whether a fault event happened in the plant or not. A fault *may have occurred* if there is at least one allowed trace leading from an initial state to any plant state that is possible according to the current state of the observer automaton. If all the traces from a possible initial state of the plant to a possible current state of the plant contain the fault, then the fault *must have occurred* for sure. Based on the set of minimal explanations of the observation generated by the plant we design in this paper a reduced diagnoser having the property that any fault that is detected by the classical diagnoser that for sure occurred in the plant is also detected by the reduced diagnoser to have happened for sure.

It should be noted that the claims in this paper do not guarantee that a fault that occurred will indeed be detected. This diagnosability property [SSL$^+$95] would require strong assumptions on the plant model, assumptions that are probably not verifiable for the large plants we consider. The on-line observers designed in this paper will detect a fault with an observable effect provided the explanation of this observable effect must include the fault. However in general no off-line method can be devised to state beforehand whether this diagnosability will indeed be satisfied or not, without a computational effort that is larger than the effort required for the on-line fault detection itself.

The paper is organized as follows. Section II introduces the mathematical notation and the preliminary definitions used throughout the paper. Then in Section III we present the monitoring of large PNs models under partial observation and we provide an algorithm to derive a reduced observer based on a backward search. In Section IV we formalize the diagnosis problem and we show how the reduced observer can be used to derive the on-line plant diagnosis of faults that must have occurred for sure. The paper is concluded in Section V with final remarks and future work.

II. Preliminaries

A. Sets and relations

Let X and Y be sets. We write $X \subseteq Y$ if X is a subset of Y, including the case $X = Y$. $X \subset Y$ denotes that $X \subseteq Y$ and $X \neq Y$. $X \setminus Y$ denotes the set of elements of X that do not belong to Y. $\mid X \mid$ denotes the cardinality of X and $Pwr(X)$ is the power set of X, that is, the set of all subsets of X. Given a function $f : X \rightarrow Y$ and $A \subseteq X$ then $f(A) = \bigcup_{x \in A} f(x)$. $\mathbb{N}$ denotes the set of natural numbers including 0. $\mathbb{N}_+$ denotes the set of natural numbers excepting 0. Given two vectors A, B of dimension m, we write $A \leq B$, if for $q = 1, \ldots, m$, $A[q] \leq B[q]$. $A < B$ means $A \leq B$ and $\exists q$ s.t. $A[q] < B[q]$.

A set X is a collection of distinct elements. Given a non-empty set X and a function $\mu : X \rightarrow \mathbb{N}$ we say that X_μ is a multi-set over X where $X_\mu = \{(x, \mu(x)) \mid x \in X\}$ and μ represents the number of appearances of x in X_μ. Thus a set can be understood as a multi-set that has no repeated elements.

Let $\preceq$ be binary relation on X. $\preceq\subseteq X \times X$ is a partial order relation on X if $i)$ $\preceq$ is reflexive, $ii)$ $\preceq$ is transitive, and $iii)$ $\preceq$ is antisymmetric $(\forall x, x' \in X)$ $(x \preceq x') \wedge (x' \preceq x) \Rightarrow (x = x')$. If $\forall x, x' \in X$ either $x \preceq x'$ or $x' \preceq x$ then $\preceq$ is a total order on X. Denote by $(X, \preceq)$ a partial order relation $\preceq$ on a nonempty set X. Then $\mathtt{max}_{\preceq}(X)$ and $\mathtt{min}_{\preceq}(X)$ denote the set of maximal respectively minimal elements of X w.r.t. $\preceq$, that is $\mathtt{max}_{\preceq}(X) = \{x \in X \mid (x' \in X \wedge x \preceq x') \Rightarrow x' = x\}$ and $\mathtt{min}_{\preceq}(X) = \{x \in X \mid (x' \in X \wedge x' \preceq x) \Rightarrow x' = x\}$.

B. Petri Nets - notation, definitions, and properties

Definition 1: A Petri Net is a structure $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ where: $i)$ $\mathcal{P}$ denotes the finite set of places, $ii)$ $\mathcal{T}$ denotes the finite set of transitions such that $\mathcal{P} \cap \mathcal{T} = \emptyset$, and $iii)$ $F \subseteq (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P})$ is the incidence (flow) relation that specifies the arcs from places to transitions and from transitions to places. F can be represented as a pair of functions $Pre : \mathcal{P} \times \mathcal{T} \rightarrow \{0, 1\}$ and $Post : \mathcal{T} \times \mathcal{P} \rightarrow \{0, 1\}$.

Denote $\mathcal{X} = \mathcal{P} \cup \mathcal{T}$. Then for $x \in \mathcal{X}$ we use the standard notations $x^{\bullet} = \{y \in \mathcal{X} \mid xFy\}$, ${}^{\bullet}x = \{y \in \mathcal{X} \mid yFx\}$, $X^{\bullet} = \bigcup_{x \in X} x^{\bullet}$, and ${}^{\bullet}X = \bigcup_{x \in X} {}^{\bullet}x$.

The incidence relation F can also be represented in a matrix form, with dimension $\mid P \mid \times \mid T \mid$, having a -1 in the (i, j)-th element if $Pre(p_i, t_j) = 1$, a 1 in the (i, j)-th element if $Post(t_j, p_i) = 1$, and a 0 everywhere else.

A *marking* M of a PN $\mathcal{N}$ is represented by a $\mid \mathcal{P} \mid$-vector that assigns to each place p of $\mathcal{P}$ a non-negative number of tokens $M : \mathcal{P} \rightarrow \mathbb{N}$.

A PN system is a pair $\langle \mathcal{N}, M_0 \rangle$ where $\mathcal{N}$ is a connected graph having at least one place and one transition and M_0 is a marking of $\mathcal{N}$ called the initial marking.

In the following we treat a marking also as a multi-set $M = \{(p, M(p)) \mid p \in \mathcal{P}$ and $M(p) \neq 0\}$ where $M(p)$ is the number of tokens present in p in the marking M ($M(p)$ stands for $\mu(p)$ when talking about a marking seen as a multi-set of tokens).

Given a PN $\mathcal{N}$ and a marking M, a transition $t \in \mathcal{T}$ is enabled in M if $\forall p \in {}^{\bullet}t$, $M(p) \geq Pre(p, t)$. Denote by $Enbl(M)$ the set of all the enabled transitions in the marking M. An enabled transition $t \in Enbl(M)$ in a marking M fires at M and produces the marking M', that is $M' = M - Pre(\cdot, t) + Post(t, \cdot)$, where abusing notation $Pre(\cdot, t)$ and $Post(t, \cdot)$ are the $\mid \mathcal{P} \mid$-vectors whose co-ordinate p is $Pre(p, t)$ respectively $Post(t, p)$.

In the following we use the notation $M \xrightarrow{t} M'$ for the firing of a transition t transforming the marking (state) of the PN from M to the new marking M'. A legal trace τ in the PN system $\langle \mathcal{N}, M_0 \rangle$ is defined as $\tau = M_0 \xrightarrow{t_1} M_1 \xrightarrow{t_2} \ldots \xrightarrow{t_\upsilon} M_\upsilon$ where inductively for $\iota = 1, 2, \ldots, \upsilon$, $M_{\iota-1} \geq Pre(\cdot, t_\iota)$. $M_0 \xrightarrow{\tau} M_\upsilon$ denotes that the enabling conditions are satisfied so that τ can be executed legally, and when τ fires at M_0 yields M_υ.

Given a PN system $\langle \mathcal{N}, M_0 \rangle$ the set of all legal traces in $\langle \mathcal{N}, M_0 \rangle$ is denoted by $\mathcal{L}_{\mathcal{N}}(M_0) = \left\{ \tau \mid M_0 \xrightarrow{\tau} M \right\}$ while the set of reachable markings is denoted by $\mathcal{R}_{\mathcal{N}}(M_0) = \left\{ M \mid \exists \tau \in \mathcal{L}_{\mathcal{N}}(M_0) \text{ s.t. } M_0 \xrightarrow{\tau} M \right\}$.

Consider a legal trace $\sigma \in \mathcal{L}_{\mathcal{N}}(M_0)$. The Parikh vector associated with σ is denoted $\overrightarrow{\sigma}$ and is a $\mid \mathcal{T} \mid - vector$ whose ι-th element corresponding to transition $t_\iota \in \mathcal{T}$ is given by $\mu_\sigma(t_\iota)$ that is the number of appearances of t_ι in the legal trace σ.

Lemma 1 (marking equation): If $M_0 \xrightarrow{\sigma} M$ then the following Marking Equation holds:

$$M_0 + F \cdot \overrightarrow{\sigma} = M \tag{1}$$

(with the incidence relation F expressed in a matrix representation).

Notice that in a general PN $\mathcal{N}$, the marking equation is a necessary but not a sufficient condition for checking if a marking M is reachable from M_0 by firing a trace σ. However if $\mathcal{N}$ is acyclic then the marking equation is a necessary and sufficient condition for the reachability problem [Mur89].

Consider two PNs $\mathcal{N}_1 = (\mathcal{P}_1, \mathcal{T}_1, F_1)$ and $\mathcal{N}_2 = (\mathcal{P}_2, \mathcal{T}_2, F_2)$. Then $\langle \mathcal{N}_1, M_{0_1} \rangle$ is called a sub-net of $\langle \mathcal{N}_2, M_{0_2} \rangle$ if: $i)$ $\mathcal{P}_1 \subseteq \mathcal{P}_2$; $ii)$ $\mathcal{T}_1 \subseteq \mathcal{T}_2$; $iii)$ $Pre_1 = Pre_2 \mid_{\mathcal{P}_1 \times \mathcal{T}_1}$; $iv)$ $Post_1 = Post_2 \mid_{\mathcal{T}_1 \times \mathcal{P}_1}$; and $v)$ $M_{0_1} = M_{0_2} \mid_{\mathcal{P}_1}$.

Conditions $i) - iv)$ state that $\mathcal{N}_1$ is a sub-graph of $\mathcal{N}_2$ where conditions $iii)$ and $iv)$ state that Pre_1 and $Post_1$ are the restriction of Pre_2, respectively $Post_2$ to the domains $\mathcal{P}_1 \times \mathcal{T}_1$, respectively $\mathcal{T}_1 \times \mathcal{P}_1$. Condition $v)$ states that the initial marking M_{0_1} is the restriction of the marking M_{0_2} to the places $\mathcal{P}_1$.

Definition 2 ([Mur89]): Given a PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ a subset of places $\mathcal{P}' \subseteq \mathcal{P}$ is a trap, respectively a siphon if $\mathcal{P}'^{\bullet} \subseteq {}^{\bullet}\mathcal{P}'$, respectively ${}^{\bullet}\mathcal{P}' \subseteq \mathcal{P}'^{\bullet}$.

A trap has the property that if it is marked (i.e. it has at least one token) under some marking, then it remains marked under each successor marking. A siphon has the property that if it has no token under some marking, then it remains token free under each successor marking.

A path of a PN $\mathcal{N}$ is a non-empty sequence $\wp = x_1 \ldots x_\upsilon$ of nodes that satisfies $(x_1, x_2), \ldots, (x_{\upsilon-1}, x_\upsilon) \in F$. A path $\wp = x_1 \ldots x_\upsilon$ is said to lead from x_1 to x_υ. A path $\wp$ leading from a node x_ι to a node x_υ is a circuit if no element occurs more than once in it and $(x_\upsilon, x_\iota) \in F$. Notice that a sequence containing one element is a path but not a circuit since $(x, x) \notin F$.

Definition 3 ([Mur89]): A PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ is called:

- *trap-circuit PN if the set of places in every directed circuit is a trap*
- *siphon-circuit PN if the set of places in every directed circuit is a siphon.*

From the definition of firing, it is straightforward to infer the following lemma.

Lemma 2 (monotonicity): Given the PN systems $\langle \mathcal{N}, M_0 \rangle$ and $\langle \mathcal{N}, M_0' \rangle$ such that $M_0 \leq M_0'$ then $\mathcal{L}_{\mathcal{N}}(M_0) \subseteq \mathcal{L}_{\mathcal{N}}(M_0')$.

Denote by $\mathcal{T}^*$ the Kleene closure of the set $\mathcal{T}$ i.e. the set of all traces of elements of $\mathcal{T}$ of arbitrary length, including the empty trace ϵ. Then let $\sigma \in \mathcal{L}_{\mathcal{N}}(M_0) \subseteq \mathcal{T}^*$ and $\mathcal{T}' \subset \mathcal{T}$. The projection $\Pi_{\mathcal{T}'} : \mathcal{L}_{\mathcal{N}}(M_0) \rightarrow \mathcal{T}'^*$ is defined as: $i)$ $\Pi_{\mathcal{T}'}(\epsilon) = \epsilon$; $ii)$ $\Pi_{\mathcal{T}'}(t) = t$ if $t \in \mathcal{T}'$; $iii)$ $\Pi_{\mathcal{T}'}(t) = \epsilon$ if $t \in \mathcal{T} \setminus \mathcal{T}'$; and $iv)$ $\Pi_{\mathcal{T}'}(\sigma t) = \Pi_{\mathcal{T}'}(\sigma)\Pi_{\mathcal{T}'}(t)$ for $\sigma \in \mathcal{L}_{\mathcal{N}}(M_0)$ and $t \in \mathcal{T}$.

A PN $\langle \mathcal{N}, M_0 \rangle$ is bounded if for every place $p \in \mathcal{P}$ there is a natural number $\mathtt{K} \in \mathbb{N}_+$ s.t. $M(p) \leq \mathtt{K}$ for any $M \in \mathcal{R}_{\mathcal{N}}(M_0)$. Given $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ and $\mathcal{T}' \subseteq \mathcal{T}$ then $\langle \mathcal{N}, M_0 \rangle$ is bounded w.r.t. $\mathcal{T}'$ if $\forall \tau \in \mathcal{T}'^*$ we have that: $(M_0 \xrightarrow{\tau} M)$ and $(M_0 \leq M) \Rightarrow (M_0 = M)$.

C. Occurrence Nets and Net Unfolding

In this section, we shall present a method of monitoring large PNs based on partial orders. This allows us to provide a rigorous mathematical definition of a minimal explanation in terms of minimal configurations in a net unfolding.

The complexity of the PN reachability analysis has been proven to be EXPSPACE-hard in the general case [Lip76],[Kos82]. This is because in the standard reachability algorithm (the Karp-Miller algorithm [KM69]) all the possible interleavings of the concurrent transitions are considered.

The proposed methods for reducing the state-space explosion problem are based on the observation that for reachability analysis not all interleavings of a given set of concurrent transitions need to be considered. Various methods that avoid considering all interleavings of concurrent transitions have been proposed, among others *stubborn sets* [Val90], *persistent sets* [GW93] and *net unfoldings* [Eng91],[McM92], [Esp94].

The unfolding of a PN is an occurrence net (i.e. a PN without cycles) that is behaviorally equivalent with the original net. Unfoldings are usually *infinite* nets since the set $\mathcal{L}_{\mathcal{N}}(M_0)$ of legal traces is usually infinite. However it is always possible to construct a finite *prefix* of the unfolding which captures its entire behavior [McM92]. The *prefix* of the unfolding has the property that it contains all the reachable states of the whole unfolding, and being finite, it can be handled by a computer. Initial prefixes can be constructed such that they are never larger and in general a lot smaller than the state space of the original PN [ERV96]. The unfolding approach is a powerful technique for attacking the state explosion problem for PN models whose degree of concurrency is high compared to their degree of (forward) branching. The unfolding method reduces the cost of the analysis from exponential in the size of Petri Net to polynomial dependence on its size.

Two nodes (places or transitions), a and b of a PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ are in conflict, denoted $a \sharp b$ if there are distinct transitions $t, t' \in \mathcal{T}$ such that $^\bullet t \cap {}^\bullet t' \neq \emptyset$ and $a \trianglelefteq t$ and $b \trianglelefteq t'$ where $\trianglelefteq$ denotes the reflexive transitive closure of F. If $\mathcal{N}$ is acyclic then $\trianglelefteq$ is a partial order.

Definition 4: An occurrence net is a net $\mathfrak{O} = (B, E, G)$ *such that:*

1) $\mathfrak{O}$ *is acyclic*
2) *every node* $a \in B \cup E$ *has a finite number of predecessors, i.e.* $\mid \{b : a \preceq b\} \mid < \infty$
3) $\mathfrak{O}$ *has no backward conflicts, i.e.* $\forall b \in B : \mid {}^\bullet b \mid \leq 1$

where $\preceq$ *denotes the reflexive transitive closure of* G.

In an occurrence net two nodes $a, b \in (B \cup E) \times (B \cup E)$ are concurrent, denoted $a \| b$, if neither $a \sharp b$ nor $a \preceq b$ nor $b \preceq a$. In the following B is referred as the set of *conditions*, E is the set of *events*, and $\mathtt{min}_{\preceq}(\mathfrak{O})$ denotes the set of minimal elements of $\mathfrak{O}$ w.r.t. $\preceq$.

Definition 5: A homomorphism from an occurrence net $\mathfrak{O} = (B, E, G)$ *to a PN system* $\langle \mathcal{N}, M_0 \rangle$ *is a mapping* $\phi : B \cup E \to \mathcal{P} \cup \mathcal{T}$ *such that:*

1) $\phi(B) \subseteq \mathcal{P}$ *and* $\phi(E) \subseteq \mathcal{T}$
2) $\forall e \in E$*, the restriction of* ϕ *to* ${}^\bullet e$ *is a bijection between* ${}^\bullet e$ *and* ${}^\bullet \phi(e)$
3) $\forall e \in E$*, the restriction of* ϕ *to* $e^\bullet$ *is a bijection between* $e^\bullet$ *and* $\phi(e)^\bullet$
4) *the restriction of* ϕ *to* $\mathtt{min}_{\preceq}(\mathfrak{O})$ *is a bijection between* $\mathtt{min}_{\preceq}(\mathfrak{O})$ *and* M_0
5) $\forall e, e' \in E : ({}^\bullet e = {}^\bullet e') \wedge (\phi(e) = \phi(e')) \Rightarrow e = e'$.

Definition 6: A branching process $\mathfrak{B}$ *of a PN* $\langle \mathcal{N}, M_0 \rangle$ *is a pair* $\mathfrak{B} = (\mathfrak{O}, \phi)$ *where* $\mathfrak{O}$ *is an occurrence net and* ϕ *is a homomorphism* $\phi : \mathfrak{O} \to \mathcal{N}$.

Definition 7: Given a PN $\langle \mathcal{N}, M_0 \rangle$ *and two branching processes* $\mathfrak{B}, \mathfrak{B}'$ *of PN* $\langle \mathcal{N}, M_0 \rangle$ *then* $\mathfrak{B}' \sqsubseteq \mathfrak{B}$ *if there exists an injective homomorphism* $\psi : \mathfrak{O}' \to \mathfrak{O}$ *s.t.* $\varphi(\mathtt{min}(\mathfrak{O}')) = \mathtt{min}(\mathfrak{O})$ *and* $\phi \circ \varphi = \phi'$.

There exists (up to an isomorphism) an unique maximum branching process (w.r.t. $\sqsubseteq$) that is the unfolding of $\langle \mathcal{N}, M_0 \rangle$ and is denoted $\mathcal{U}_{\mathcal{N}}(M_0)$ [McM92],[Esp94].

Definition 8: A configuration $C = (B_C, E_C, G_C)$ *in the occurrence net* $\mathfrak{O}$ *is defined as follows:*

1) C *is a proper sub-net of* $\mathfrak{O}$ *(*$C \subseteq \mathfrak{O}$*)*
2) C *is conflict free, i.e.* $\forall a, b \in (B_C \cup E_C) \times (B_C \cup E_C) \Rightarrow \neg(a \sharp b)$
3) C *is causally upward-closed, i.e.* $\forall b \in B_C \cup E_C : a \in B \cup E$ *and* $a \preceq b \Rightarrow a \in B_C \cup E_C$
4) $\mathtt{min}_{\preceq}(C) = \mathtt{min}_{\preceq}(\mathfrak{O})$
5) *and* G_C *is the restriction of* G *to* $(B_C \cup E_C) \times (E_C \cup B_C)$

Denote by $C^\perp = (B_{C^\perp}, E_{C^\perp}, G_{C^\perp})$ the initial configuration of the occurrence net $\mathfrak{O}$. $B_{C^\perp} = \{b \in B : {}^\bullet b = \emptyset\}$ is the set of condition-nodes in $\mathfrak{O}$ that correspond to the places that contain a token in initial marking ($B_{C^\perp} = \mathtt{min}_{\preceq}(\mathfrak{O})$) and $E_{C^\perp} = \emptyset$.

For a configuration C in $\mathfrak{O}$ denote by $CUT(C)$ the maximal (w.r.t. set inclusion) set of conditions in C that have no successors in C, i.e. $CUT(C) = ((\bigcup_{e \in E_C} e^\bullet) \cup (\mathtt{min}_{\preceq}(O)) \setminus (\bigcup_{e \in E_C} {}^\bullet e)$ Denote by $mark(C)$ the marking in $\mathcal{N}$ that corresponds to $CUT(C)$ ($mark(C) = \phi(CUT(C))$). Obviously we have that $CUT(C^\perp) = B_{C^\perp} = \mathtt{min}_{\preceq}(\mathfrak{O})$ and $mark(C^\perp) = \phi(CUT(C^\perp)) = M_0$ (where a marking is seen as a multi-set of tokens).

Denote by $Enbl(C)$ the set of transitions that are enabled in $\mathcal{N}$ from the marking $mark(C)$. For an enabled transition $t \in Enbl(C)$ append to C an event e s.t. $t = \phi(e)$. We say that C is extended by e and denote the configuration that is thus obtained by $C' = C \odot e$. We have that $C' = (B_{C'}, E_{C'}, G_{C'})$ where $B_{C'} = B_C \cup \{e^{\bullet}\}$ and $E_{C'} = E_C \cup \{e\}$.

Consider two configurations C and C' s.t. C' is obtained from C by appending the events $e_1, \ldots, e_q$ ($C' = C \odot e_1 \odot \ldots \odot e_q$). Then C is a proper sub-net of C' and we say that C is a prefix of C'. This is denoted $C \sqsubset C'$. Denote by $\mathcal{C}$ the set of all the configurations in $\mathcal{U}_{\mathcal{N}}(M_0)$.

Definition 9: Given the unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$ of a PN $\langle \mathcal{N}, M_0 \rangle$ then $\underline{C}(e) = (B_{\underline{C}(e)}, E_{\underline{C}(e)}, E_{\underline{C}(e)})$ is the minimal configuration that explains the execution of e if $E_{\underline{C}(e)} = \left\{e' \in E : e' \preceq_{\underline{C}(e)} e\right\}$.

As already mentioned unfoldings are usually *infinite* nets. As shown in [McM92] it is always possible to construct a finite *initial prefix* of the unfolding which captures its entire behavior by deriving the set of cut-off events. An event e is a cut-off event in the unfolding if there exists another event e' such that $i)$ $\phi(\underline{C}(e)) = \phi(\underline{C}(e'))$ and $ii)$ $\underline{C}(e') \sqsubset \underline{C}(e)$. The idea is that the continuations of $\mathcal{U}_{\mathcal{N}}(M_0)$ from $\underline{C}(e)$ and $\underline{C}(e')$ are isomorphic.

Definition 10: Given a partially ordered set $(\Sigma, \preceq)$, the string $s = a_1 a_2 \ldots a_\upsilon$ is a linearization of $(\Sigma, \preceq)$ if $\upsilon = \mid \Sigma \mid$ and $\forall a_\iota, a_\lambda \in \Sigma$ then $i)$ $a_\iota = a_\lambda \Rightarrow \iota = \lambda$ and $ii)$ for $\iota \neq \lambda$, if $a_\iota \preceq a_\lambda$ then $\iota < \lambda$. In words, s is a string obtained considering all the symbols of the set Σ, where each symbol appears once in the string s and for any two different elements of Σ s.t. $a_\iota \preceq a_\lambda$ then a_ι is considered in s before a_λ. Denote by $\langle \Sigma \rangle_{\preceq}$ the set of all the strings s that are linearizations of $(\Sigma, \preceq)$.

Consider a configuration of $C = (B_C, E_C, G_C)$ in the net unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$. $\preceq_C$ is the reflexive transitive closure of G_C, i.e. a partial order defined onto the set of elements $B_C \cup E_C$. Denote by $\langle E_C \rangle_{\preceq_C}$ the set of all the linearizations of the partial ordered set $(E_C, \preceq_C)$. Strings that belong to $\langle E_C \rangle_{\preceq_C}$ can be obtained one from the other by shuffling (interleaving) the order of the concurrent events.

Let σ be a linearization of $(E_C, \preceq_C)$, i.e. $\sigma \in \langle E_C \rangle_{\preceq_C}$. We have that $\tau = \phi(\sigma)$ is a legal trace in $\langle \mathcal{N}, M_0 \rangle$ where for $\sigma = e_1 \ldots e_k$, $\phi(\sigma) = \phi(e_1) \ldots \phi(e_k)$. Denote by $\mathcal{L}_{\mathcal{N}}(C)$ the set of all the traces that are obtained via ϕ from the linearizations of the partial ordered set $(E_C, \preceq_C)$.

We have that all the traces in the set $\mathcal{L}_{\mathcal{N}}(C)$ have the same Parikh vector, i.e. $\forall \tau, \tau' \in \mathcal{L}_{\mathcal{N}}(C)$, $\vec{\tau} = \vec{\tau'}$. Thus a configuration C compactly represents a set of traces that are equivalent under the interleaving of concurrent events. Since $\mathcal{L}_{\mathcal{N}}(M_0) = \bigcup_{C \in \mathcal{C}} \mathcal{L}_{\mathcal{N}}(C)$, we have that the PN system $\langle \mathcal{N}, M_0 \rangle$ and its unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$ are behaviorally equivalent.

In the following, whenever clear from the context, we drop the lower index C when we refer to the partial order relation $\preceq_C$ defined in a configuration C.

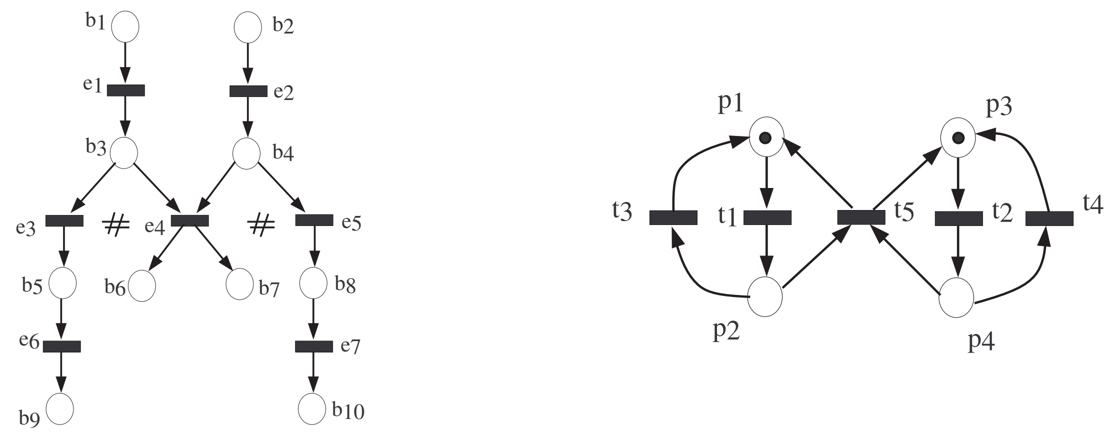

Fig. 1.

Example 1: Consider the occurrence net $\mathfrak{O} = (B, E, G)$ displayed in Fig. 1-left. We have that:

- $b_1 \preceq e_1 \preceq b_3 \preceq e_3$, *etc.*
- $e_3 \sharp e_4$; $e_4 \sharp e_5$; $e_6 \sharp e_4$ *since* $e_3 \preceq e_6$ *and* $e_3 \sharp e_4$, *etc.*
- $e_1 \parallel e_2$, $e_3 \parallel e_2$, *etc.*
- $\mathtt{min}(\mathfrak{O}) = \{b_1, b_2\}$

Consider the PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ displayed in Fig. 1-right. Let ϕ be a homomorphism from the occurrence net $\mathfrak{O}$ onto $\langle \mathcal{N}, M_0 \rangle$ be defined as:

- $\phi(b_1) = p_1$; $\phi(b_2) = p_3$; $\phi(b_3) = p_2$; $\phi(b_4) = p_4$
- $\phi(b_5) = \phi(b_6) = p_1$; $\phi(b_7) = \phi(b_8) = p_3$; $\phi(b_9) = p_2$; $\phi(b_{10}) = p_4$
- $\phi(e_\iota) = t_\iota$ *for* $\iota = 1, 2, 3$, $\phi(e_4) = t_5$, $\phi(e_5) = t_4$
- $\phi(e_6) = t_1$; $\phi(e_7) = t_2$

We have that $(\mathfrak{O}, \phi)$ is a branching process for $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$. C_1 is the initial configuration with $B_{C_1} = \{b_1, b_2\}$ and $E_{C_1} = \emptyset$. C_2 is the configuration obtained appending to C_1 the event e_1 i.e. $B_{C_2} = \{b_1, b_2, b_3\}$ and $E_{C_2} = \{e_1\}$.

C_3 is the configuration obtained appending to the initial configuration the events e_1 and e_2 i.e. $B_{C_3} = \{b_1, b_2, b_3, b_4\}$ and $E_{C_3} = \{e_1, e_2\}$. $CUT(C_1) = \{b_1, b_2\}$; $CUT(C_2) = \{b_2, b_3\}$; $CUT(C_3) = \{b_3, b_4\}$.

C_4 is obtained appending to C_3 the event e_4, i.e. $B_{C_4} = \{b_1, b_2, b_3, b_4, b_6, b_7\}$ and $E_{C_4} = \{e_1, e_2, e_4\}$. The partially ordered set $(E_{C_4}, \preceq)$ has two linearizations, $\sigma_1 = e_1 e_2 e_4$ and $\sigma_1 = e_2 e_1 e_4$ that differ by the way the concurrent events e_1 and e_2 are interleaved. Thus two traces $\tau_1 = t_1 t_2 t_5$ and $\tau_1 = t_2 t_1 t_5$ are compactly represented by C_4 without interleaving the two concurrent transitions t_1 and t_2.

III. The on-line monitoring of Petri Net models under partial observation

In this section we shell present firstly the standard algorithm for constructing a classical observer-automaton of a given PN model. The method is similar to the construction of a classical observer for

DES modeled as automata [MBL00]. Then we show how to construct a reduced observer automaton (basis reachability tree in [GCS05]) that is based on the computation of the set of minimal explanations of the observation generated by the plant. The idea is simple. The set of all the markings that can be reached from the initial markings by firing strings of transitions that obey the received observation (explanations) can be characterized as follows. First a set of (basis) markings is computed calculating backwards the set of minimal explanations of the received observation. All the markings that are not included in this set of basis markings are then reachable from a basis marking firing a string of unobservable transitions.

A. The classical observer automaton

Consider a PN model $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ and partition the set of transitions into disjoint subsets of observable $\mathcal{T}_o$ and unobservable transitions $\mathcal{T}_{uo}$, i.e. $\mathcal{T} = \mathcal{T}_o \cup \mathcal{T}_{uo}$ and $\mathcal{T}_o \cap \mathcal{T}_{uo} = \emptyset$. Given an arbitrary marking M denote by $UR_{\mathcal{N}}(M)$ the unobservable reach of M that is the set of markings that can be reached starting from M by firing only strings of unobservable transitions:

$$UR_{\mathcal{N}}(M) = \left\{ M' \mid \exists \sigma_{uo} \in \mathcal{T}_{uo}^* \text{ s.t. } M \xrightarrow{\sigma_{uo}} M' \right\} \tag{2}$$

For a set of markings $\mathcal{M}$, define: $UR_{\mathcal{N}}(\mathcal{M}) = \bigcup_{M \in \mathcal{M}} UR_{\mathcal{N}}(M)$.

Consider a PN $\langle \mathcal{N}, M_0 \rangle$ with labeling function $\ell_o : \mathcal{T}_o \to \Omega$ where Ω is a set of labels that are emitted by the observable events. The definition of ℓ_o extends to strings in the obvious manner i.e. for $\sigma \in \mathcal{T}_o^*$, $\sigma = t_1 t_2 \ldots t_\lambda$ we have $\ell_o(\sigma) = \ell_o(t_1)\ell_o(t_2)\ldots\ell_o(t_\lambda)$.

Definition 11: The classical observer-automaton $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$ *for the partially observable PN* $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{T} = \mathcal{T}_o \cup \mathcal{T}_{uo}$ *is* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle) = (X_{co}, E_{co}, f_{co}, x_0^{co}, \varrho_{co})$ *where:*

- X_{co} *is the set of states of* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$
- $\varrho_{co} : X_{co} \to Pwr(\mathcal{R}_{\mathcal{N}}(M_0))$ *is a function that associates to each state* $x_{co} \in X_{co}$ *a set of reachable markings* $\varrho_{co}(x_{co}) \in Pwr(\mathcal{R}_{\mathcal{N}}(M_0))$
- $E_{co} = \Omega$ *is the set of events of the classical observer* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$
- $\varrho_{co}(x_0^{co}) = UR_{\mathcal{N}}(M_0)$ *is the set of markings in* $\langle \mathcal{N}, M_0 \rangle$ *estimated in the initial state of the classical observer* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$
- $f_{co} : X_{co} \times E_{co}^* \to X_{co}$ *is the transition function of* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$ *that is defined as follows: for* $x_\iota^{co} \in X_{co}$ *a state of* $\mathtt{CO}(\langle \mathcal{N}, M_0 \rangle)$ *and a string of observable labels* $\omega \in E_{co}^*$ *we have* $f_{co}(x_0^{co}, \omega) = x_\iota^{co}$ *if* $\varrho_{co}(x_\iota^{co}) \neq \emptyset$ *where* $\varrho_{co}(x_\iota^{co}) = \left\{ M_\iota : M_0 \xrightarrow{\tau} M_\iota \wedge \ell_o(\Pi_{\mathcal{T}_o}(\tau)) = \omega \right\}$.

B. The reduced observer automaton

It is possible to obtain an observer automaton that is easier to work with by modifying the read-out map ϱ_{co} to another, simpler read-out map ϱ_{ro} that, for the same observed trace ω enumerates only a

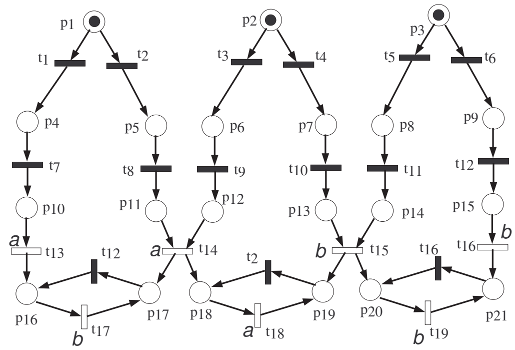

Fig. 2.

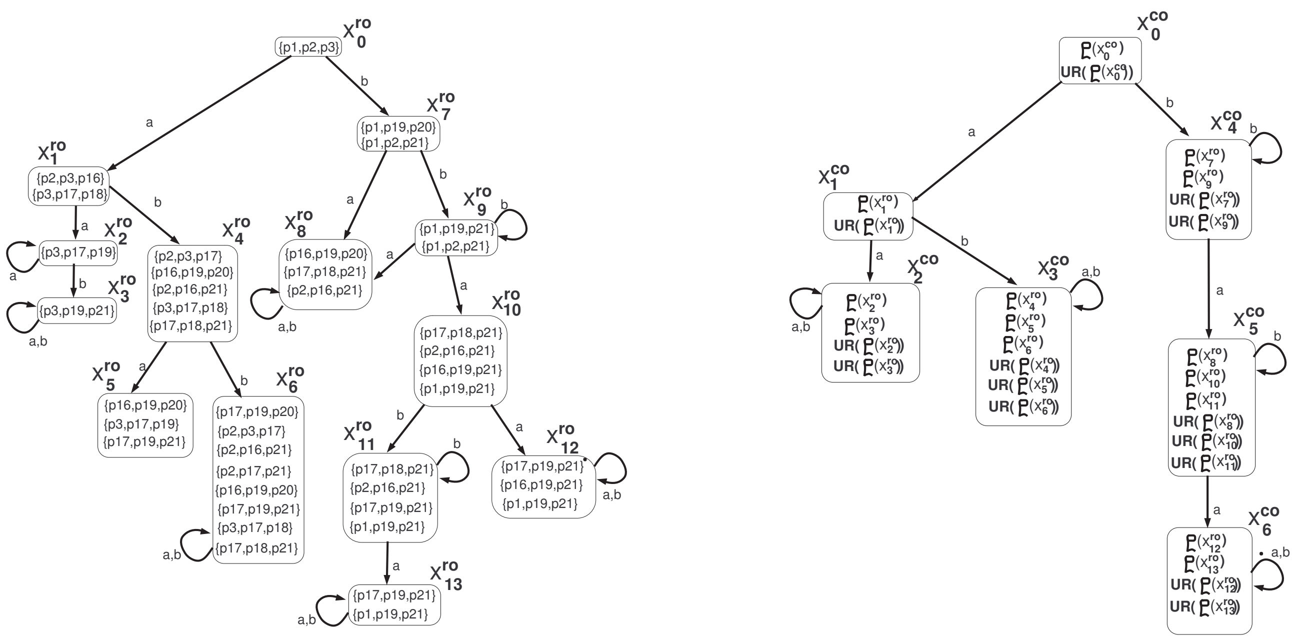

Fig. 3. The reduced observer (left) and the classical observer (right) for PN model of Example 2

subset of the PN $\langle \mathcal{N}, M_0 \rangle$-markings in ϱ_{co}. This modification may require a change in the state space of the observer automaton as well. To illustrate the rationale behind the construction of a reduced observer automaton $\texttt{RO}$ consider a state $x_\iota^{co} \in X_{co}$ of the $\texttt{CO}$ and then let $\mathcal{M}'(x_\iota^{co})$ be a subset of the set of markings $\varrho_{co}(x_\iota^{co})$ of $\texttt{CO}$ corresponding to x_ι^{co} s.t. $UR_{\mathcal{N}}(\mathcal{M}'(x_\iota^{co})) = \varrho_{co}(x_\iota^{co})$. We follow [GS02] and call $\mathcal{M}'(x_\iota^{co})$ a set of basis markings for $\varrho_{co}(x_\iota^{co})$ if $UR_{\mathcal{N}}(\mathcal{M}'(x_\iota^{co})) = \varrho_{co}(x_\iota^{co})$.

Definition 12: $\texttt{RO}(\langle \mathcal{N}, M_0 \rangle) = (X_{ro}, E_{ro}, f_{ro}, x_{ro_0}, \varrho_{ro})$ *is a reduced observer-automaton of the PN* $\langle \mathcal{N}, M_0 \rangle$ *if* $\forall \omega \in E_{ro}^*$, $f_{ro}(x_0^{ro}, \omega) = x_\iota^{ro}$ *implies that:*

1) $\forall M_\iota \in \varrho_{ro}(x_\iota^{ro}), \exists \tau \in \mathcal{L}_{\mathcal{N}}(M_0)$ *s.t.* $M_0 \xrightarrow{\sigma} M_\iota$ *and* $\ell_o(\Pi_{\mathcal{T}_o}(\tau)) = \omega$
2) *and* $UR(\varrho_{ro}(x_\iota^{ro})) = \varrho_{co}(x_\lambda^{ro})$ *where* $f_{co}(x_0^{ro}, \omega) = x_\lambda^{co}$

Example 2: Consider the PN model displayed in Fig. 2. The set of observable transitions is $\mathcal{T}_o = \{t_{13}, t_{14}, t_{15}, t_{16}, t_{17}, t_{18}, t_{19}\}$ *while all the other transitions are unobservable. The observation labeling* *function is defined as follows:* $\ell_o(t_{13}) = \ell_o(t_{14}) = \ell_o(t_{18}) = a$, $\ell_o(t_{15}) = \ell_o(t_{16}) = \ell_o(t_{17}) = \ell_o(t_{19}) = b$.

For the initial state the $\mathtt{RO}$ *(Fig. 3-left) considers only the initial marking* $M_0 = \{p_1, p_2, p_3\}$*, i.e.* $\varrho(x_0^{ro}) = \{M_0\}$ *(since the PN model is 1-safe we denote the initial marking by simply enumerating the places that contain a token).*

The $\mathtt{CO}$ *(Fig. 3-right) considers for its initial state all the markings that can be reached from* M_0 *firing strings of unobservable transitions, i.e.* $\varrho(x_0^{co}) = UR(M_0)$*. In total* $\varrho(x_0^{co})$ *comprises* 125 *markings that correspond to all the triples* (p_i, p_j, p_q) *of marked places that belongs to* $(p_1, p_4, p_5, p_{10}, p_{11}) \times (p_2, p_6, p_7, p_{12}, p_{13}) \times (p_3, p_8, p_9, p_{14}, p_{15})$.

Suppose the first observed label is a*. The state* x_1^{ro} *of* $\mathtt{RO}$ *considers the markings* $M_1 = \{p_2, p_3, p_{16}\}$ *and* $M_2 = \{p_3, p_{17}, p_{18}\}$*, i.e.* $\varrho(x_1^{ro}) = \{M_1, M_2\}$.

The state x_1^{co} *of the* $\mathtt{CO}$ *on the other hand considers all the markings that can be reached from* M_0 *firing strings that contain only one observable transition that is labeled* a. $\varrho(x_1^{co}) = UR(M_1) \cup UR(M_2)$*. In total* $\varrho(x_1^{co})$ *contains* 35 *markings that correspond to all the triples of marked places* (p_i, p_j, p_q) *that belong to either* $p_{16} \times (p_2, p_6, p_7, p_{12}, p_{13}) \times (p_3, p_8, p_9, p_{14}, p_{15})$ *or* $(p_{16}, p_{17}) \times p_{18} \times (p_3, p_8, p_9, p_{14}, p_{15})$.

Notice that the set of markings considered by the states of the $\mathtt{RO}$ x_2^{ro} *and* x_3^{ro} *have the same unobservable reach, i.e.* $UR(M_3) = UR(M_4)$ *where* $M_3 = \{p_3, p_{17}, p_{19}\}$ *and* $M_4 = \{p_3, p_{19}, p_{21}\}$*. The* $\mathtt{CO}$ *considers only one state* x_2^{co} *where* $\varrho(x_2^{co}) = UR(\varrho(x_2^{ro})) = UR(\varrho(x_3^{ro}))$*. Similarly for* x_3^{co} *we have that* $\varrho(x_3^{co}) = UR(\varrho(x_4^{ro})) = UR(\varrho(x_5^{ro})) = UR(\varrho(x_6^{ro}))$*, for* x_4^{co} *we have that* $\varrho(x_4^{co}) = UR(\varrho(x_7^{ro})) = UR(\varrho(x_9^{ro}))$*, for* x_5^{co} *we have that* $\varrho(x_5^{co}) = UR(\varrho(x_8^{ro})) = UR(\varrho(x_{10}^{ro})) = UR(\varrho(x_{11}^{ro}))$ *and for* x_6^{co} *we have that* $\varrho(x_6^{co}) = UR(\varrho(x_{12}^{ro})) = UR(\varrho(x_{13}^{ro}))$.

Denote by $\mathcal{L}(\mathtt{RO}(\langle\mathcal{N}, M_0\rangle))$ and $\mathcal{L}(\mathtt{CO}(\langle\mathcal{N}, M_0\rangle))$ the language of the reduced observer $\mathtt{RO}(\langle\mathcal{N}, M_0\rangle)$ respectively the language of the classical observer $\mathtt{CO}(\langle\mathcal{N}, M_0\rangle)$. Both languages are subsets of $\mathcal{T}_o^*$ and must be identical since they both must accept the set of all possible observed sequences of labels generated by the PN model $\langle\mathcal{N}, M_0\rangle$. By construction of the $\mathtt{RO}$ we have that the reduced observer $\mathtt{RO}(\langle\mathcal{N}, M_0\rangle)$ and the classical observer $\mathtt{CO}(\langle\mathcal{N}, M_0\rangle)$ of a partial observable PN model $\langle\mathcal{N}, M_0\rangle$ are such that:

1) $\mathcal{L}(\mathtt{RO}(\langle\mathcal{N}, M_0\rangle)) = \mathcal{L}(\mathtt{CO}(\langle\mathcal{N}, M_0\rangle))$
2) $\mid X_{co} \mid \leq \mid X_{ro} \mid$
3) and $\forall \omega \in \mathcal{L}(\mathtt{RO}(\langle\mathcal{N}, M_0\rangle))(= \mathcal{L}(\mathtt{CO}(\langle\mathcal{N}, M_0\rangle))) \Rightarrow \varrho(x_\iota^{ro}) \subseteq \varrho(x_\lambda^{co})$) and $UR(\varrho(x_\iota^{ro})) = \varrho(x_\lambda^{co})$

where x_ι^{ro} and x_λ^{co} are the states that are reached from the initial state x_0^{ro} of $\mathtt{RO}$ respectively the initial state x_0^{co} of $\mathtt{CO}$ by executing the string of labels ω, i.e. $x_\iota^{ro} = f_{co}(x_0^{ro}, \omega)$ and $x_\lambda^{co} = f_{co}(x_0^{co}, \omega)$.

Thus the number of states of a classical observer is smaller than the number of states of the reduced observer but the number of markings that are considered by the classical observer for any of its states is usually a lot bigger than the number of markings considered by the reduced observer for its corresponding state(s). This means that the state space of the reduced observer may be larger than the state space of the classical observer. However this drawback vanishes when the observer is derived on-line, computing the current state of the observer based on the last observation generated by the plant.

C. On-line monitoring of PNs under partial observation

Assume from now on that the size of the plant under investigation is large. This typically means that the off-line derived CO considers for any of its states a very large number of reachable markings and moreover that the size of CO (seen as automaton) may be also very big. Furthermore assume that changes of the plant structure (e.g. changes in $\mathcal{T}_o$ when a sensor fails) occur from time to time. This implies that the CO, derived off-line, must be modified from time to time. Given the effort required for the off-line synthesis of the CO and given the size of the CO this is hardly possible in practice.

This difficulty can be partly avoided by constructing the observer on-line, constructing only those branches of the observer graph that are necessary according to the observation, and doing that only when those branches become necessary. This leads to the following recursive on-line implementation of an observer (both for CO and for RO):

1) the on-line observer starts in the initial state x_0
2) as soon an observable transition $t^o \in \mathcal{T}_o$ is executed in the plant (we assume that no two observable events are executed exactly at the same time) the sensor associated with t^o immediately informs the supervisory system (we assume that the sensor output $\ell_o(t'^o)$ that is emitted, when any observable transition $t'_o \in \mathcal{T}$ is executed, is never lost and never delayed)
3) a new state of the observer is calculated by enumerating a set of possible markings the plant can be in after observing the label $\ell_o(t^o)$
4) return to 2 with the newly calculated state (the set of new markings) as the initial state

Basically an on-line observer is obtained by deriving only the branch of the off-line observer-automaton that explains the on-line plant observation.

From now on we make the following assumption unless otherwise stated:

Assumption 1: The labeling function ℓ_o is injective, i.e. $\ell(t_\iota) = \ell(t_\nu) \Rightarrow t_\iota = t_\nu$.

It is easy to extend the algorithms derived in this paper to the case where ℓ_o is not injective, by carrying out the backwards search for all the transitions that share the same label, and tacking unions of sets of reachable markings.

Below $\mathcal{O}_n = t_1^o \ldots t_n^o$ denotes a string of n observable events ($\mathcal{O}_n \in \mathcal{T}_o^*$) that are known to have happened in the plant.

The classical on-line observer $\texttt{CO}(\langle \mathcal{N}, M_0 \rangle)$ considers for its initial state x_0^{co} in step $1)$ the set of markings given by $\varrho_{co}(x_0^{co})$ where: $\varrho_{co}(x_0^{co}) = \left\{M : M_0 \xrightarrow{\sigma_{uo}} M \wedge \sigma_{uo} \in \mathcal{T}_{uo}^*\right\}$. Then inductive calculations in step 2), 3), 4) evaluate, for an observed string $\mathcal{O}_k = t_1^o \dots t_k$, the state x_k^{co}:

$$\varrho_{co}(x_k^{co}) = \left\{M : M_0 \xrightarrow{\tau} M \wedge \Pi_{\mathcal{T}_o}(\tau) = \mathcal{O}_k\right\}$$

The on-line computation of a (minimal) reduced observer $\mathtt{RO}(\langle\mathcal{N}, M_0\rangle)$ can be performed backwards by calculating the minimal explanations of the received string of observations $\mathcal{O}_n = t_1^o \dots t_n$. Recall that a minimal explanation is a trace that considers only transitions that must have happened *prior to* the execution of the last received observation.

Consider the PN model $\langle\mathcal{N}, M_0\rangle$ and its net unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$ as defined in Section II-C.

Definition 13: Given the unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$ of a PN $\langle\mathcal{N}, M_0\rangle$ and the first observed event $\mathcal{O}_1 = t_1^o$ then $\underline{C}(t_1^o) = (B_{\underline{C}(t_1^o)}, E_{\underline{C}(t_1^o)}, G_{\underline{C}(t_1^o)})$ is a minimal configuration that allows for the execution of t_1^o if:

i) $e \in E_{\underline{C}(t_1^o)}$ *s.t.* $\phi(t_1^o) = e_1^o$ *and*

ii) $\forall e \in E_{\underline{C}(t_1^o)}$, *if* $e \neq e_1^o$ *then* $\phi(e) \in \mathcal{T}_{uo}$ *and* $e \preceq e_1^o$.

Denote by $\underline{\mathcal{C}}(\mathcal{O}_1)$ the set of all minimal configurations that satisfy Definition 13 for observation $\mathcal{O}_1 = t_1^o$ and denote by $\underline{\mathcal{E}}(\mathcal{O}_1)$ the set of all minimal explanations of $\mathcal{O}_1$:

$$\underline{\mathcal{E}}(\mathcal{O}_1) = \left\{\sigma \mid \sigma \in \langle E_{\underline{C}(t_1^o)}\rangle \wedge \underline{C}(\mathcal{O}_1) \in \underline{\mathcal{C}}(\mathcal{O}_1)\right\}$$

Denote by $\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_1)$ the set of traces in $\langle\mathcal{N}, M_0\rangle$ that correspond to the minimal explanations $\underline{\mathcal{E}}(\mathcal{O}_1)$:

$$\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_1) = \{\tau \mid \tau = \phi(\sigma) \wedge \sigma \in \underline{\mathcal{E}}(\mathcal{O}_1)\}$$

Definition 13 can be extended recursively for a given a sequence of observed events $\mathcal{O}_n = t_1^o \dots t_n^o$, as follows.

Definition 14: Given the unfolding $\mathcal{U}_{\mathcal{N}}(M_0)$ and a sequence of observed events $\mathcal{O}_n = t_1^o \dots t_n^o$ then $\underline{C}(\mathcal{O}_n) = (B_{\underline{C}(\mathcal{O}_n)}, E_{\underline{C}(\mathcal{O}_n)}, G_{\underline{C}(\mathcal{O}_n)})$ is a minimal configuration that obeys the observation $\mathcal{O}_n$ if:

1) *there are n events in $E_{\underline{C}(\mathcal{O}_n)}$ that have images via ϕ observable transitions and $\forall k$, $1 \leq k \leq n$, there exists an unique $e_k^o \in E_{\underline{C}(\mathcal{O}_n)}$ s.t. $\phi(e_k^o) = t_k^o$*
2) *($\forall q, k : \quad 1 \leq q < k \leq n$) $\Rightarrow$ ($e_q^o \prec e_k^o$ or $e_q^o \| e_k^o$)*
3) $\forall e \in E$, *$\phi(e) \in \mathcal{T}_{uo} \Rightarrow \exists e_k^o$ such that $e \preceq e_k^o$.*

Denote by $\underline{\mathcal{C}}(\mathcal{O}_n)$ the set of all minimal configurations that minimally explain $\mathcal{O}_n$ and let $\underline{\mathcal{E}}(\mathcal{O}_n)$, respectively $\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_n)$ be defined as above.

D. Backward computation of an on-line reduced observer

The backward computation of the minimal explanations can be seen as a forward search in the reverse net $\mathcal{N}_{rev}$ (obtained from $\mathcal{N}$ by reversing the direction of all the arcs) using modified firing and enabling rules. The backward search algorithm that we use for deriving the reduced observer is an adaptation of the algorithm presented in [AIN00], [FRSB02] for checking the coverability property. The problem in [AIN00], [FRSB02] is to check (backwards) if, given a bad marking M_{bad}, there is a trace allowable from the initial marking M_0 that leads to a marking greater than M_{bad} or equal. The difference is that for observer design and for fault detection one must calculate all the minimal traces, whereas in checking the coverability property, it is sufficient to prove the existence of one single trace.

Formally we have the following way of defining the reverse net dynamics. Define $a \ominus b = a - b$ if $a \geq b$, and $a \ominus b = 0$ otherwise and extend "$\ominus$" to multi-sets in the natural manner [AIN00].

Definition 15: Backwards enabling rule: *A transition t is backward enabled in a marking $M \in \mathbb{N}^{|\mathcal{P}|}$ if $\exists p \in t^{\bullet}$ s.t. $M(p) \geq 1$.* Backwards firing rule: *A backward enabled transition t in a marking $M \in \mathbb{N}^{|\mathcal{P}|}$ fires backwards from M producing M' (denoted $M \overset{t}{\rightsquigarrow} M'$) where $M' = M \ominus Post(t, \cdot) + Pre(\cdot, t)$.*

A sequence of transitions $\tau = t_\upsilon \ldots t_1$ is backward allowable from M_υ (denoted $M_\upsilon \overset{\tau}{\rightsquigarrow} M_0$) if for $\iota = \upsilon, \ldots, 0$, $\tau_\iota = t_\upsilon \ldots t_{\iota+1}$, and t_ι is backward enabled in M_ι where $M_\upsilon \overset{\tau_\iota}{\rightsquigarrow} M_\iota$ i.e. $\exists M_{\upsilon-1}, \ldots M_{\iota+1}$ such that: $M_\upsilon \overset{t_\upsilon}{\rightsquigarrow} M_{\upsilon-1} \overset{t_{\upsilon-1}}{\rightsquigarrow} M_{\upsilon-2} \ldots \overset{t_{\iota+1}}{\rightsquigarrow} M_\iota$.

Definition 16: Given a PN $\langle \mathcal{N}, M_0 \rangle$ and a marking M, then M is covered by M_0 if $\exists M' \leq M_0$ s.t $M \overset{\sigma}{\rightsquigarrow} M'$.

Definition 17: Consider a PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ and a partition $\mathcal{T} = \mathcal{T}_o \cup \mathcal{T}_{uo}$. Then given an initial marking M_0 and a final marking M_{fin} denote by $\mathcal{UC}_{\mathcal{N}}(M_{fin}, M_0)$ the set of all markings $M \leq M_0$ that cover M_{fin} by finite unobservable strings:

$$\mathcal{UC}_{\mathcal{N}}(M_{fin}, M_0) = \left\{ M \leq M_0 \mid M_{fin} \overset{\sigma_{uo}}{\rightsquigarrow} M \wedge \sigma_{uo} \in \mathcal{T}_{uo}^* \right\}$$

Let $\mathcal{UL}_{\mathcal{N}}(M_{fin}, M_0)$ be the set of unobservable strings that are backwards feasible from M_{fin} and lead to a marking $M \leq M_0$:

$$\mathcal{UL}_{\mathcal{N}}(M_{fin}, M_0) = \left\{ \sigma_{uo} \in \mathcal{T}_{uo}^* \mid \exists M \in \mathcal{UC}_{\mathcal{N}}(M_{fin}, M_0) \;\; s.t. \;\; M_{fin} \overset{\sigma_{uo}}{\rightsquigarrow} M \right\}$$

Proposition 1: We have that:

(a) *Given a PN $\langle \mathcal{N}, M_0 \rangle$ and a marking M that is not covered by M_0 then $\forall M' > M$, M' is not covered by M_0.*

(b) *Given a PN $\mathcal{N}$, a partition $\mathcal{T} = \mathcal{T}_o \cup \mathcal{T}_{uo}$, a final marking M_{fin}, and an initial marking M_0 then:*

$$\mathcal{UC}_{\mathcal{N}}(M_{fin}, M_0) \neq \emptyset \quad \textit{if} \quad \forall M'_{fin} < M_{fin} \quad \mathcal{UC}_{\mathcal{N}}(M'_{fin}, M_0) \neq \emptyset \tag{3}$$

Proof: The proof is straightforward. ■

Let $t^o \in \mathcal{T}$ be the first observable event. We explain below how the set of minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ is calculated backwards.

Alg_min_exp: Algorithm to calculate the set of minimal explanations

INPUT: $\langle \mathcal{N}, M_0 \rangle$, $\mathcal{T}_o$, $\mathcal{T}_{uo}$, t^o

OUTPUT: $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$

1 label M_{fin} as the root and tag it *"new"* where $M_{fin} = Pre(\cdot, t^o)$

2 while new markings exist do the following:

2.1 select a new marking M s.t. :

2.1.1 there does not exist another marking M' s.t. $M' \leq M$ and M' is tagged either as *"new"* or *"unknown"*

2.1.2 M has no a predecessor marking M' such that there exists a marking M'' tagged as *"unknown"* and $M'' \leq M$

2.2 if no unobservable transitions are backwards enabled at M then

2.2.1 if $M \leq M_0$ then tag M as *"solution end"* and tag all the markings from the root to M as *"solution"*

2.2.2 else tag M as *"no solution"*

2.2.3 repeat until no more markings are tagged with *"no solution"*

2.2.3.1 for all visited markings s.t. $M' \geq M$, tag M' as *"no solution"*

2.2.3.2 remove from the tree the markings that are reached backwards only from markings tagged as *"no solution"*

2.2.3.3 for all markings M'' that have the tag *"unknown"* and have their successors tagged as *"no solution"* tag them as *"no solution"*

2.3 else tag M as *"solution-end"* if $M \leq M_0$, otherwise tag M as *"unknown"* and for each unobservable transition t enabled at M:

2.3.1 calculate $M \overset{t}{\leadsto} M'$

2.3.2 if there exits a marking M'' such that $M' \geq M''$ then

2.3.2.1 if M'' is tagged as *"no solution"* then remove M'

2.3.2.2 else if $M' = M''$ then draw an arc from M to M''

2.3.3 else introduce M' as a node, draw an arc with label t from M to M', and tag M'' as *"new"*

2.4 if no new marking can be selected at 2.1 then a fix point is achieved and the calculation ends

3 $\mathcal{UL}_{\mathcal{N}}(M_{fin}, M_0)$ is obtained as the set of unobservable strings that start in a marking tagged as *solution-end* and end in the *root* marking.

4 $\underline{\mathcal{L}}_{\mathcal{N}}(t^o) = \{\tau \mid \tau = \sigma t^o$ and $\sigma \in \mathcal{UL}_{\mathcal{N}}(M_{fin}, M_0)\}$

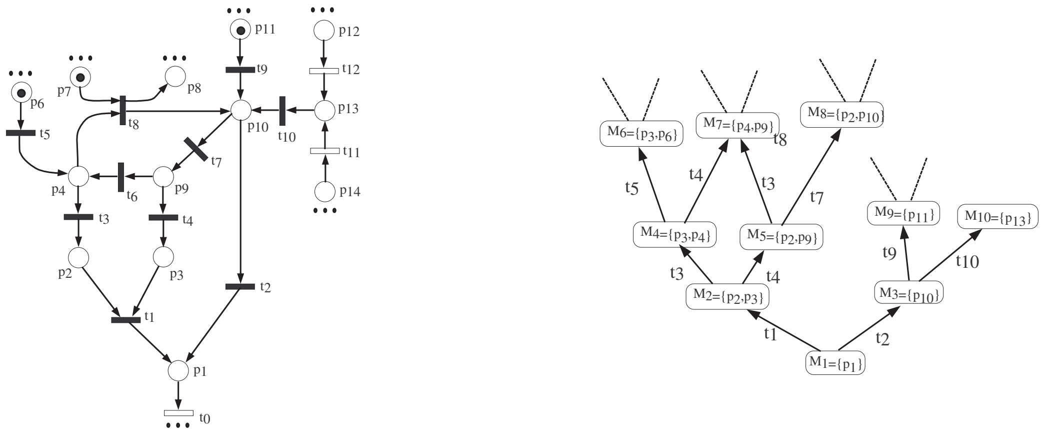

Fig. 4. A part of a PN model and the backward coverability tree for $M_1 = Pre(\cdot, t^o)$

Example 3: Consider in Fig. 4-left a part of a PN model (the dots placed next to $p_6, p_7, p_8 p_{11}, p_{12}, p_{14}$ *and* t_0 *indicate this). Transitions* t_0, t_{11} *and* t_{12} *are observable transitions whereas all the other transitions that are displayed are unobservable.*

Let t_0 *be the transition that is observed first.* $M_1 = \{p_1\}$ *is the root marking, i.e. the partial marking that allows for the execution of* t_0.

At M_1 *we have* t_1 *and* t_2 *as the backwards enabled transitions and they lead to* M_2 *respectively* M_3. *Assume that* M_2 *is selected at step* 2.1. t_3 *and* t_4 *are backfired from* M_2 *in step* 2.3 *of* **Alg_min_exp** *and lead to* M_4 *and* M_5 *respectively. The set of markings tagged as* ”new” *is* $\{M_3, M_4, M_5\}$. *Consider that* M_5 *is selected.* t_3 *and* t_7 *are backfired and we obtain* M_7 *and* M_8. *At this point the set of markings tagged as* ”new” *is now equal to* $\{M_3, M_4, M_7, M_8\}$. *Notice that at this step* M_8 *cannot be selected since* $M_8 \geq M_3$ *and* M_3 *has the tag* ”new”.

Next select M_4 *in step* 2.1. t_5 *is backfired from* M_4 *obtaining* M_6 *while the marking obtained backfiring* t_4 *is* M_7. *After the next execution of step* 2 *the set of markings tagged as* ”new” *is* $\{M_3, M_6, M_8\}$. *Select in the iteration of* 2.1 *the marking* M_3. t_9 *is backfired in step* 2.3 *and we obtain the marking* $M_9 = \{p_{11}\}$ *that is smaller than the initial marking. Thus* $\tau = t_9 t_2 t_0$ *is a minimal explanation.* M_9 *is tagged as* ”solution-end” *and* M_3 *is marked as* ”solution”. t_{10} *is backfired and we obtain the marking* $M_{10} = \{p_{13}\}$.

The set of markings tagged as ”new” *is* $\{M_6, M_8, M_9, M_{10}\}$. *Now* M_8 *can be selected at step* 2.1 *since* M_3 *is tagged as* ”solution”.

At M_{10} *there are no unobservable transitions backwards enabled and in the initial marking* p_{13} *is unmarked. Thus* M_{10} *is tagged* ”no solution”. *The computation continues in this manner until either the set of new markings becomes empty or no selection can be made at step* 2.1 *(a fix point is achieved and the computation ends).*

Remark 1: In the algorithm **Alg_min_exp**, the interleaving of the concurrent events is not filtered out. I.e. at M_2, t_3 and t_4 are concurrent events that are interleaved obtaining backwards M_7. In the same manner as the unfolding method [McM92] is used to search forward from the initial marking, the backwards unfoldings proposed in [AIN00] can be used to filter out the interleaving of concurrent events.

Remark 2: In step 2 of **Alg_min_exp** the computation continues even if a marking smaller than M_0 is found and a minimal explanation is derived. This is necessary because we must calculate the entire set of minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ to guarantee that the unobservable reach $UR(\underline{\mathcal{M}}_{\mathcal{N}}(t^o))$ of the subset of current possible markings that we derive $\underline{\mathcal{M}}_{\mathcal{N}}(t^o)$ for the current state of the reduced observer is indeed equal to the entire set of possible current markings $\mathcal{M}_{\mathcal{N}}(t^o)$, the current state of the classical observer. In order to guarantee this, it is necessary to explore also paths that lead from one initially marked place to another initially marked place, since these paths may provide additional minimal explanations for the observed event.

The preceding remarks indicate that the on-line computational cost of calculating all the minimal explanations for a reduced order observer is still quite large. However we show in Section IV-D that for a subclass of PN models, namely PN models with trap unobservable circuits, there is possible to design a very efficient algorithm that derives a subset of minimal explanations $\underline{\mathcal{L}}'_{\mathcal{N}}(t^o) \subseteq \underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ s.t. $UR(\underline{\mathcal{M}}'_{\mathcal{N}}(t^o)) = UR(\underline{\mathcal{M}}_{\mathcal{N}}(t^o)) = \mathcal{M}_{\mathcal{N}}(t^o)$.

Theorem 1: Given a partially observed PN model $\langle \mathcal{N}, M_0 \rangle$ that is bounded w.r.t. the unobservable evolution, and given any observable event t^o that can be generated first by the plant, then **Alg_min_exp** *derives in finite time the entire set of minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$.*

Proof: First we prove that **Alg_min_exp** terminates. Since the PN model $\langle \mathcal{N}, M_0 \rangle$ is bounded w.r.t. the unobservable evolution we have that the set of minimal explanations is finite (notice that the unobservable cycles that repeat the markings are filtered out). Any infinite sequence created from a finite number of elements must include a copy of an element, infinitely many often. This is in contradiction with *"all the predecessor markings are either bigger or incomparable"*. Hence the algorithm must stop after a finite number of steps. Thus after a finite number of markings have been generated by **Alg_min_exp**, the algorithm either finds a minimal explanations or cannot select a new marking at step 2.1). Since the number of minimal explanations is finite it results that **Alg_min_exp** cannot select a new marking and terminates.

To prove that **Alg_min_exp** computes $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ requires i) to prove that any trace that is calculated is a minimal explanation and ii) to prove that all the minimal explanations are calculated.

i) can be proven straightforwardly by induction constructing a minimal configuration such that the trace that is calculated by **Alg_min_exp** is a linearization of its set of events. The proof of ii) is straightforward since at any step we consider all the unobservable transitions that are backwards enabled. ■

Given the received observation $\mathcal{O}_n = t_1^o \dots t_n^o$ the computation of an on-line reduced observer $\mathtt{RO}(\mathcal{O}_n)$

is performed recursively as follows:

1) initialize the initial state in the reduced observer $\underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_0) = \{M_0\}$ and $\varrho(x_0^{ro}) = \underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_0)$
2) then for $k = 1, \ldots, n$
   a) $M_{fin_k} = Pre(\cdot, t_k^o)$
   b) for all $M_{k-1} \in \underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_{k-1})$
      i) compute $\mathcal{UC}_{\mathcal{N}}(M_{fin_k}, M_{k-1})$ that is the set of markings that cover unobservably M_{fin_k}, with initial marking M_{k-1} executing **Alg_min_exp**
      ii) derive $\mathcal{UL}_{\mathcal{N}}(M_{fin_k}, M_{k-1})$ that is the set of minimal unobservable traces that can be executed from M_{k-1} s.t. the resulting marking covers M_{fin_k}
      iii) derive the set of minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_k)$ and the set of markings $\underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_k)$:

$$\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_k) = \{\tau_k \mid \tau_k = \tau_{k-1}\sigma_{uo}t_k^o, \tau_{k-1} \in \underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_k) \wedge \sigma_{uo} \in \mathcal{UL}_{\mathcal{N}}(M_{fin_k}, M_{k-1})\}$$
$$\underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_k) = \left\{M_k \mid M_0 \xrightarrow{\tau_k} M_k \wedge \tau_k \in \underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_k)\right\}$$

3) create the new state x_k^{ro} in $\texttt{RO}(\mathcal{O}_k)$, $\varrho(x_k^{ro}) = \underline{\mathcal{M}}_{\mathcal{N}}(\mathcal{O}_k)$ and draw an arc from x_{k-1}^{ro} to x_k^{ro} labeled t_k^o

The main drawbacks of the backward search methods are that the computation terminates when a fix-point is achieved and that unreachable states are visited during the computation. Even though incomparable to the forward search (since the backward and the forward search explore different state spaces) the backward search was found more efficient than the forward search for DES models of large size [NAH$^+$98]. As shown in [FRSB02], the computational efficiency of the backward search can be increased by using place invariants (i.e. the visited markings satisfy the P-invariants) or other heuristics to avoid unreachable markings as well as the backward unfolding technique [AIN00] to avoid the consideration of all the possible interleavings of the concurrent events. Moreover for real-life applications, the size of the unobservable sub-net that is processed is in general small, so that the calculation is efficient.

IV. The diagnosis of PN models

In this section we present two algorithms for the centralized diagnosis of a large plant. We present in Section IV-B the classical diagnosis algorithm based on the calculation of the complete explanations of the received observation. We call it classical since the diagnosis is performed based on the calculations derived by a classical observer as presented in Section III-A.

Then in Section IV-C we propose a diagnosis algorithm based on the calculations of the minimal explanations of the received observation (see Section III-B). We show that the diagnosis result based on minimal explanations is sufficient for reliably detecting the faults that happened for sure in the plant.

A. The setting and problem formulation

The plant model represents the normal plant behavior as well as the abnormal (usually undesirable) behavior that can occur after a fault has occurred. The abnormal behavior is initiated by the occurrence of some unobservable (silent) transitions that represent the fault events that may happen in the plant. A diagnoser uses the plant model, the plant observation, and in the distributed setting of [JBB] the information received from its neighbouring agents, in order to answer the following questions: *"Did a fault happen or not?"*(fault detection), *"Which kind of fault happened if any?"* (fault isolation) and *"How did it happen?"*(explanations [McI98]).

The diagnosis task should be seen as part of a centralized supervisory architecture where the diagnosis result is used on-line for taking some control action that guarantee the safe operation of the plant. In this respect and taking into account that the plant under investigation is assumed to have a large size it is important to specify, before designing the algorithms, what are the specifications for the plant diagnostic. For example, the user should specify whether the diagnostic is concerned with finding all the fault-events that *"could have happened in the plant without contradicting the plant observation"* or with finding only the fault events that *"necessarily must have happened for explaining the received observation"*. With the `CO` diagnoser the first specification can be specified, while the `RO` diagnoser can only satisfy the second type of specification.

We consider in this section the synthesis of on-line `CO` and `RO` diagnosers, under the following structural and functional assumptions:

- the PN model of the overall plant $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ is completely known, and it is bounded w.r.t. $\mathcal{T}_{uo}$; in particular we assume that the model is completely correct, without any errors, and that there are no unmodelled (hidden) external interactions (the closed world assumption)
- the initial marking M_0 is precisely known
- the plant observation is represented by a subset of observable transitions $\mathcal{T}_o \subseteq \mathcal{T}$
- the occurrence of an observable transition $t \in \mathcal{T}_o$ is always reported correctly and without delays
- the faults are represented by a subset $\mathcal{T}_f$ of unobservable (silent) transitions ($\mathcal{T}_f \subseteq \mathcal{T}_{uo}$)
- *no-fault-masking* i.e. the occurrence of a fault transition must have effects on the resulting marking and consequently on the future plant behavior.

In this paper we do not formalize the last assumption since we do not deal with diagnosability in itself. This paper answers the question of when and in what sense there exists a reduced observer `RO` that detects those faults that, according to the `CO` observer, must have happened for sure. Conditions for the `CO` to detect all the faults that must have happened for sure can be found in papers on `CO` [SSL$^+$95].

In this work the faults in the PN models are represented as (fault) transitions whose occurrence indicates a malfunction in the plant behavior [SSL+95]. Obviously the set of fault transitions (denoted $\mathcal{T}_f$) is a subset of the set of unobservable transitions ($\mathcal{T}_f \subseteq \mathcal{T}_{uo}$) since otherwise the fault detection problem would be trivial.

Beside the fact that a fault must be unobservable, it must also be unpredictable, i.e. for any state the plant can be in before the occurrence of a fault at least one no-fault event must be legal according to the plant model $\langle \mathcal{N}, M_0 \rangle$ used for synthesizing the diagnoser; otherwise the imminent fault would be predictable and, consequently, the model would not correctly represent the occurrence of the fault (an earliest event should have been labeled as a fault). We formalize this as follows.

Assumption 2: Given a PN model $\langle \mathcal{N}, M_0 \rangle$ and $\mathcal{T}_f$ ($\mathcal{T}_f \subseteq \mathcal{T}_{uo}$) the set of fault transitions, then for any reachable state $M \in \mathcal{R}_{\mathcal{N}}(M_0)$, at least one non-fault transition t, $t \in \mathcal{T} \setminus \mathcal{T}_f$ is enabled, that is:

$$\forall M \in \mathcal{R}_{\mathcal{N}}(M_0),\ Enbl(M) \not\subseteq \mathcal{T}_f$$

In words Assumption 2 says that: *"a fault is the choice of the plant of not respecting the good (designed) behavior"* which is a subset of the behaviour that is legal according to the model used for designing the diagnoser. Since the condition that the faults are unpredictable requires to check for every reachable marking if there are enabled non-fault transitions or not, it is computationally impossible to check for a large PN model whether Assumption 2 holds true or not. However it is very natural to assume that for every fault event $t \in \mathcal{T}_f$, there exists a non-fault event $t' \in \mathcal{T} \setminus \mathcal{T}_f$ such that ${}^\bullet t' \subseteq {}^\bullet t$. This is a sufficient condition for the fault to be unpredictable since whenever a fault event is enabled, at least one non-fault event is enabled as well.

B. Centralized diagnosis based on complete explanations

Consider the plant model given as a PN $\mathcal{N} = (\mathcal{P}, \mathcal{T}, F)$ with given initial marking M_0. Then consider the partition of the transition set $\mathcal{T}$ in two disjoint subsets $\mathcal{T}_o$ observable and respectively $\mathcal{T}_{uo}$ unobservable transitions and let $\mathcal{T}_f \subset \mathcal{T}_{uo}$ be the subset of the unobservable transitions that model the faults. The plant observation available at time θ_n is given by the ordered sequence of observable events $\mathcal{O}_n = t_1^o \dots t_n^o$.

Since $\mathcal{O}_n$ is correctly and without any delay received by the diagnoser-agent, the possible plant evolutions up to the time θ_n are given by the set of all the possible traces in the PN model $\mathcal{N}$ that start from the known initial marking M_0 and that obey the observation $\mathcal{O}_n$:

$$\mathcal{L}_{\mathcal{N}}(\mathcal{O}_n) = \{\tau \in \mathcal{L}_{\mathcal{N}}(M_0) \mid \Pi_{\mathcal{T}_o}(\tau) = \mathcal{O}_n\}$$

The set of the possible states (markings) the plant can be in is:

$$\mathcal{M}_{\mathcal{N}}(\mathcal{O}_n) = \left\{ M \mid \exists \tau \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n) \text{ s.t. } M_0 \xrightarrow{\tau} M \right\}$$

Consequently the plant diagnosis after observing $\mathcal{O}_n$ is obtained by projecting the set of possible evolutions onto the set of fault events $\mathcal{T}_f$:

$$\mathcal{D}_{\mathcal{N}}(\mathcal{O}_n) = \left\{\sigma_f \mid \sigma_f = \Pi_{\mathcal{T}_f}(\tau) \wedge \tau \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)\right\} \tag{4}$$

The centralized diagnosis result is:

$$\mathcal{DR}_{\mathcal{N}}(\mathcal{O}_n) = \begin{cases} \mathtt{N} & \text{if } \mathcal{D}_{\mathcal{N}}(\mathcal{O}_n) = \{\epsilon\} \\ \mathtt{F} & \text{if } \epsilon \notin \mathcal{D}_{\mathcal{N}}(\mathcal{O}_n) \\ \mathtt{UF} & \text{if } \{\epsilon\} \subsetneq \mathcal{D}_{\mathcal{N}}(\mathcal{O}_n) \end{cases} \tag{5}$$

where N, F and UF represent the diagnoser state *normal* (no fault has happened), *sure fault* and respectively *uncertain* (a fault may have happened) [SSL$^+$95].

C. Centralized diagnosis based on minimal explanations

Let the set of minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_n)$ and the set of estimated markings of $\underline{\mathcal{M}}(\mathcal{O}_n)$ be as presented in Section III-B. The minimal plant diagnosis after observing $\mathcal{O}_n$ (denoted $\underline{\mathcal{D}}_{\mathcal{N}}(\mathcal{O}_n)$) is obtained by projecting the set of minimal explanations onto the set of fault events $\mathcal{T}_f$:

$$\underline{\mathcal{D}}_{\mathcal{N}}(\mathcal{O}_n) = \left\{\sigma_f \mid \sigma_f = \Pi_{\mathcal{T}_f}(\tau) \wedge \tau \in \underline{\mathcal{L}}_{\mathcal{N}}(O_n)\right\} \tag{6}$$

Then the diagnosis result based on the set of minimal explanations is:

$$\underline{\mathcal{DR}}_{\mathcal{N}}(\mathcal{O}_n) = \begin{cases} \mathtt{N} & \text{if } \underline{\mathcal{D}}_{\mathcal{N}}(\mathcal{O}_n) = \{\epsilon\} \\ \mathtt{F} & \text{if } \epsilon \notin \underline{\mathcal{D}}_{\mathcal{N}}(\mathcal{O}_n) \\ \mathtt{UF} & \text{if } \epsilon \subsetneq \underline{\mathcal{D}}_{\mathcal{N}}(\mathcal{O}_n) \end{cases} \tag{7}$$

Theorem 2: If the plant model $\mathcal{N}$ obeys Assumption 2 then we have the following relationship between the diagnosis result $\mathcal{DR}_{\mathcal{N}}(\mathcal{O})$ derived based on the set of complete explanations and the diagnosis result $\underline{\mathcal{DR}}_{\mathcal{N}}(\mathcal{O})$ derived based on the set of minimal explanations:

$$\begin{tikzcd}
\mathcal{L}_{\mathcal{N}}(\mathcal{O}_n) \arrow[d] \arrow[r, phantom, "\supseteq"] & \underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_n) \arrow[d] \\
\Pi_{\mathcal{T}_f}(\mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)) \arrow[d] \arrow[r, phantom, "\supseteq"] & \Pi_{\mathcal{T}_f}(\underline{\mathcal{L}}_{\mathcal{N}}(\mathcal{O}_n)) \arrow[d] \\
\mathcal{DR}_{\mathcal{N}}(\mathcal{O}_n) \arrow[r, phantom, "\sim_F"] & \underline{\mathcal{DR}}_{\mathcal{N}}(\mathcal{O}_n)
\end{tikzcd}$$

$\mathcal{DR}_{\mathcal{N}}(\mathcal{O}_n)$	$\sim_{\mathtt{F}}$	$\underline{\mathcal{DR}}_{\mathcal{N}}(\mathcal{O}_n)$
$\{\mathtt{N}\}$	$\Rightarrow$	$\{\mathtt{N}\}$
$\{\mathtt{N}, \mathtt{UF}\}$	$\Leftarrow$	$\{\mathtt{N}\}$
$\{\mathtt{UF}\}$	$\Rightarrow$	$\{\mathtt{N}, \mathtt{UF}\}$
$\{\mathtt{UF}\}$	$\Leftarrow$	$\{\mathtt{UF}\}$
$\{\mathtt{F}\}$	$\Leftrightarrow$	$\{\mathtt{F}\}$

Fig. 5.

Proof: Given the observation $\mathcal{O}_1 = t_1^o \dots t_n^o$, consider the set of configurations $\mathcal{C}(t^o)$ in $\mathcal{U}_\mathcal{N}(\mathcal{O}_n)$. We have that a fault is diagnosed that for sure happened based on the received observation $\mathcal{O}_n$ and using the set of explanations generated by CO iff $\forall C \in \mathcal{C}(t^o)$, $\exists e \in E_C$ s.t. $\phi(e) = t_f$ and $e \preceq e_q^o$ for some event $e_q^o \in E_C$ that corresponds to an event that was observed ($\phi(e_q^o) = t_q^o$, $1 \leq q \leq n$).

This is true because by Assumption 2 in any reachable marking at least a non-fault event is enabled thus the necessary condition for a fault event to be diagnosed that for sure happened is that for every configuration $C \in \mathcal{C}(t^o)$ there exists at least an event e that is the image of fault transition t_f ($\phi(e) = t_f$) that is a predecessor ($e \preceq e_q^o$) of an observed event e_q^o.

Hence by deriving only the set of minimal configuration $\underline{\mathcal{C}}(t^o)$ all the faults that can be diagnosed that for sure have happened are indeed detected. Thus $\mathcal{DR}_\mathcal{N} = \{\texttt{F}\} \Leftrightarrow \underline{\mathcal{DR}}_\mathcal{N} = \{\texttt{F}\}$. The other relations between $\mathcal{DR}_\mathcal{N}$ and $\underline{\mathcal{DR}}_\mathcal{N}$ are trivial. ■

$\underline{\mathcal{L}}_\mathcal{N}(\mathcal{O}_n)$ is in general a lot smaller than $\mathcal{L}_\mathcal{N}(\mathcal{O}_n)$, thus the efficiency of RO diagnoser relies on the computational effort for enumerating backwards the set of minimal explanations. This computational complexity depends on the size of the backward reachable state space for unobservable sub-nets, explaining the different faults one is interested in. Even though the computational effort for deriving $\underline{\mathcal{L}}_\mathcal{N}(\mathcal{O}_n)$ is not explicitly comparable to the computational effort for deriving $\mathcal{L}_\mathcal{N}(\mathcal{O}_n)$ (since the forward respectively the backward search explore different and incomparable state spaces), in practice one can expect that the backwards implementation of the minimal explanations will be quite efficient. Indeed in many applications there are no sub-nets $\mathcal{N}'$ of the PN model $\mathcal{N}$ having a large size and comprising only unobservable events. Moreover the efficiency of the diagnosis algorithm based on the (backward) calculation of the minimal explanations - the on-line reduced order observer algorithm - of the plant observation can be further improved if there is *a priori* knowledge of plant dynamics that allows the use of some heuristics to drive the backward search [FRSB02].

D. The case of PNs with unobservable trap circuits

In this section we treat the case when all the unobservable circuits in the PN model are traps (see Definition 3) showing that this class of PNs allows to compute a (often small) subset of minimal explanations such that the diagnoser designed based on this subset of minimal explanations has the same performance as the diagnoser designed based on the entire set of minimal explanations. We show how additional termination conditions can be used in algorithm **Alg_min_exp** presented in Section III-D (that calculates the set of minimal explanations) in order to calculate this small subset of minimal explanations.

Theorem 3: Consider a trap circuit PN $\langle \mathcal{N}, M_0 \rangle$. Then, given a trace σ that is legal from the initial marking M_0, $\sigma \in \mathcal{L}_\mathcal{N}(M_0)$ we have that:

i) $\sigma' \in \mathcal{L}_{\mathcal{N}}(M_0)$ *and* $\overrightarrow{\sigma'} < \overrightarrow{\sigma}$ *together imply that*

ii) $\exists\ \sigma''$ *s.t.* $\sigma'\sigma'' \in \mathcal{L}_{\mathcal{N}}(M_0)$ *and* $\overrightarrow{\sigma'} + \overrightarrow{\sigma''} = \overrightarrow{\sigma}$

(where $\sigma'\sigma''$ is the trace obtained by catenation of σ' and σ'').

To prove Theorem 3 we need the following result that can be found as Theorem 17 in [Mur89].

Theorem 4: ([Mur89]) *In a trap-circuit net $\mathcal{N}$, M_d is reachable from M_0 iff:*

i) *there exists $\overrightarrow{\sigma}$ a non-negative integer solution of the marking equation Eq. 1*

ii) *and $\langle \mathcal{N}_{\overrightarrow{\sigma}}, M_{0_{\overrightarrow{\sigma}}} \rangle$ has no token-free siphons*

where $\mathcal{N}_{\overrightarrow{\sigma}}$ denotes the sub-net of $\mathcal{N}$ consisting of transitions t s.t. $\overrightarrow{\sigma}(t) > 0$ together with their input and output places and $M_{0_{\overrightarrow{\sigma}}}$ denotes the sub-vector of M_0 for places in $\mathcal{N}_{\overrightarrow{\sigma}}$.

Proof: [Theorem 4]-sketch. $\Leftarrow$ We have that $\mathcal{N}_{\overrightarrow{\sigma}}$ has not a token-free siphon. Then inductively one can prove that after firing a sequence of transitions σ', the remaining sub-net $\mathcal{N}_{\overrightarrow{\sigma''}}$ of $\mathcal{N}_{\overrightarrow{\sigma}}$ with $\overrightarrow{\sigma} = \overrightarrow{\sigma'} + \overrightarrow{\sigma''}$ has not a token-free siphon.

$\Rightarrow$ The proof is trivial. ■

Remark 3: In the following we present the proof of Theorem 3 that basically constitutes a detailed proof of the induction step of the proof of Theorem 4.

Proof: [Theorem 3] Since $\sigma \in \mathcal{L}_{\mathcal{N}}(M_0)$ denote by M_d the marking obtained firing σ from M_0 ($M_0 \xrightarrow{\sigma} M_d$). We have that $\exists \overrightarrow{\sigma''}$ s.t. $M_0 + F \cdot \overrightarrow{\sigma'} + F \cdot \overrightarrow{\sigma''} = M_d$. Then $\sigma' \in \mathcal{L}_{\mathcal{N}}(M_0)$ and $M_0 \xrightarrow{\sigma'} M'$ imply that: $M' + F \cdot \overrightarrow{\sigma''} = M_d$.

To prove that there exists a legal trace σ'' that can be executed from M' we need to prove that $\langle \mathcal{N}_{\overrightarrow{\sigma''}}, M'_{\overrightarrow{\sigma''}} \rangle$ has no token-free siphons where $\mathcal{N}_{\overrightarrow{\sigma''}}$ is the sub-net of $\mathcal{N}$ consisting of transitions that are executed in σ'' together with their input and output places and $M'_{\overrightarrow{\sigma''}}$ is the sub-vector marking of M' for places in $\mathcal{N}_{\overrightarrow{\sigma''}}$.

For $M_0 \xrightarrow{\sigma} M_d$ we have that $M_d(p) \geq 0$ and:

$$\sum_{t \in {}^\bullet p} \overrightarrow{\sigma}(t) + M_0(p) \geq \sum_{t \in p^\bullet} \overrightarrow{\sigma}(t) \tag{8}$$

that in words means that for any place $p \in \mathcal{P}$ the number of executions of the transitions that remove tokens from p in σ is smaller than or equal to the number of tokens plus the number of executions of transitions in σ that add tokens to p.

Consider now a set of places Q in the sub-net $\langle \mathcal{N}_{\overrightarrow{\sigma''}}, M'_{\overrightarrow{\sigma''}} \rangle$ s.t. Q is a siphon in $\langle \mathcal{N}_{\overrightarrow{\sigma''}}, M'_{\overrightarrow{\sigma''}} \rangle$. i.e. Assume that Q is token-free in the marking that results after firing σ' from M_0, i.e. $M'_{\overrightarrow{\sigma'}}(Q) = 0$. We would have then for any place $p \in Q$ that:

$$\sum_{t \in {}^\bullet p} \overrightarrow{\sigma'}(t) + M_0(p) = \sum_{t \in p^\bullet} \overrightarrow{\sigma'}(t) \tag{9}$$

From (8) and (9) we obtain:

$$\sum_{t \in {}^\bullet p} \overrightarrow{\sigma''}(t) \geq \sum_{t \in p^\bullet} \overrightarrow{\sigma''}(t) \tag{10}$$

Now consider a place $p_1 \in Q$. We have that $p \in \mathcal{P}_{\vec{\sigma''}}$ thus in $\mathcal{N}_{\vec{\sigma''}}$ either ${}^\bullet p_1 \neq \emptyset$ or $p_1^\bullet \neq \emptyset$.

From (10) we have that $\sum_{t \in {}^\bullet p_1} \vec{\sigma''}(t) > 0$. Since Q is a siphon we have that:

$$\forall t \in \mathcal{T},\ (\vec{\sigma''}(t) > 0 \text{ and } p_i \in {}^\bullet t) \Rightarrow p_i \in Q$$

The for each p_i we have that $\sum_{t \in {}^\bullet p_i} \vec{\sigma''}(t) > 0$. Two cases must be considered:

Case 1 p_1 and p_i belong to a circuit.

Case 2 there exists a place p_j such that $\vec{\sigma''}(t) > 0$, and $t \in {}^\bullet p_i \cap p_j^\bullet$.

Case 1 We have the following two cases:

Case 1.1 neither p_1 nor p_i have input transitions in $\mathcal{N}_{\vec{\sigma''}}$ other than transitions that are part of the circuit in $\mathcal{N}_{\vec{\sigma''}}$

Case 1.2 either p_1 or p_i has input transitions in $\mathcal{N}_{\vec{\sigma''}}$ other than transitions that are part of the circuit in $\mathcal{N}_{\vec{\sigma''}}$

Case 1.1 We have the following two cases:

Case 1.1.1 neither p_1 nor p_i have input transitions that belong to $\mathcal{N}_{\vec{\sigma'}}$

Case 1.1.2 either p_1 or p_i has input transitions that belong to $\mathcal{N}_{\vec{\sigma'}}$

Case 1.1.1 In this case $\{p_1, p_i\}$ is an empty siphon in $\mathcal{N}_{\vec{\sigma}}$ that contradicts the initial assumption.

Case 1.1.2 In this case either p_1 or p_i would have become marked firing the transitions that belong to σ'. Since all the circuits in $\mathcal{N}$ are traps it results that Q contains tokens.

For *Case 1.2* and *Case 2* consider a place p_j and apply the same reasoning as above. By a simple induction argument one can prove considering all the places of Q that either a place that belongs to circuit has been marked firing a transition considered in the string σ' and thus Q is not empty in $\mathcal{N}_{\vec{\sigma''}}$ or there is a siphon $Q' \subseteq Q$ that was empty in $\mathcal{N}_{\vec{\sigma}}$. Thus the statement of the theorem is proven by contradiction since Q was assumed empty and $\mathcal{N}_{\vec{\sigma}}$ cannot contain an empty siphon. ■

Then we have the following corollary:

Corollary 1: Consider a trap circuit PN $\langle \mathcal{N}, M_0 \rangle$. Then, given two traces σ_1 and σ_2 that are legal from the initial marking M_0, $\sigma_1 \in \mathcal{L}_{\mathcal{N}}(M_0)$ and $\sigma_2 \in \mathcal{L}_{\mathcal{N}}(M_0)$ we have that:

$\sigma_1\sigma_2 \in \mathcal{L}_{\mathcal{N}}(M_0)$ *implies that*

$\exists \sigma_1'$ *s.t.* $\sigma_2\sigma_1' \in \mathcal{L}_{\mathcal{N}}(M_0)$ *and* $\vec{\sigma_1'} = \vec{\sigma_1}$

Proof: Straightforward applying Theorem 3. ■

We now show that the following additional assumption greatly reduces the computational effort required to calculate all the minimal explanations for an observed sequence of events $\mathcal{O}_n$.

Assumption 3: All the unobservable circuits in the PN model of the plant are trap circuits.

Based on Assumption 3 and Theorem 3 we obtain the following result:

Proposition 2: Consider a PN $\langle \mathcal{N}, M_0 \rangle$ *satisfying Assumption 3; the first observed event in the plant is* t_1^o*. Then, given two unobservable strings* $\sigma_{uo_1}, \sigma_{uo_2} \in \mathcal{T}_{uo}^*$ *that are both legal from the initial marking* M_0 *(*$\sigma_{uo}, \sigma'_{uo} \in \mathcal{L}_{\mathcal{N}}(M_0)$*), s.t.:*

1) $M_0 \xrightarrow{\sigma_{uo}} M \geq Pre(\cdot, t_1^o)$
2) $M_0 \xrightarrow{\sigma'_{uo}} M' \geq Pre(\cdot, t_1^o)$
3) *and* $\vec{\sigma'_{uo}} < \vec{\sigma_{uo}}$

there always exists an unobservable string $\exists \sigma''_{uo} \in \mathcal{T}_{uo}^*$ *s.t.* $i)$ $\sigma'_{uo}\sigma''_{uo} \in \mathcal{L}_{\mathcal{N}}(M_0)$ *and* $ii)$ $\vec{\sigma'_{uo}} + \vec{\sigma''_{uo}} = \vec{\sigma_{uo}}$*.*

Proof: The proof is straightforward applying Corollary 1 to $\langle \mathcal{N}_{uo}, M_0^{uo} \rangle$ where $\mathcal{N}_{uo}$ denotes the sub-net of $\mathcal{N}$ comprising the unobservable transitions $\mathcal{T}_{uo}$ and M_0^{uo} denotes the sub-vector of M_0 restricted to places in $\mathcal{N}_{uo}$. ■

Consider the set $\underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ of minimal explanations of the first observed event t^o executed in the plant $\mathcal{N}$. We say that $\tau \in \underline{\mathcal{L}}_{\mathcal{N}}(t^o)$ is a strictly minimal explanation of t^o if $\forall \tau' \in \underline{\mathcal{L}}_{\mathcal{N}}(t^o)$, $\vec{\tau'} \leq \vec{\tau} \Rightarrow \vec{\tau'} = \vec{\tau}$. Denote by $\underline{\mathcal{L}}_{\mathcal{N}}^s(t^o)$ the set of strictly minimal explanations of t^o. For a sequence of observed events $\mathcal{O}_n$ denote by $\underline{\mathcal{L}}_{\mathcal{N}}^s(\mathcal{O}_n)$ the set of strictly minimal explanations of the received observation. Denote by $\underline{\mathcal{M}}_{\mathcal{N}}^s$ the set of markings that result firing strictly minimal explanations from the initial marking:

$$\underline{\mathcal{M}}_{\mathcal{N}}^s(\mathcal{O}_n) = \left\{ M \mid M_0 \xrightarrow{\tau} M \wedge \tau \in \underline{\mathcal{L}}_{\mathcal{N}}^s(\mathcal{O}_n) \right\}$$

Denote by $\underline{\mathcal{DR}}_{\mathcal{N}}^s(\mathcal{O}_n)$ the diagnosis result based on the set of strictly minimal explanations $\underline{\mathcal{L}}_{\mathcal{N}}^s(\mathcal{O}_n)$.

Theorem 5: Consider a PN model that has the property that all the unobservable circuits in $\langle \mathcal{N}, M_0 \rangle$ are traps and any observation $\mathcal{O}_n$ that can be generated by the plant. We have that:

1) $\underline{\mathcal{DR}}_{\mathcal{N}}^s(\mathcal{O}_n) = \{\texttt{F}\} \Leftrightarrow \mathcal{DR}_{\mathcal{N}}(\mathcal{O}_n) = \{\texttt{F}\}$
2) and $UR(\underline{\mathcal{M}}_{\mathcal{N}}^s(\mathcal{O}_n)) = \mathcal{M}_{\mathcal{N}}(\mathcal{O}_n)$

where $\mathcal{DR}_{\mathcal{N}}(\mathcal{O}_n)$ is the diagnosis result based on the entire set of explanations and $\mathcal{M}_{\mathcal{N}}(\mathcal{O}_n)$ is the entire set current estimated of markings.

Proof: The proof is straightforward using Proposition 2 and Assumption 2. ■

We can derive the set of strictly minimal explanations of the first observed event $\underline{\mathcal{L}}_{\mathcal{N}}^s(t^o)$ running the algorithm **Alg_min_exp** with the additional termination condition:

> if there exist two markings M_i, M_j that are reached backwards from M_{fin} by firing σ_i and σ_j ($M_{fin} \overset{\sigma_i}{\rightsquigarrow} M_i$ and $M_{fin} \overset{\sigma_j}{\rightsquigarrow} M_j$) such that $\vec{\sigma_i} \geq \vec{\sigma_j}$ then M_i is deleted

This condition implies that:

1) the computation does not continue backwards from a *solution-end* node
2) if a marking M_i is reached backwards from M_{fin} firing σ_i and there is a minimal explanation σ_k that is already derived such that $\vec{\sigma_i} \geq \vec{\sigma_k}$ then M_i is deleted

The extension to a sequence of observed events is then straightforward.

Remark 4: Notice that the tabular algorithm proposed in [GCS05] to calculate the set of strictly minimal explanations for a PN with acyclic unobservable sub-nets can be easily adapted for PN with trap unobservable circuits.

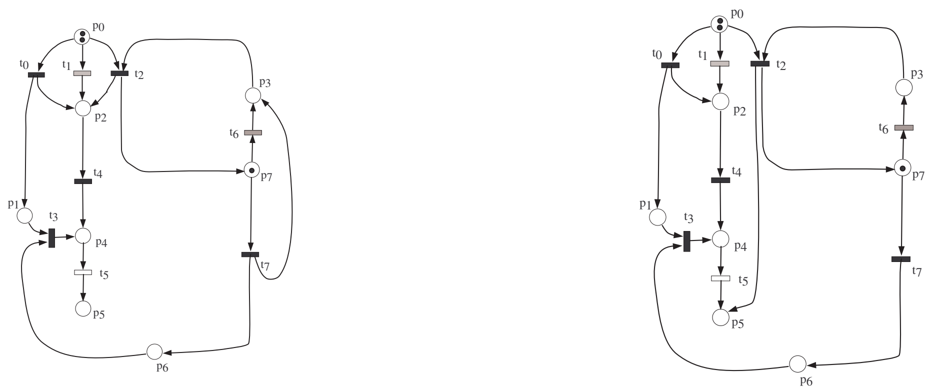

Fig. 6. A a PN with trap unobservable circuits (left) and a general PN (right)

Example 4: Consider the PN model $\langle \mathcal{N}, M_0 \rangle$ displayed in Fig. 6-left. $\mathcal{N}$ is a PN with unobservable trap circuits since $\mathcal{N}$ has two unobservable circuits $p_3t_2p_7t_6$ and $p_3t_2p_7t_7$ and both circuits contain the set of places $\{p_3, p_7\}$ that is a trap. t_5 is the only observable transition. t_1 and t_6 are the fault transitions. The set of minimal explanations of the first occurrence of t_5 is:

$$\underline{\mathcal{L}}_{\mathcal{N}}(t^o) = \left\{ \tau_1 = t_0t_4t_5; \tau_2 = t_1t_4t_5; \tau_3 = t_0t_7t_3t_5; \tau_4 = t_6t_2t_4t_5; \tau_5 = t_6t_2t_0t_7t_3t_5; \tau_6 = t_6t_2t_6t_2t_4t_5 \right\}$$

The set of strictly minimal explanations is:

$$\underline{\mathcal{L}}^s_{\mathcal{N}}(t^o) = \left\{ \tau_1 = t_0t_4t_5; \tau_2 = t_1t_4t_5; \tau_3 = t_0t_7t_3t_5; \tau_4 = t_6t_2t_4t_5; \right\}$$

τ_5 is not a strictly minimal explanation because $\vec{\tau}_3 \leq \vec{\tau}_5$. The strictly minimal explanation τ_3 can be extended by firing the string $\sigma = t_2t_6$ and $\vec{\tau}_3 + \vec{\sigma} = \vec{\tau}_5$.

Similarly τ_6 is not a strictly minimal explanation because $\vec{\tau}_4\vec{\tau}_6$. The strictly minimal extension τ_4 can be extended by firing the string $\sigma = t_2t_6$ and $\vec{\tau}_4 + \vec{\sigma} = \vec{\tau}_6$.

Consider now the PN model $\langle \mathcal{N}', M_0 \rangle$ displayed in Fig. 6-right which contains an unobservable circuit that is not a trap. t_5 is the only observable transition. t_1 and t_6 are the fault transitions. The set of minimal explanations of the first occurrence of t_5 is:

$$\underline{\mathcal{L}}'_{\mathcal{N}}(t^o) = \{\tau_1 = t_0t_4t_5; \tau_2 = t_1t_4t_5; \tau_3 = t_0t_7t_3t_5; \tau_4 = t_6t_2t_0t_7t_3t_5; \tau_5 = t_6t_2t_6t_2t_0t_7t_3t_5\}$$

The set of strictly minimal explanations is:

$$\underline{\mathcal{L}}^s_{\mathcal{N}}(t^o) = \{\tau_1 = t_0t_4t_5; \tau_2 = t_1t_4t_5; \tau_3 = t_0t_7t_3t_5\}$$

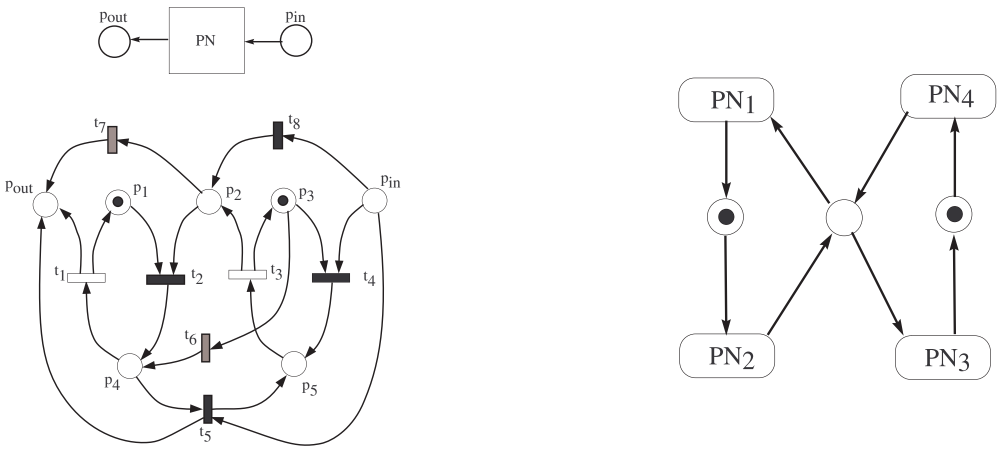

Fig. 7. The PN model of a component -left. Four components that interact via common places - right.

We have that τ_3 cannot be extendable neither by the string $\sigma' = t_2 t_6$ nor by the string $\sigma = t_6 t_2$ and consequently $UR'(\underline{\mathcal{M}}^s) \neq \underline{\mathcal{M}}$. This illustrates why Theorem 5 is not valid for PNs with unobservable circuits that are not traps.

E. Final remarks

We have discussed in this section the detection for sure of a single fault. The extension to the detection of the occurrence for sure of multiple faults is straightforward. Consider the set of fault events partitioned as $\mathcal{T}_f = \mathcal{T}_{\mathtt{F}_1} \cup \ldots \mathcal{T}_{\mathtt{F}_m}$, where a subset of fault transitions $\mathcal{T}_{\mathtt{F}_i}$ $i = 1, \ldots, m$ model a fault of kind $\mathtt{F}_\mathtt{i}$. Given the observation generated by the plant $\mathcal{O}_n$, we say that a fault of kind $\mathtt{F}_i$ happened for sure in the plant at least $\mathtt{k}_i$ times if any explanation $\tau \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)$ contains at least $\mathtt{k}_\mathtt{i}$ appearances of fault transitions that belong to $\mathcal{T}_{\mathtt{F}_i}$, and equality holds for at least one explanation $\tau' \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)$, i.e.: $\forall \tau \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)$, $\sum_{t \in \mathcal{T}_{\mathtt{F}_i}} (\overrightarrow{\tau}(t)) \geq \mathtt{k}_i$ and $\exists \tau' \in \mathcal{L}_{\mathcal{N}}(\mathcal{O}_n)$ s.t. $\sum_{t \in \mathcal{T}_{\mathtt{F}_i}} (\overrightarrow{\tau}(t)) = \mathtt{k}_i$. Then it is easy to show that if the classical diagnoser detects e.g. that a fault of kind $\mathtt{F}_\mathtt{i}$ occurred for sure at least $\mathtt{k}_\mathtt{i}$ times, that a fault of kind $\mathtt{F}_\mathtt{j}$ occurred for sure at least $\mathtt{k}_\mathtt{j}$ times etc. then the reduced diagnoser detects the same thing.

In a companion paper [JBB] we will show why the backward search for minimal explanations of local observations allows for a distributed implementation of the fault diagnosis algorithms. The main property of the backward search that enables this decomposition is the fact that the initial marking does not have to be known completely in order to apply the algorithm, unlike what is needed in the case of the forward search and the centralized observer. In order to illustrate this distributed implementation, consider the following simple scheme (see Fig. 7) where the overall plant description is given by a set of components (each component modeled as a PN) and their interactions (modeled as shared places). This example illustrates the application of the method that we presented in this paper to the modular/distributed monitoring of a large plant.

Fig. 7-right represents a plant comprising 4 interacting components. The components are similar except for the partition of the set of observable and unobservable transitions of the components that may be different. Consider an arbitrary component of the model (Fig. 7-left).

The normal sequence of operations of a component transfers a token received at p_{in} to p_{out} executing $t_4t_3t_2t_1$. However due to some internal failures each component may fail to accomplish this task. E.g. the fault event t_6 removes the token from p_3 and makes impossible the transfer of the input token via the desired sequence of operations. Transitions t_5 and t_8 model recovery actions, whereas t_7 also models a fault.

As already mentioned each component has its own partition of the transitions into observable and unobservable transitions, as well as its own observation labeling function. Assume for the component displayed in Fig. 7-left that the set of observable transitions is $\mathcal{T}_o = \{t_1, t_3\}$ while all the other transitions are unobservable. Assume that the local agent that monitors $Comp_j$ displayed in 7-left observes the label of t_1 and this is the first observation generated by the plant.

We have the set of minimal explanations of t_1 in the local component given by $\{\tau_1 = t_8t_2t_1; \tau_2 = t_6t_1\}$ where τ_1 requires that one token has entered p_{in}, but τ_2 is a valid minimal explanation whatever happens outside of $Comp_j$.

Similarly if the label of t_3 is observed first in the plant, the set of minimal explanations of t_3 in the local component is given by $\{\tau_3 = t_4t_3; \tau_4 = t_6t_5t_3; \tau_5 = t_8t_2t_5t_3\}$ where τ_3 and τ_4 require that one token was delivered from a neighbouring component whereas τ_5 requires that two tokens were delivered from neighbouring components.

We analyze then in the neighbouring components how the required token(s) can be delivered. For a plant that comprises a large number of components there will be typically a small number of components that need to be analyzed, i.e. only those components that contain places from which there are oriented paths comprising only unobservable transitions that lead to the input places of the observable transition whose label was emitted.

In a distributed setting where each component is supervised by a local agent [JB05], the agents exchange information about the tokens that could have exited/entered different components, computing the set of minimal explanations of the observation of the overall plant by consistent pairs of locally allowable traces. E.g. for the local minimal explanations that require tokens via p_{in}, the local agent that supervises the component displayed in Fig. 7-left must ask the neighbouring agents about the possibility that these neighbouring components sent the required tokens to place p_{in} of the local component $Comp_j$.

Notice that local computations are possible even though the marking of a component is only partially known (e.g. the marking of p_{in} is not precisely known). Moreover under some technical conditions we have shown in [Jir06] that a local agent can derive in absence of any communication with its neighbours, a local preliminary diagnosis that is an overdiagnosis of the diagnosis result derived by a centralized agent for that component w.r.t. the detection of the faults that for sure happened in that component. This will be the subject of a paper in preparation [JBB].

V. Conclusions

The research is motivated by the need to designing distributed fault diagnosis algorithms for large and complex systems where inputs/output signals are sent/received by components placed in different locations [GL03],[FBHJ05],[JB05]. The lack of observation of the interactions of a component with its neighbors, the unreliability of the communication channels, as well as the requirement that the local agents should be able to provide the diagnosis of their component in any situation make the distributed diagnosis problem very difficult.

Beside its use for designing a distributed diagnosis algorithm the backward analysis obtained in this paper can be deployed for the centralized monitoring of large PN models. It is well known that for large plants a diagnoser-automaton may become too large to handle. This is because for a given sequence of observed labels the centralized monitoring requires the calculation of the entire set of complete explanations, involving the enumeration of very large sets of markings. The method for the centralized monitoring of large PN models proposed in this paper relies on the construction of a reduced observer that considers in a given state fewer markings than the classical observer. However, all the markings considered by the classical observer can be obtained from the markings considered by a reduced observer, by firing unobservable transitions. The size of the set of markings considered as states of the reduced observer is in general a lot smaller than the size of the set of markings considered as states of the classical observer. Moreover, it is possible at any time if required to derive the set of markings estimated by the classical observer.

We have shown that backward search for deriving the set of minimal explanations of the received observation leads to a plant diagnosis result that equals the centralized diagnosis result based on the set of complete explanations at least for the detection of the faults that for sure happened in the plant. This makes possible the centralized monitoring of very large plants since the complexity of the calculations does not depend on the entire plant size but only on the size of the largest sub-net that contains only unobservable events.

In this paper we have considered the case of untimed PN models where an abstract notion of time is introduced via the partial order relation between the events in the net unfolding. As a future work we plan to extend the methodology presented in this paper for PN models that explicitly consider the time as a continuous and quantifiable parameter (e.g. Time Petri Nets).

Another direction to extend this research is to consider the case of a large plant with uncertain observation [LZ02], i.e. the plant observation may be corrupted, randomly delayed or lost.

VI. Acknowledgments

This research was partially sponsored by The Belgian Found for Scientific Research (BOF) and by the Belgian Program on Inter-university Poles of Attraction (IAP) initiated by the Belgium State, Prime Minister's Office for Science, Technology and Culture. The first author was also supported by a Doctoral Fellowship from the Research Council of Ghent University (BOF doctoraatbeurs).

References

[AIN00] P.A. Abdulla, S.P. Iyer, and A. Nylen. On unfolding unbounded Petri Nets. In Springer-Verlag, editor, *Computer Aided Verification, LNCS*, volume 1855, 2000.

[CKV95] J. Cardoso, L.A. Kunzle, and R. Valette. Petri Net based reasoning for the diagnosis of Dynamic Discrete Event Systems. In *6th International Fuzzy Systemes Association World Congress*, pages 333–336, July 1995.

[DRvB04] G. Delzanno, J-F. Raskin, and L. van Begin. Covering sharing trees: A compact data structure for parameterized verification. *Software Tools for Technology Transfer*, 5(2):268–297, 2004.

[Eng91] J. Engelfriet. Branching processes of Petri Nets. *Acta Informatica*, 28(6):575–591, 1991.

[ERV96] J. Esparza, S. Romer, and W. Volger. An improvement of McMillan's unfolding algorithm. *LNCS*, 1055:87–106, March 1996. Springer-Verlag.

[Esp94] J. Esparza. Model checking using net unfoldings. *Science of Computer Programming*, 23(2):151–194, 1994.

[FBHJ05] E. Fabre, A. Benvensite, S. Haar, and C. Jard. Distributed monitoring of concurrent and asynchronous systems. *Journal of Discrete Event Dynamic Systems*, 15(1):33–84, March 2005.

[FRSB02] A. Finkel, J-F. Raskin, M. Samuelidis, and L. Van Begin. Monotonic extensions of Petri Nets: forward and backward search revisited. *Electronic Notes on Theoretical Computer Science*, 68(6), 2002.

[GCS05] A. Giua, D. Corona, and C. Seatzu. State estimation of λ-free labeled Petri Nets with contact-free nondeterministic transitions. *Journal of Discrete Event Dynamic Systems*, 15(1):85–108, March 2005.

[GL03] S. Genc and S. Lafortune. Distributed diagnosis for DES using Petri Nets. In *Conference on Applications and Theory of Petri Nets (ATPN'03)*, Eindhoven, The Nederlands, 2003.

[GS02] A. Giua and C. Seatzu. Observability of Place/Transition Nets. *IEEE Transactions On Automatic Control*, Vol. 47(9):1424–1437, 2002.

[GW93] P. Godefroid and P. Wolper. Using partial orders for efficient verification of deadlock freedom and safety properties. In *Formal Methods in Systems Design*, volume 2(2), pages 149–164, 1993.

[JB04] G. Jiroveanu and R. K. Boel. Contextual analysis of Petri Nets for distributed applications. In *MTNS'04 Conference*, Leuven, Belgium, 2004.

[JB05] G. Jiroveanu and R.K. Boel. Distributed diagnosis for Petri Net models with unobservable interactions via common places. In *44th Conference on Decision and Control (CDC'05)*, Sevilla, Spain, 2005.

[JBB] G. Jiroveanu, R.K. Boel, and B. Bordbar. Distributed on-line monitoring for networks of Petri Nets, with unobservable interactions via common places. in preparation.

[Jir06] G. Jiroveanu. *Fault diagnosis for large Petri nets*. PhD thesis, Ghent University, Gent, Belgium, 2006.

[KM69] R.M. Karp and R.E. Miller. Parallel Program schemata. *Journal of Computer and Systems Science*, 3(2):147–195, May 1969.

[Kos82] S.R. Kosaraju. Decidability of reachability in vector addition systems. *Proceedings of 14th Annual ACM Symphosium Theory Computing*, San Francisco:267–281, May 1982.

[LA94] L.Portinale and C. Anglano. B-W analysis: a backward reachability analysis for diagnostic problem solving suitable to parallel implementation. In *15th Int. Conference on Application and Theory of Petri Nets, LNCS*, volume 815, pages 39–58, Zaragoza, Spain, 1994.

[Lip76] R.J. Lipton. The reachability problem requires exponential space. *New Haven, CT, Yale University, Department of Computer Science, Res. Rep. 62*, January 1976.

[LZ02] G. Lamperti and M. Zanella. Diagnosis of Discrete Event Systems from uncertain temporal observations. *Artificial Intelligence*, 138:91–137, 2002.

[MBL00] H. Marchand, O. Boivineau, and S. Lafortune. Optimal Control of Discrete Event Systems under Partial Observation. Technical report, INRIA Rennes, 2000.

[McI98] S. McIlraith. Explanatory diagnosis: Conjecturing actions to explain observation. In 6^{th} *Int. Conf. on Principles of Knowledge Representation and Reasoning*, 1998.

[McM92] K. L. McMillan. Using unfoldings to avoid the state space explosion problem in verification of asynchronous circuits. In Springer-Verlag, editor, *4th International Workshop on Computer Aid Verification, LNCS*, volume 663, 1992.

[Mur89] T. Murata. Petri Nets: Properties,Analysis and Applications. *Proceedings IEEE*, 77(4):541–580, April 1989.

[NAH+98] J.L. Nielsen, H.R. Andersen, H. Hulgaard, G. Behrmann, K. Kristoffersen, and K.G. Larsen. Verification of large state/event systems using compositionality and dependency analysis. In *International Conference on Tools and Algorithms for Construction and Analysis*, pages 201–216, 1998.

[OW90] C.M. Ozvern and A.S. Willsky. Observability of Discrete Event Systems. *IEEE Transactions on Automatic Control*, Vol. 35(7):797–806, 1990.

[SJ94] V.S. Srinivasan and M.A. Jafari. Fault detection/monitoring using Timed Petri Nets. *IEEE Transactions On Systems Managment and Cybernetics*, 23(4):1155–1162, 1994.

[SSL+95] M. Sampath, R. Sengupta, S. Lafortune, S. Sinnamohideen, and D. Teneketzis. Diagnosability of Discrete Event Systems. *IEEE Transactions on Automatic Control*, 40(9):1555–1575, 1995.

[Val90] A. Valmari. Stubborn sets for reduced state space generation. In Lecture Notes in Computer Science, editor, *In Advances in Petri Nets*, volume 483, pages 491–515. Springer Verlag, 1990.