Types of Asynchronous Diagnosability and the Reveals-Relation in Occurrence Nets

Stefan Haar

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Types of Asynchronous Diagnosability and the Reveals-Relation in Occurrence Nets

Stefan Haar

N° 6902

April 15, 2009

Thème COM

ISSN 0249-6399 ISRN INRIA/RR--6902--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY - ÎLE-DE-FRANCE

Types of Asynchronous Diagnosability and the Reveals-Relation in Occurrence Nets

Stefan Haar*

Thème COM — Systèmes communicants
Équipes-Projets MExIco

Rapport de recherche n° 6902 — April 15, 2009 — 31 pages

Abstract: We consider asynchronous diagnosis in (safe) Petri net models of distributed systems, using the partial order semantics of occurrence net unfoldings. Unlike the classical case, observability and diagnosability will appear in two different forms each: a strong form associated to interleaving semantics, and a weak form characteristic of nonsequential processes, and requiring an asynchronous progress assumption on those processes. We give algebraic characterizations for both types, and give verification methods. Sufficient conditions for strong diagnosability are derived from linear semiflows. The study of weak diagnosability leads us to the analysis of a relation in occurrence nets, first presented in [14]: given the occurrence of some event a that *reveals* b, the occurrence of b is inevitable; here b may be concurrent to, or even in the future of a. We show that the *reveals*-relation can be effectively computed on a suitable bounded prefix of the unfolding, and show its use in asynchronous diagnosis. Based on this relation, a decomposition of the Petri net unfolding into *facets* is defined, yielding an abstraction technique that preserves and reflects maximal partially ordered runs.

Key-words: Asynchronous Systems, partial order semantics, Petri Nets, Diagnosability

* Work done in part while the author was with School of Information Technology and Engineering, University of Ottawa, Canada. Email: Stefan.Haar@inria.fr or haar@lsv.ens-cachan.fr. INRIA Saclay - LSV ENS Cachan

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

Types of Asynchronous Diagnosability and the Reveals-Relation in Occurrence Nets

Résumé : L'article étudie le diagnostic asynchrone dans des modèles de Réseaux de Petri (saufs), sous la sémantique en ordre partiel des dépliages en réseaux d'occurrences. Contrairement au cas classique, les propriétés d'observabilité et de diagnosticabilité y apparaissent chacune sous deux formes différentes: une version forte associée à la sémantique d'entrelacement et une forme faible, caractéristique des processus non-séquentiels, qui nécessite d'imposer une hypothèse de progrès asynchrone sur ces processus. Nous donnons des caractérisations algébriques pour les deux types, et développons des méthodes de vérification. Des conditions suffisantes pour la diagnosticabilité forte seront obtenues grâce aux invariantes linéaires. L'étude de la diagnosticabilité faible conduit à l'analyse d'une relation intrinsèque aux réseaux d'occurrences, que nous avons présentée dans [14]: étant donné l'occurrence d'un évènement a qui *revèle* b, l'occurrence de b est inévitable; dans ce cas, b peut intervenir en parallèle avec a, voire dans le futur de a. Nous démontrons que la rélation *reveals* peut être effectivement calculée à partir d'un préfixe adéquat du dépliage, et indiquons son utilisation en diagnostic asynchrone. A partir de cette relation, nous definissons une décomposition du dépliage du réseaux de Petri en *facettes*, ce qui donne une technique d'abstraction qui conserve et reflète les exécutions partiellement ordonnées maximales.

Mots-clés : Systèmes asynchrones, Sémantique en ordre partiel, Réseaux de Petri, Diagnosticabilité

Contents

1 Introduction

In highly distributed networked systems, events occur in an asynchronous way; moreover, the supervisor needs to receive alarms from sensors that are generally at a non-negligible distance. Due to asynchronicity between the system and its supervision, alarms collected at different distant sensors can not be meaningfully given a temporal precedence. This generates particular challenges for model-based fault diagnosis, as one has to revise the mathematical representation of systems and their behaviour. In particular, it is appropriate to leave the usual *interleaving* semantics which describes system behaviour by sequences of events: throughout this paper, we will follow the approach of [8, 9] in which

- the system is modeled as a (safe) Petri net, thus taking into account the local and asynchronous nature of states and transitions, and
- the semantics on which diagnosis operates is that of *partially ordered executions* as obtained through the partial order *unfolding* of Petri nets.

Petri nets (see e.g. [27, 24, 16]) and their partial order unfoldings [23, 6, 18] have been increasingly used in recent years for both fault diagnosis [8, 9, 13] and control (see e.g. [12]) of asynchronous discrete event systems. The advantage of partial order semantics lies in the space reduction for representing nonsequential processes that have a high degree of parallelism. In unfoldings, sets of concurrent events are not ordered, which means they have to be represented only once (by one partial order) rather than by giving all their interleavings whose number is exponential in the size of the concurrent set. See also the discussion in [9] and the discussion in the reference [7], entirely dedicated to the necessity of *true concurrency* in the study of distributed discrete event systems.

The purpose of the present article is to investigate *diagnosability* for Petri net models under the partial order perspective. Not surprisingly, the work of

Sampath et al.'s [28] classical characterization of diagnosability in languages of words obtained as *sequential* runs of *automata* will carry over - partly- to the asynchronous setting where the languages are formes by *nonsequential* runs of *Petri nets*. However, important differences will become apparent between diagnosis in interleaving semantics on the one hand and in partial order semantics on the other. Our analysis leads us to distinguish *weak* and *strong* versions of both observability and diagnosability. In short, strongly diagnosable systems allow fault diagnosis under any policy of execution, even those in which some subprocesses may move on quickly while others halt; for weak diagnosability, diagnosis needs only be succesful in executions that have all parts progress in a balanced way. This *progress* problem is completely absent from automata models.

We will also consider different methods for verification of these properties. In the context of strong diagnosability, we will give *sufficient* conditions derived from the theory of net *invariants*. The case of weak diagnosability is different, since it does not allow to focus on interleaved behaviour; it exhibits phenomena that are intrinsic to *concurrency* in system behaviour. It motivates a deeper analysis of the structure of occurrence nets, leading to the *reveals* relation $\triangleright$ which we first pointed out (under the name of *covering* relation) in [14]. It connects pairs (a, b) of events such that a *reveals* b in the sense that whenever a occurs, b must have occurred or will eventually occur as well. We will define the relation $\triangleright$, prove its key properties, and show that it can be effectively computed off-line on a bounded prefix of the model unfolding.

Once the $\triangleright$-relation is known, it can be used, e.g., to detect and identify invisible fault events: the observation of a allows to deduce that any b revealed by a either has already occurred, or will inevitably eventually occur (possibly in the future of a, or in parallel). In a similar way, the design of a controller for the PN system can use the fact that in order to prevent an event b, it is sufficient to prevent some a (by forcing occurrence of some x that is incompatible with a) that reveals b; here, a and b can be *concurrent*. This fact allows, in principle, to formulate diagnosis in terms of eventual occurrence, which generalizes both (a posteriori) diagnosis and prediction.

A further application of the *reveals* relation is in a possible reduction of the size of occurrence net representations by suitable abstractions. *Facets* are subnets of the unfolding in which *any* two events reveal one another. As a consequence, if some event in a facet occurs, eventually all other events of the facet have to occur. Facets enjoy some nice structural properties; their study opens the way to a new topic of *qualitative diagnosability* which is the subject of future work.

The paper is organized as follows:

Section 2 gives basic definitions; Section 3 recalls the asynchronous diagnosis methodology from [8,9,13], and defines weak and strong diagnosability concepts. The characterizations for the two properties are given in Section 4. Section 5 investivates gives procedures for its effective verification. The *reveals* relation is introduced and studied in Section 6; Section 7 presents and analyzes abstractions into facets and associated diagnosability issues, and Section 8 concludes.

2 Definitions

Nets and homomorphisms A *net* is a triple $N = (\mathcal{P}, T, F)$, where $\mathcal{P}$ and T are disjoint sets of *places* and *transitions*, respectively, and $F \subset (\mathcal{P}\times T)\cup(T\times\mathcal{P})$ is the *flow relation.* In figures, places are represented by circles, and marked places are highlighted in thick; rectangular boxes represent transitions, and arrows represent F. Let $<$ be the transitive closure of F and $\leqslant$ the reflexive closure of $<$. For node $x \in P \cup T$, call ${}^\bullet x \triangleq \{x' \mid F(x', x)\}$ the *preset*, and $x^\bullet \triangleq \{x' \mid F(x, x')\}$ the *postset* of x; further, let $\lceil x \rceil \triangleq \{x' \mid x' < x\}$ be the *prime configuration* (see below) or *cone* of x, and $\lfloor x \rfloor \triangleq \lceil x \rceil \backslash \{x\}$ the *pre-cone* of x.

A *net homomorphism* from N to N' is a map $\pi : P \cup T \longmapsto P' \cup T'$ such that:

1. $\pi(P) \subseteq P'$, $\pi(T) \subseteq T'$, and
2. $\pi_{|{}^\bullet e} : {}^\bullet e \to {}^\bullet\pi(e)$ and $\pi_{|e^\bullet} : e^\bullet \to \pi(e)^\bullet$ induce bijections, for every $e \in E$.

Homomorphisms between nets allow to formalize branching processes, see below.

Definition 1 *Two nodes x, x' of a net N are* in conflict, *written $x\#x'$, if there exist $t, t' \in T$ such that* (i) $t \neq t'$, (ii) ${}^\bullet t \cap {}^\bullet t' \neq \emptyset$, *and* (iii) $t \leqslant x$ *and* $t' \leqslant x'$. *A node x is said to be in* self-conflict *iff $x\#x$. An* occurrence net *(ON) is a net $ON = (B, E, F, \mathbf{c}_0)$, with the elements of B called* conditions *and those of E* events*, satisfying the additional properties :*

1. *no self-conflict:* $\forall x \in B \cup E : \neg[x\#x]$;
2. *$\leqslant$ is a partial order:* $\forall x \in B \cup E : \neg[x < x]$;
3. $\forall x \in B \cup E : |\lceil x \rceil| < \infty$;
4. *no backward branching:* $\forall b \in B : |{}^\bullet b| \leq 1$.
5. *the set $\mathbf{c}_0 \triangleq \min(ON)$ of $\leqslant$-minimal nodes of ON is contained in B.*

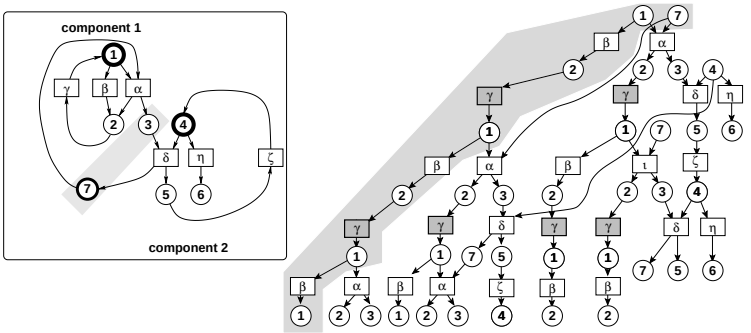

Figure 1: A Petri net (left) and a prefix of its unfolding

A prefix *of ON is any subnet spanned by a downward closed subset $\mathcal{R} \subseteq B \cup E$, i.e. such that for every $x \in \mathbf{K}$, $\lceil x \rceil \subseteq \mathcal{R}$. Prefix $\mathbf{K}$ is a* configuration *iff it is conflict-free, i.e. $x \in \mathbf{K}$ and $x \# y$ imply $y \notin \mathbf{K}$. Denote as $\mathbf{Con}(ON)$ the set of ON's configurations. Call any $\subseteq$-maximal element of $\mathbf{Con}(ON)$ a* ***run*** *of ON; the set of runs is denoted as $\Omega(ON)$ or simply Ω if no confusion can occur.*

The right hand side of Figure 1 shows an occurrence net with one configuration highlighted.

In the above definition, we have added restriction 5.) which is not required, e.g., in [5], without loss of generality and for convenience. Note further that, as a consequence of property 3), $B \cup E$ is well-ordered by $\leqslant$, i.e. there exist no infinite strictly decreasing sequences. Occurrence nets are useful to represent executions of Petri nets, see below: essential dynamical properties are visible via the topological structure of the acyclic graph. Nodes x and x' are *concurrent,* written x **co** x', if neither $x \leqslant x'$, nor $x' \leqslant x$, nor $x \# x'$ hold. A *co-set* is a set $\mathcal{X}$ of pairwise concurrent conditions; a maximal co-set $\mathcal{X}$ w.r.t. set inclusion is called a *cut*, and generically denoted by the symbol $\mathbf{c}$; in particular, $\mathbf{c}_0$ is a cut, called the *initial* cut of ON.

We note for future reference that occurrence nets are a special case of *event structures* [25]:

Definition 2 *A tuple $\mathcal{E} = (E, <, \#)$ is a* ***prime event structure*** *or* ***PES*** *iff:*

1. *$(E, <)$ is a countable, partially ordered set,*
2. *$\lceil e \rceil$ is finite for all $e \in E$,*
3. *$\# \subseteq E \times E$ is symmetric and irreflexive, and $\forall\ x, y, z \in E$: $x \# y$ and $y < z$ together imply $x \# z$.*

Petri Nets Let $N = (P, T, F)$ be a finite net. A *marking* of net N is a multi-set $M \in \mathfrak{M}(P)$. A *Petri net* (PN) is a pair $\mathcal{N} = (N, M_0)$, where $M_0 \in \mathfrak{M}(P)$ is an *initial* marking. $T \in T$ is *enabled* at M, written $M \xrightarrow{t}$, if for all $p \in {}^\bullet t$, $M(p) \geqslant 1$. If $M \xrightarrow{t}$, then t can *fire*, leading to $M' = (M - 1_{{}^\bullet t}) + 1_{t^\bullet}$, where symbol 1 denotes the set indicator function; write in that case $M \xrightarrow{t} M'$. The set $\mathbf{R}(M_0)$ contains the markings of $\mathcal{N}$ *reachable* through $\longrightarrow$. A Petri net $\mathcal{N} = (N, M_0)$ is *k-safe* if $M(p) \leqslant k$ for all $M \in \mathbf{R}(M_0)$ and all $p \in \mathcal{P}$. 1-safe nets are simply called *safe*. Only safe nets are considered in this article; their reachable markings will be represented as *sets* $M \subseteq P$.

Example: In Figure 1, the left hand side shows a safe Petri net which will be used as an example throughout. The marked places $1, 4$, and 7 are indicated by thick circles; in this initial marking, the enabled transitions are α, β and η. As the inscriptions suggest, the net represents a simple model of fault propagation between two components. Initially, both components are in an *ok* state reflected by the initial marking. Then, one may have occurrences of fault α or β or η. In the latter case, component 2 will remain permanently in a faulty state (reflected by place 6), regardless of the actions in component 1. On the side of comoponent 1, fault β has no outside effect; it can be repaired by occurrence of γ. Fault α, on the other hand, marks place 3 and thus enables induced fault δ on the side

of component 2, thus exhibiting propagation of a fault; in this model, that fault can be repaired on either component, through transitions γ and ζ, respectively.

Branching Processes and Unfoldings The branching process semantics reflects the partial order behavior of Petri nets in occurrence nets, thus allowing for structural analysis.

Definition 3 *A* branching process *of the safe Petri net $\mathcal{N} = (P, T, F, M_0)$ is given by a pair $\pi = (ON, \pi)$, where $ON = (B, E, G, \mathbf{c}_0)$, and π is a homomorphism from ON to N, such that:*

1. *π is injective on $\mathbf{c}_0$, and $\pi(\mathbf{c}_0) = M_0$;*
2. *for all $e, e' \in E$, ${}^\bullet e = {}^\bullet e'$ and $\pi(e) = \pi(e')$ together imply $e = e'$.*

For π_1, π_2 two branching processes, π_2 is a prefix *of π_1, written $\pi_2 \sqsubseteq \pi_1$, if there exists an injective homomorphism ψ from ON_2 into a prefix of ON_1, such that ψ induces a bijection between the initial cuts $\mathbf{c}_0^1$ and $\mathbf{c}_0^2$, and the composition $\pi_1 \circ \psi$ coincides with π_2.*

By theorem 23 of [5], there exists a unique (up to an isomorphism) $\sqsubseteq$-maximal branching process, called the *unfolding* of $\mathcal{N}$ and denoted $\mathcal{U}(\mathcal{N})$; by abuse of notation, we will also use $\mathcal{U}(\mathcal{N})$ for the occurrence net obtained by the unfolding.

Following [6], the unfolding of $\mathcal{N}$ can be computed using the canonical algorithm given below (we omit any cut-off criteria here since they are not essential for our purposes). For any branching process $\pi = (ON_\pi, \pi_\pi)$ of $\mathcal{N} = (P, T, F)$ - with $ON_\pi = (B_\pi, E_\pi, G_\pi)$ - , denote as $pex(\pi) \subseteq T \times \mathfrak{P}(B)$ the set of *possible extensions* of π, i.e. of the pairs $(t, \mathcal{X})$ such that

- $\mathcal{X}$ is a co-set of ON_π,
- ${}^\bullet t = \pi_\pi(\mathcal{X})$,
- $\in E_\pi$ contains no event e such that $\pi_\pi = t$ and ${}^\bullet e = \mathcal{X}$.

The unfolding procedure adapted from [6] for safe Petri net $\mathcal{N} = (N, M_0)$ is then:

- Let $\mathbf{c}_0 \triangleq M_0 \times \{\emptyset\}$ and initialize $\pi = (\mathbf{c}_0, \emptyset, \emptyset)$.
- For given $\pi = (ON_\pi, \pi_\pi)$ with $ON_\pi = (B_\pi, E_\pi, G_\pi)$, compute $pex(ON_\pi)$ and replace
  - E_π by $E_\pi \cup pex(ON_\pi)$,
  - B_π by $B_\pi \cup V$, where $V \triangleq \{(P, e) \mid e \in pex(ON_\pi), p \in \pi_\pi(e)^\bullet\}$, and
  - G_π by $G_\pi \cup U$, where

$$U \triangleq \{(b, (t, \mathcal{X})) \mid (t, \mathcal{X}) \in pex(ON_\pi), b \in \mathcal{X}\} \cup \left\{e, ((P, e)) \mid e \in pex(ON_\pi), p \in \pi_\pi(e)^\bullet\right\}.$$

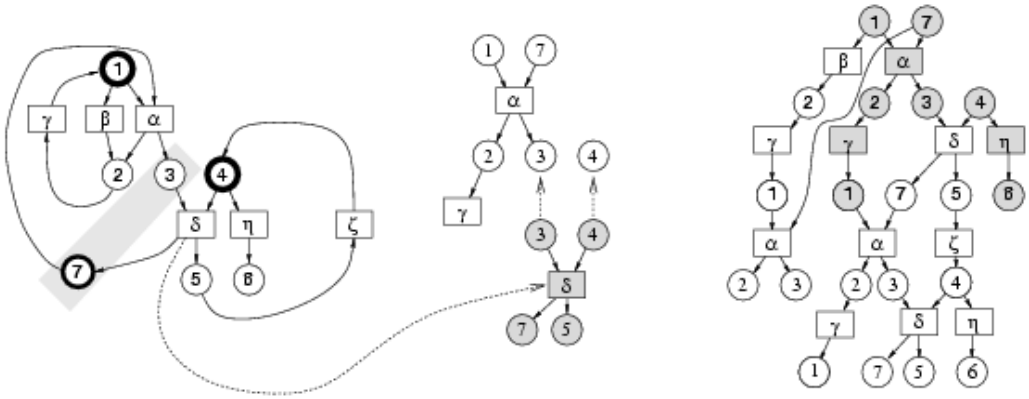

Figure 2: Procedure of Unfolding for a Petri net; example taken from [8]

Fig. 2 gives an illustration, taking up the running example from [8, 13]. A Petri net $\mathcal{N}$ is shown on the left, and a branching process $\pi = (ON, \pi)$ of $\mathcal{N}$ on the right hand side. Conditions are labeled by places, events by transitions. A configuration is shown in grey. The mechanism for constructing the unfolding of $\mathcal{N}$ is illustrated in the middle.

We note the following technical properties for future reference:

Lemma 1 ([27, 5]) *If $\mathcal{U} = (ON, \pi)$ is the unfolding of safe Petri net $\mathcal{N} = (P, T, F, M_0)$, then:*

1. *If $\mathbf{c} \subseteq B$ is a cut then so is $\mathbf{c}' \triangleq (\mathbf{c} \backslash {}^\bullet e) \cup e^\bullet$ for every e such that ${}^\bullet e \subseteq \mathbf{c}$;*
2. *for any two conditions x, y, x **co** y implies $\pi(x) \neq \pi(y)$.*
3. *π maps all cuts of ON into $\mathcal{N}$-markings in $\mathbf{R}(M_0)$, and every marking in $\mathbf{R}(M_0)$ is the π-image of a cut of ON.*

Every *finite* configuration κ terminates at a cut, which we denote $\mathbf{c}_\kappa$. The mapping $\kappa \mapsto \mathbf{c}_\kappa$ is bijective; for each cut $\mathbf{c}$, the union of the cones of all conditions in $\mathbf{c}$ yield the unique configuration κ such that $\mathbf{c} = \kappa_\mathbf{c}$. Moreover, one has the following correspondences:

- If κ is a configuration of $\mathcal{U}_\mathcal{N}$ with $\mathcal{N} = (N, M_0)$, then every occurrence sequence σ obtained as a linear order extension of the partial order $\leqslant_\kappa$ yields a firable transition sequence of $\mathcal{N}$. Conversely, every firable transition sequence of $\mathcal{N}$ corresponds to a linear order extension of some configuration of $\mathcal{U}_\mathcal{N}$. To sum up: the nonsequential executions of $\mathcal{N}$ are in one-to-one correspondence with the configurations of $\mathcal{U}(\mathcal{N})$. We will therefore speak of $\mathcal{N}$*'s configurations* and write $\mathbf{Con}(\mathcal{N}) \triangleq \mathbf{Con}(\mathcal{U}_\mathcal{N})$ and $\Omega(\mathcal{N}) \triangleq \Omega(\mathcal{U}_\mathcal{N})$.
- for every reachable marking $M \subseteq P$ of $\mathcal{N}$, there exists at least one cut $\mathbf{c}$ of $\mathcal{U}(\mathcal{N})$ such that $\pi(\mathbf{c}) = M$ for all p, and the unique configuration κ such that $\mathbf{c}_\kappa = \mathbf{c}$ is such that execution of κ takes M_0 to M; write $M_0 \xrightarrow{\kappa} M$ for this.

- Conversely, every finite configuration κ corresponds to a unique reachable marking $M(\kappa)$ given by $M(\kappa) \triangleq \pi(\mathbf{c}_\kappa)$. We call configurations that lead to the same marking *marking equivalent*, and write $\kappa \equiv_M \kappa'$ iff $M(\kappa) = M(\kappa')$.

3 Asynchronous Diagnosis and Diagnosability

3.1 The Diagnosis Procedure

We focus on extending the diagnosis approach developped in [8]. Its purpose is the identification of possible system runs (the *explanations*) that are compatible with a partially ordered alarm pattern (the *observation*). The heart of its algorithm is the unfolding of the synchronized product net $\mathcal{N} \times \mathcal{A}$ obtained from $\mathcal{N}$ and an alarm pattern $\mathcal{A}$, where $\mathcal{A}$ is given as a Petri net whose net is an unbranched finite occurrence net. The synchronized product glues together transitions of $\mathcal{N}$ with corresponding alarms in $\mathcal{A}$. The unfolding $\mathcal{U}_{\mathcal{N}\times\mathcal{A}}$ then yields all the explanations that $\mathcal{N}$ can give for $\mathcal{A}$. In fact, the configurations κ of $\mathcal{N}$ that *explain* $\mathcal{A}$ are those for which $\mathcal{U}_{(\mathcal{N}\times\mathcal{A})}$ contains a corresponding configuration $\overline{\kappa}$ whose projection (i) to the alarm set yields $\mathcal{A}$, and (ii) to $\mathcal{N}$-nodes yields κ.

Here, we are interested in the capacity of the product approach to detect occurrence of a *fault event* ϕ, that is, the shape of alarm patterns and their explanations are secondary. Instead, we focus on the question whether and when the observations allow to deduce that ϕ must have occurred.

Reminder: Diagnosability for interleaved sequences Before introducing *fault diagnosis* and the diagnosability problem for the asynchronous setting, let us recall the formal definition of Sampath et al. [28] for diagnosis in interleaved models (see also Lin [22]): let $\mathcal{L}$ be a prefix-closed language (the behavior of the system to be diagnosed) over the event alphabet $\mathbf{Alph}$, denote $O \subseteq \mathbf{Alph}$ the set of *observable* and $UO \triangleq \mathbf{Alph} \backslash O$ that of *unobservable* events[1]. Denote $P : \mathbf{Alph}^* \to O^*$ the projection to observable words, that is, the homomorphism that erases all unobservable events and leaves observable ones unchanged; moreover, let $\phi \in UO$ be a *fault*[2]. Then $\mathcal{L}$ is *diagnosable* iff there exists $n \in \mathbb{N}$ such that, for any word $\mathcal{L} \ni w = w'\phi$, any $v \in \mathbf{Alph}^*$ s. th. $wv \in \mathcal{L}$ and $|v| \geqslant n$ satisfies

$$x \in P^{-1}\left[P\left(wv\right)\right] \quad \Rightarrow \quad |x|_\phi \geqslant 1. \tag{1}$$

Here, $|u|$ denotes total length, and $|u|_\phi$ the number of ϕ-occurrences in word u. Condition (1) means that every behavior x that produces the same sequence of observable events as wv does, contains at least one fault event: all extensions of w of at least length n will make the fault apparent. A polynomial time algorithm for testing diagnosability is given by Kumar et al. [17]; see also Yoo and Lafortune [30].

[1] see Kumar and Shayman [19] on observability and co-observability.

[2] for simplicity, we assume there is only one *fault type* in the sense of [28]; the developments given below extend to the general case.

Asynchronous Diagnosis We shall be using analogous terminology and symbols here.

Definition 4 *Let* $\mathcal{N} = (P, T, F, M_0)$ *be a Petri net with unfolding* $\mathcal{U} = (B, E, G, \pi)$, *and* **Alph** *an alarm alphabet containing the empty symbol* ε*; further, let* $\chi : T \to$ **Alph**, *for* **Alph** *some non-empty alphabet, be a* labeling function *associating alarms to system transitions. Call* silent *or* unobservable *transitions the elements of* $UO \triangleq \chi^{-1}(\varepsilon)$, *and let* $O \triangleq T \backslash UO$ *be the set of* observable *transitions, and* $\phi \subseteq UO$ *the* fault *to be diagnosed.*

Here, $\mathcal{N} = (\mathcal{P}, T, F, M_0)$ is the underlying "true" system, with the places in P representing the local states. This framework allows for *erasing* (i.e. labeling by ε) and *ambiguity* (the same label for distinct events). Without loss of generality, $\phi \in UO$; in fact, a fault that is indicated by an alarm needs not be diagnosed; the diagnosis problem concerns *silent* faults, whose associated "alarm" is ε. Set $E_\phi \triangleq \pi^{-1}(\{\phi\})$, $E_O \triangleq \pi^{-1}(O)$, and $E_{UO} \triangleq E \backslash E_O$. The approach carries over to sets of faults without deep changes, yet we will focus on the case with one fault event to keep notations simpler. We will illustrate below the effect of different labeling functions on the same net; that is, for $\mathcal{N}$ fixed, we will ask what constraints λ must satisfy to achieve observability and diagnosability. Requiring that e.g. transition α of the net on the left hand side of figure 1 be observable, means in practice that an active sensor needs to be put on the corresponding plant part, allowing to record some alarm $\lambda(\alpha)$ on each occurrence of α. Conversely, if we determine that visibility of α is not necessary, then such a sensor need not be deployed (or, if it is already in place, we need not record its alarms).

Let us return to the basic ingredients. The main difference between the asynchronous setting and the state machine framework of e.g. [2] is in the *languages* considered, that is, in the notion of *behaviour* that underlies the approach. Since the asynchronous semantics of $\mathcal{N}$ is given by the set of nonsequential processes, i.e. the *configurations* of its partial order unfolding $\mathcal{U}_\mathcal{N}$, these take over the role that is played by the word-language for automata in the above. Let therefore

$$\mathcal{L} \quad \triangleq \quad \{\kappa \cap E \mid \kappa \in \mathbf{Con}(\mathcal{N})\};$$

we will consider configurations as sets of *events*.

Height and Progress As Fig. 3 shows, concurrent systems may exhibit non-sequential processes whose local parts do not progress at the same pace. Suppose the fault to be diagnosed is γ. On some interleaved behaviors, γ may go undetected: if the net performs an infinite number of cycles involving α and β, no decision on γ will be available. However, it is clear that γ occurs with certainty under this behavior unless the right hand part of the net remains idle forever. In most applications, the assumption that "something will eventually happen", is realistic: in particular, if a transition is enabled, it will eventually either fire or become disabled by another transition. In order to parallel the interleaved case, we therefore consider two different notions of diagnosability over unfoldings:

- the restrictive one of **strong diagnosability** which requires faults to be detected by *all* infinite executions;

- and **weak diagnosability** which requires that all faults be detectable at least on those executions which progress in a balanced way on all local components.

The examples will show that the two notions do not coincide. To formalize things, we have to dwell on the notion of *height*, which is the measure for progress of the system in logical time. Measuring progress for concurrent processes can be done by counting events, like for sequences; this leads to a notion of *length*, see [10]. This length is to be contrasted with *height*, in which the causal relations between events are taken into account: the height of a prefix, e.g. a configuration, is the length of its longest causal chain; call this the *upper height*. A more sophisticated height function measures, so to speak, the advancement of the slowest parts of the process. This concept - which we will call *lower height* - is based on the "measuring scale" of prefixes formed by the prefixes $\mathcal{R}_n$, see below, which are formed by all nodes whose *upper height* is at most n; these prefixes grow uniformly on 'all ends' as n grows.

Let us formalize things now. Set recursively $\|\lceil e \rceil\| \triangleq 1 + \|\lfloor e \rfloor\|$ and

$$\|\mathcal{R}\| \quad \triangleq \quad \sup\{\|\lceil e \rceil\| \mid e \in E \cap \mathcal{R}\} \quad \Big(\textit{where } \sup(\emptyset) \triangleq 0\Big)\,. \tag{2}$$

Using definition (2), we define, for $n \in \mathbb{N}_0$, $\mathcal{R}_n$ to denote the maximal prefix whose height does not exceed n; call $\mathcal{R}_n$ $\mathcal{N}$'s *nth prime prefix*. Now, applying (2) directly to configurations - seen as special prefixes - yields the *upper height* $\|\kappa\| \in \mathbb{N} \cup +\infty$ for $\kappa \in \mathbf{Con}(\mathcal{N})$. We define the *lower height* $\langle\langle\kappa\rangle\rangle$ of κ as follows :

$$\langle\langle\kappa\rangle\rangle \quad \triangleq \quad \sup\{n \in \mathbb{N} \mid \exists\ \omega \in \Omega :\ \omega \cap \mathcal{R}_n = \kappa \cap \mathcal{R}_n\}\,,\ \textit{where } \sup(\emptyset) = 0 \tag{3}$$

this height defines a metric that is standard in partial order semantics, see e.g. [4, 21, 20]. Of course, $\langle\langle\kappa\rangle\rangle \leqslant \|\kappa\|$, with equality iff either $\langle\langle\kappa\rangle\rangle = +\infty$ or

$$\forall\ \omega,\ \omega' \in \Omega\ :\ \big[(\kappa \sqsubseteq \omega\ \wedge\ \kappa \sqsubseteq \omega') \quad \Rightarrow \quad \big(\omega \equiv_{\mathcal{R}_{\|\kappa\|}} \kappa\big)\big]\ ; \tag{4}$$

call the finite configurations that satisfy (4) **progressive**. By extension, call an arbitrary configuration $\mathbf{K}$ progressive iff all its finite truncations $(\mathbf{K} \cap \mathcal{R}_n)_{n \in \mathbb{N}}$ are progressive A non-progressive configuration $\mathbf{K}$ may allow an extension by events whose height is inferior to $\|\mathbf{K}\|$; progressive configurations cannot be extended without increasing the lower height.

The term of 'progressive' configurations is justified by the fact that their local processes all progress in a *fair* way, none of them lagging behind indefinitely[3]

Example: In Fig. 3, we have $\|\kappa\| = \langle\langle\kappa'\rangle\rangle = 2$, but $\|\kappa_2\| = 3$ and $\langle\langle\kappa_2\rangle\rangle = 2$. Clearly, κ is progressive and κ' is not; however, κ' can be extended into progressive configurations, e.g. $\kappa' \cup \{\beta, b\}$.

Live and dead configurations In analogy with the *liveness* requirement in [28], let us say that a configuration κ is **dead** iff $\kappa \sqsubseteq \kappa'$ implies $\kappa' \in \mathbf{FCon}$. On a finite run, absence or presence of faults can eventually be verified, so they can be discarded from our analysis, so we assume henceforth no dead configuration exists.

[3]In fact, progressive executions for safe nets are necessarily fair in the sense that any transition which is enabled an infinite number of times must also fire an infinite number of times, i.e. cannot be ignored indefinitely long. The converse is not true; fair executions do not necessarily lead to progressive configurations.

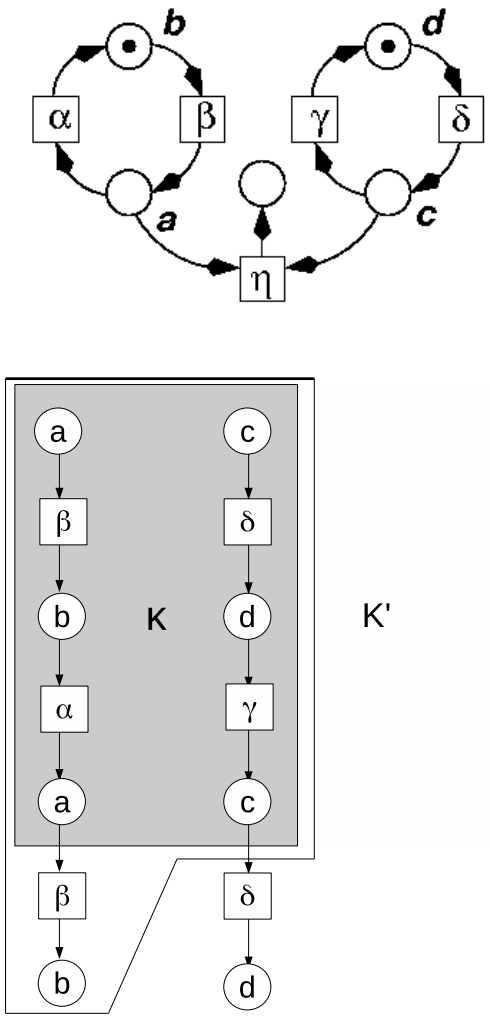

Figure 3: Left: a Petri Net ; right : some of its configurations

Faulty configurations Denote as $\mathcal{L}_{\mathbf{prog}}$ the set of *progressive* configurations; observe that $\mathcal{L}$ and $\mathcal{L}_{\mathbf{prog}}$ are both prefix closed. For $\kappa \in \mathcal{L}$ let κ_O be the labeled partial order induced by κ on $\kappa \cap E_O$. $\kappa \sim_O \kappa'$ iff κ_O and κ'_O are isomorphic. Let $\equiv_\phi$ be the equivalence on $\mathcal{L}$ given by $\kappa \equiv_\phi \kappa'$ iff $[\kappa \cap E_\phi = \emptyset \iff \kappa' \cap E_\phi = \emptyset]$; that is, two configurations are ϕ-equivalent if either both contain a fault, or neither of them does. This finishes our preparations.

Definition 5 *Let $\mathcal{L}$ be a configuration language, i.e. a set of finite partially ordered configurations such that $\kappa \in \mathcal{L}$ and $\kappa' \sqsubseteq \kappa$ imply $\kappa' \in \mathcal{L}$, and height measure $H : \mathcal{L} \to [0, \infty)$ be either $H \equiv \langle\langle \bullet \rangle\rangle$ or $H \equiv \| \bullet \|$. $\mathcal{L}$ is H-**diagnosable** w.r.t. O and ϕ iff there exists $n \in \mathbb{N}$ such that for all $\kappa_\phi \in \mathcal{L}$ having a maximal event $e \in E_\phi$, it holds that every $\kappa \in \mathcal{L}$ such that* (a) $\kappa_\phi \sqsubseteq \kappa$, (b) κ *is not dead, and* (c) $H(\kappa) \geqslant H(\kappa_\phi) + n$, *satisfies:*

$$\forall \kappa' \in \mathcal{L} : \ \kappa' \sim_O \kappa \quad \Rightarrow \quad E_\phi \cap \kappa' \neq \emptyset. \tag{5}$$

Now, we lift diagnosability from languages to nets:

Definition 6 *Let $\mathcal{N} = (P, T, F, M_0)$ a safe Petri net, $\mathcal{U}_\mathcal{N} = (B, E, G, \mathbf{c}_0)$ its unfolding, and $\mathcal{L}$ and $\mathcal{L}_{\mathbf{prog}}$ as above. Further, let $E_O \triangleq \pi^{-1}(O) \subseteq E$ be the set of observable events, $\phi \notin E_O$ a and $E_\phi \triangleq \pi^{-1}(\{\phi\})$. Then:*

1. *$\mathcal{N}$ satisfies($\mathcal{OBS}$) (for O) iff for all $\kappa, \kappa' \in \mathcal{L}$,*

$$(\kappa \sqsubseteq \kappa') \wedge (\kappa \neq \kappa') \wedge (M_\kappa = M_{\kappa'}) \quad \Rightarrow \quad (\kappa \not\sim_O \kappa') \tag{6}$$

2. *$\mathcal{N}$ satisfies $\mathcal{WOBS}$ (for O) iff (6) holds for all $\kappa, \kappa' \in \mathcal{L}_{\mathbf{prog}}$.*
3. *$\mathcal{N}$ is called* **(strongly) diagnosable** *(satisfies $\mathbb{D}$) w.r.t. O and ϕ iff*
   (a) *$\mathcal{N}$ satisfies $\mathcal{OBS}$ (for O), and*
   (b) *$\mathcal{L}$ is $\langle\langle \bullet \rangle\rangle$-diagnosable in the sense of Definition 5 w.r.t. O and E_ϕ.*
4. *$\mathcal{N}$ is called* **weakly diagnosable** *(satisfies $\mathbb{W}$) w.r.t. O and ϕ iff*
   (a) *$\mathcal{N}$ satisfies ($\mathcal{WOBS}$), and*
   (b) *$\mathcal{L}_{\mathbf{prog}}$ is $\| \bullet \|$-diagnosable w.r.t. O and ϕ.*

Some remarks are in order. First, *strongly* diagnosable nets are also diagnosable in the sense of [28] (see (1)), and vice versa: Consider the interleavings of its runs instead of the partially ordered runs. The existence of the constant bound n, such that the fault can be decided with certainty at most n actions after occurrence of the fault, corresponds to the fact that only a finite number of invisible transition firings can occur concurrently to any visible transition.

Secondly, note that while strong diagnosability trivially implies weak diagnosability, the converse is not true[4]: In Fig. 3, suppose β is the fault action $\phi \triangleq \beta$, $O = \{\alpha\}$, and for $m \in \mathbb{N}$, let $\kappa^{(m)}$ be the smallest configuration such that *(i)* β never occurs on $\kappa^{(m)}$, and *(ii)* δ occurs exactly m times on $\kappa^{(m)}$. Then $\|\kappa^{(m)}\| = 2m + 1$, yet $\kappa^{(m)} \sim_O \kappa^{(1)}$ for all m, so the system is not strongly diagnosable. Note that the $\kappa^{(m)}$ are not progressive; all *progressive* configurations of height at least $2k + 1$ contain at least k instances of $\alpha \in O$, from which it follows directly that the system is weakly diagnosable.

[4] similarly for weak and strong *observability*

4 Characterization of Diagnosability

After these preparations, we are now ready to state and prove our characterizations of weak and strong diagnosability. As in the classical setting, diagnosability is *violated* iff the system is able to perform two indiscernible, non-fault-equivalent cycles. That is, there must be O-equivalent configurations κ_1 and κ_2 having O-equivalent extensions κ_1' and κ_2' such that $M(\kappa_i) \leqslant M(\kappa_i')$, and such that κ_1' and κ_2' are not ϕ-equivalent; then the system may repeat that cyclic behavior indefinitely, without a decision about occurrence of faults. In fact:

Theorem 1 *With labeling $\lambda : T \to$ **Alph**, and ϕ, O, UO, $\mathcal{L}$ and $\mathcal{L}_{\mathbf{prog}}$ as above, a safe Petri net $\mathcal{N} = (P, T, F, M_0)$ is* **strongly diagnosable** *w.r.t. O and ϕ iff it satisfies $\mathcal{OBS}$ and*

$$\forall\ \kappa_1, \kappa_2, \kappa_1', \kappa_2' \in \mathcal{L} : \left[\left\{ \begin{array}{rcl} & \kappa_1 \sim_O \kappa_2 \ \wedge\ \kappa_1' \sim_O \kappa_2' \ \wedge\ \kappa_1 \neq \kappa_1' & \\ \wedge & \forall\ i \in \{1,2\} : \left(\begin{array}{rcl} & M_{\kappa_i} = M_{\kappa_i'} & \\ \wedge & \kappa_i \sqsubseteq \kappa_i' & \end{array} \right) & \end{array} \right\} \Rightarrow \kappa_1' \equiv_\phi \kappa_2' \right] . \quad (7)$$

$\mathcal{N}$ is ***weakly diagnosable*** *w.r.t. O and ϕ iff $\mathcal{WOBS}$ the restriction of (7) to $\mathcal{L}_{\mathbf{prog}}$ hold.*

Note, before we proceed to the proof, that (7) allows $\kappa_2 = \kappa_2'$ in the assumption. In preparation of the proof below, denote as $\kappa_1 \circ \kappa_2$ the *concatenation* configuration obtained from κ_1 in $\mathcal{N} = (N, M_0)$ and κ_2 in $(N, M(\kappa_1))$ appended after κ_1. Define powers of configurations by $\kappa^1 \triangleq \kappa$ and $\kappa^{k+1} \triangleq \kappa^k \circ \kappa$.

Proof: We show the strong diagnosability case; the result for weak diagnosability is obtained by replacing $\mathcal{L}$ by $\mathcal{L}_{\mathbf{prog}}$. For the **"only if"** part, let $\kappa_i \sqsubseteq \kappa_i'$, $i \in \{1,2\}$, constitute a violation of (7), i.e.

1. without loss of generality, $\kappa_2' \cap E_\phi \neq \emptyset$ and $\kappa_1' \cap E_\phi = \kappa_1 \cap E_\phi = \emptyset$;
2. $\kappa_i' = \kappa_i \circ \mu_i$, where μ_1 contains at least one event, and finally
3. $\kappa_i' \sim_O \kappa_i$ and $M_{\kappa_i} = M_{\kappa_i'}$.

From 2, it follows that a copy of μ_i can be appended to κ_i' as well, and so forth; let $\kappa_i^k \triangleq \kappa_i \circ \mu_i^k$ be the configuration obtained after appending k copies of μ_i to κ_i. Observe that $\|\kappa_1^k\| \geqslant \max(k, \|\kappa_1\|)$. Thus $\|\kappa_1^k\| \to \infty$ as $k \to \infty$. Now, by assumption we have $\kappa_2^k \sim_O \kappa_2$; further, by construction, $\mu_2 \cap E_\phi$ and therefore $\kappa_2^k \cap E_\phi = \emptyset$. It follows that (5) is violated.

To show the **"if"** part, suppose (5) does *not* hold: for every $n \in \mathbb{N}$, there exists $\kappa(n) \in \mathcal{L}$ such that
(1) some $e \in E_\phi$ is $\leqslant$-maximal in $E \cap \kappa(n)$, and *(2)* there exist $\kappa_1(n), \kappa_2(n) \in \mathcal{L}$ such that

$$(\kappa(n) \sqsubseteq \kappa_1(n)) \ \wedge\ (\|\kappa_1(n)\| \geqslant \|\kappa(n)\| + n) \ \wedge\ (\kappa_2(n) \sim_O \kappa_1(n)) \ \wedge\ (\kappa_2 \cap E_\phi = \emptyset) .$$

Assume first that one can choose κ_1' with $\kappa_1 \sqsubseteq \kappa_1' \sqsubseteq \kappa_1(n)$ such that $M_{\kappa_1} = M_{\kappa_1'}$, $\kappa_1 \sim_O \kappa_1'$, and $\kappa_1 \neq \kappa_1'$; then we are done by setting $\kappa_2' \triangleq \kappa_2$. Thus assume

$$\forall\ \kappa_1' : \left[\left\{ \begin{array}{l} \kappa_1 \sqsubseteq \kappa_1' \sqsubseteq \kappa_1(n) \\ \kappa_1 \sim_O \kappa_1' \\ M_{\kappa_1} = M_{\kappa_1'} \end{array} \right\} \Rightarrow \kappa_1 = \kappa_1' \right] . \quad (8)$$

For any $\kappa_1 \sqsubseteq \kappa_1(n)$, let $U(\kappa_1, n)$ be the set of configurations $\kappa_2 \sqsubseteq \kappa_2(n)$ such that $\kappa_2 \sim_O \kappa_1$. For any reachable marking M of $\mathcal{N}$, let $S_1(M, n)$ be the set of configurations κ_1 such that *(i)* $\kappa_1 \sqsubseteq \kappa_1(n)$ and *(ii)* $M = M(\kappa_1)$. Let $\mathcal{K}$ be the number of all reachable markings of $\mathcal{N}$. Then for all $n > \mathcal{K}$, there is at least one marking M such that $|S_1(M, n)| \geqslant 2$; repeating the argument, one finds using (8) that for all $n > \mathcal{K}^2$ there exists a marking M such that $|S_1(M, n)| > \mathcal{K}$. With

$$U_2(M,n) \quad \triangleq \quad \left\{ \begin{array}{l} \kappa_2 \in \mathcal{L}, \\ \kappa_2 \sqsubseteq \kappa_2(n) \end{array} \;\middle|\; \begin{array}{r} \exists \kappa_1 \in S_1(M,n): \\ \kappa_1 \sim_O \kappa_2 \end{array} \right\},$$

we therefore have $|U_2(M, n)| > \mathcal{K}$. Thus there exist $\kappa_2, \kappa_2' \in U_2(M, n)$ such that $\kappa_2 \neq \kappa_2'$ and $M\kappa_2 = M_{\kappa_2'}$. By definition of $U_2(M, n)$, $\kappa_1 \sim_O \kappa_2$ and $\kappa_1' \sim_O \kappa_2'$. Since, by construction, $\kappa_1 \sqsubseteq \kappa_1' \sqsubseteq \kappa_1(n)$ and $M_{\kappa_1} = M_{\kappa_1'}$, property (7) is violated, q.e.d. □

Note that in the above, the treatment of the progressive and non-progressive cases does not require a different proof: the difference is only in the set of configurations over which the different κ-variables in the proof may range. However, strong and weak diagnosability are not equivalent, and they allow for very different verification methods, see below.

5 Verification of Diagnosability

We will now describe criteria for strong and weak observability and diagnosability, respectively. First we will turn to the *strong* case; after that, the ideas emerging from the discussion of the weak case will lead to the investigation of the *reveals* relation in the next section. For strong Diagnosability, we will derive sufficient conditions from Petri net invariants; we follow the terminology and notation of [3]. For a net $N = (P, T, F)$, the **incidence matrix** $\mathbf{N} : (P \times T) \to \{-1, 0, 1\}$ is given by

$$\mathbf{N}(p,t) \quad \triangleq \quad \left\{ \begin{array}{rcl} 0 & : & (pFtFp) \vee \neg(pFt \vee tFp) \\ 1 & : & (pFt) \wedge \neg(tFp) \\ -1 & : & (tFp) \wedge \neg(pFt) \end{array} \right. .$$

For a sequence $\sigma \in T^*$ of transitions, the **Parikh vector** $\overline{\sigma} : T \to \mathbb{N}$ is given by $\overline{\sigma}(t) \triangleq |\sigma|_t$, i.e. the number of occurrences of transition t in σ. The action of transitions of N can be described by $\mathbf{N}$ (this is Lemma 2.12 in [3]): For $\sigma \in T^*$ and markings M, M' of N such that $M \xrightarrow{\sigma} M'$, one has the following **Marking Equation**:

$$M' \quad = \quad M + \mathbf{N}\overline{\sigma}. \tag{9}$$

Note that $\mathbf{N}$ is independent of the marking, i.e. represents a *net* (P, T, F) rather than a *Petri net*.

Let $N = (P, T, F)$ be a net. A **T-invariant** (also called **T-semiflow**) of N is a rational-valued solution of the equation $\mathbf{N} \cdot x = 0$. Equivalently ([3], Proposition 2.36), a mapping $J : T \to \mathbb{Q}$ is a T-Invariant of N iff for all $p \in P$,

$$\sum_{t \in {}^\bullet p} J(t) \quad = \quad \sum_{t \in p^\bullet} J(t). \tag{10}$$

The importance of T-invariants lies in the following property ([3], Proposition 2.37): Suppose M is a marking of N and $\sigma \in T^*$ such that $M \xrightarrow{\sigma}$. Then $\overline{\sigma}$ is a T-invariant of N iff it reproduces M, i.e. $M \xrightarrow{\sigma} M$.

We thus know that $M_\kappa = M_{\kappa'}$ holds iff the "Parikh vector" $\overline{(\kappa', \kappa)}$ given by

$$\overline{(\kappa', \kappa)}(t) \quad \triangleq \quad |\{e \in \kappa' \backslash \kappa \mid \pi(e) = t\}|,$$

satisfies Equation (10). In fact, any linearization σ of the events in $\kappa' \backslash \kappa$ has same Parikh vector, and so the above results apply simultaneously to any such σ. Therefore, (10) can be used to check whether a given marking can *possibly* be reproduced in an unobservable way: in that case, Equation (10) must have a semi-positive solution (i.e. with all entries non-negative and at least one positive entry). Now, any violation of strong diagnosability must be a realization of some unobservable firing sequence that can be repeated arbitrarily often. Since the net is finite, the existence of such a sequence entails that some T-invariant must be fired in that sequence. As a consequence, we have:

Lemma 2 *If for all semi-positive solutions $v \in \mathbb{N}^T$ of (10), there exists $t \in O$ such that $v(t) > 0$, then $\mathcal{N}$ satisfies $\mathcal{OBS}$.*

However, a given $\overline{\tau} \in \mathbb{N}^T$ may satisfy (10) without corresponding to any firing sequence enabled in M; Fig. 4 gives an example, see below. The solutions of (10) are only *candidates* for cycles. The purely structural condition of Lemma 2 is sufficient, but not necessary.

Examples

1. In the net from Fig. 3, the T-invariants are (with coordinates ordered by alphabetic order on $\{\alpha, \beta, \gamma, \delta, \eta\}$) $(1, 1, 0, 0, 0)$, $(0, 0, 1, 1, 0)$ and their positive linear combinations. Thus η is covered by none of them. Weak observability requires that at least one of $\{\alpha, \beta, \gamma, \delta\}$ be observable; for strong observability, one each out of $\{\alpha, \beta\}$ and out of $\{\gamma, \delta\}$ must be observable.

2. For the net in Fig. 5, again with alphabetic ordering, the incidence matrix is

$$\mathbf{N} \quad = \quad \begin{pmatrix} -1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} ;$$

   the T-invariants are $(0, 1, 1, 0, 0, 0)$, $(1, 0, 1, 1, 0, 1)$, and their positive linear combinations. Suppose e.g. α is observable and none else; then we need also β to be observable, since otherwise an infinite loop with β is possible and unobservable. For strong observability and strong diagnosability, we obtain from Lemma 2 the following sufficient criterion:

$$[\gamma \in O] \quad \vee \quad [\beta \in O \wedge (|O \cap \{\alpha, \delta, \zeta\}| \geqslant 1)] \,. \tag{11}$$

Following the reasoning for α and β above, we see - by inspection of all cases - that this criterion is also *necessary*

3. The net in Fig. 4 has incidence matrix

$$\mathbf{N} = \begin{pmatrix} -1 & -1 & 0 & 1 & 1 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix},$$

and thus all T-invariants are of the form $(2u, 2u, u, 2u, 2u)^\top$ with $u, v \in \mathbb{N}$. However, these invariants are not realizable (i.e. firable), as γ will never fire. On the other hand, observability and diagnosability are satisfied for any choice of O, even $\emptyset$; compare the unfolding given on the right hand side. The example illustrates that T-invariants provide a much coarser analysis of Petri net behavior than unfoldings.

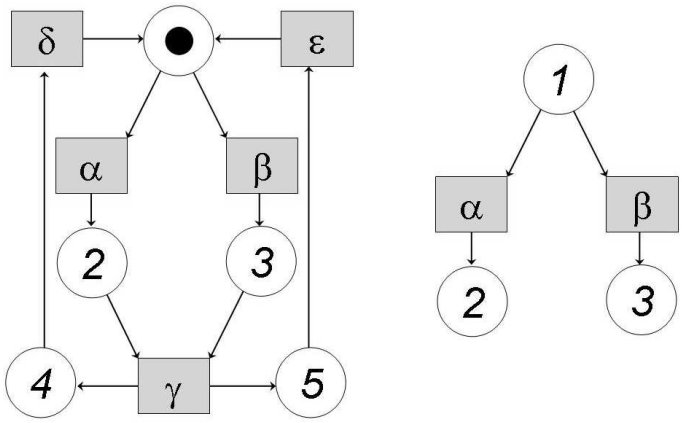

Figure 4: A Petri net $\mathcal{N}$ whose net has a single T-invariant (left), and $\mathcal{N}$'s unfolding (right).

5.1 Using Unfoldings for Checking Weak Diagnosability

Even if unfoldings are infinite in general, any safe Petri net admits finite complete prefixes that contain every reachable marking; this is what allows using branching processes in Model Checking [6, 23]. Methods for obtaining and optimizing such *complete prefixes* have received considerable attention in the literature, see e.g. [18].

Finite Complete Prefix The runs of $\mathcal{U}(\mathcal{N})$ represent all maximal nonsequential executions. That is, any firing *sequence* of $\mathcal{N}$ is obtained as the linear order extension of (some prefix of) some run $\omega \in \Omega(\mathcal{U}(\mathcal{N}))$. If $\mathcal{U}(\mathcal{N})$ is infinite, we are naturally interested in *finite* prefixes of $\mathcal{U}(\mathcal{N})$ that are *complete* in the sense that their analysis allows to derive results for all of $\mathcal{U}(\mathcal{N})$. The definition

and size of such prefixes varies with the intended purpose; see [4, 18] for a systematic treatment. We use here the following definition, similar to that in [12]:

Definition 7 *The order 1 unfolding, denoted $\mathcal{U}_1(\mathcal{N})$, is a finite prefix of the unfolding obtained by stopping the construction of the unfolding when we reach a* ***cut-off*** *event e, i.e., an event such that:*

- *EITHER firing of $\lceil e \rceil$ brings back to the initial marking: $M(\lceil e \rceil) = M_0$;*
- *OR there exists another event e' with the following properties:*
  1. *The prime configuration for e' is a prefix of that of e: $\lceil e' \rceil \subseteq \lceil e \rceil$;*
  2. *the two congurations are marking-equivalent: $M(\lceil e \rceil) = M(\lceil e' \rceil)$.*

In the following we call e' the mirror transition *of e in $\tilde{\mathcal{N}}_1(M_0)$. Once we have constructed $\mathcal{U}_1(\mathcal{N})$, assume we continue the unfolding until we reach an event e such that there exist another event e' with the following properties:*

- *either e' does not belong to $\mathcal{U}_1(\mathcal{N})$ or it is a cut-off event of $\mathcal{U}_1(\mathcal{N})$;*
- *The prime configuration for e' is a prefix of that of e: $\lceil e' \rceil \subseteq \lceil e \rceil$;*
- *the two configurations are marking-equivalent: $M(\lceil e \rceil) = M(\lceil e' \rceil)$.*

The resulting net, denoted $\mathcal{U}_2(\mathcal{N})$, is called order 2 unfolding*; by iterating the above, one obtains a nested family $(\mathcal{U}_n(\mathcal{N}))_{n\in\mathbb{N}}$ of* n-th order unfoldings*.*

Note that the initial definition from [23] used as cutoff criterion the *cardinality*, i.e. $|\lceil e' \rceil| < |\lceil e \rceil|$, which would lead to a shorter prefix in general yet not guarantee completeness w.r.t. computing the *reveals* relation below.In our generalized setting, one can only hope for finite prefixes whose size can be bounded given the net structure, and sufficient to decide diagnosability. The following results show that such prefixes exist, and thus effective offline verification of diagnosability is possible. We have:

Theorem 2 *For a given net $N = (P, T, F)$, there exists a finite number $Z = Z(\mathcal{N})$ such that for any 1-safe marking $M_0 \subseteq P$ of N, the Z-th prefix $\mathcal{R}_Z$ of the unfolding of $\mathcal{N} = (N, M_0)$ is sufficient to verify (strong or weak) diagnosability: if there exist any $\kappa_1, \kappa_1', \kappa_2, \kappa_2'$ such that (7) is violated, one can choose them with this property such that* $\max(\|\kappa_1'\|, \|\kappa_2'\|) \leqslant Z$.

Proof: Call an alarm pattern $\mathcal{A}$ *reducible* iff for all $\kappa \in \mathbf{expl}(\mathcal{A})$, there exist $\kappa_1, \kappa_2, \kappa_3$ such that *(i)* $\kappa = \kappa_1 \circ \kappa_2 \circ \kappa_3$, *(ii)* $\|\kappa_2\| > 0$, *(iii)* $\kappa_1 \circ \kappa_2^+ \circ \kappa_3 \subseteq \mathcal{L}$, and *irreducible* otherwise. Now the result follows, since $\mathcal{K}$ is finite, from the pigeonhole principle once the following claim is proved: the number J of irreducible alarm patterns of $\mathcal{N}$ is bounded above by $2^{\mathcal{K}}$. For this, note that the height $\|\kappa\|$ of any configuration $\kappa \in \mathcal{L}$ that does not contain two comparable markings, i.e. such that $\kappa_1 \sqsubseteq \kappa_2 \sqsubseteq \kappa$ and $M_{\kappa_1} = M_{\kappa_2}$ imply $\kappa_1 = \kappa_2$, is bounded above by $\mathcal{K}$. Hence, all alarm patterns $\mathcal{A}$ whose height exceeds m are reducible, since $\kappa \in \mathbf{expl}(\mathcal{A})$ implies $\|\kappa\| \geqslant \|\mathcal{A}\|$; finally, the number of patterns of height $\mathcal{K}$ or less is bounded above by $2^{\mathcal{K}}$. □

More can be said of the configurations in a prefix of the unfolding of $\mathcal{N}$:

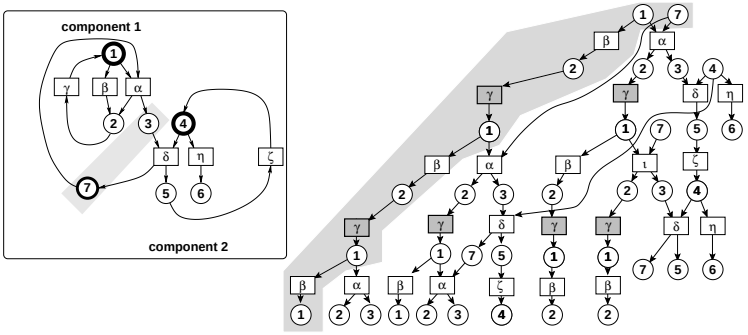

Figure 5: A complete prefix (right) for the net on the left

Lemma 3 *Let $\mathcal{R}$ be any prefix of the unfolding $\mathcal{U}_{ON}$. If there exist* **witnesses of non-diagnosability** *in $\mathcal{R}$, configurations κ_i, κ'_i for $i \in \{1,2\}$ such that the left hand side of (7) holds, but $\kappa'_1 \not\equiv_\phi \kappa'_2$, then κ'_1, κ_2 can be chosen maximal for ON.*

Proof: By assumption, there exist κ''_i such that $\kappa_i \sqsubseteq \kappa''_i \sqsubseteq \kappa'_i$ (and therefore $\kappa''_i \sim_O \kappa'_i$ and $\kappa''_1 \sim_O \kappa''_2$) and κ'''_i a maximal configuration of ON such that $M_{\kappa''_i} = M_{\kappa'''_i}$; hence $\kappa''_1, \kappa''_2, \kappa'''_1, \kappa'''_2$ are also witnesses of non-diagnosability. □

One obtains thus the following algorithm for checking weak diagnosability of $\mathcal{N} = (P, T, F, M_0)$:

(A) Compute a complete prefix γ as above, and its set $max_{ON} \triangleq \Omega(\gamma)$ of maximal configurations.

(B) For any pair κ'_1, κ'_2 of maximal configurations such that $\kappa_1 \sim_O \kappa_2$, check whether there exist $\kappa_i \sqsubseteq \kappa'_i$ such that $\kappa_1 \sim_O \kappa'_1$, $M_{\kappa_1} = M_{\kappa'_1}$ and $\kappa_1 \sqsubseteq \kappa'_1$.

Examples

1. In the context of Fig. 5, we ask under which choices of O the net $\mathcal{N}$ satisfies $\mathcal{OBS}$, and if so, whether $\mathcal{N}$ is then diagnosable for that O and a given fault ϕ. First, we claim that $\mathcal{OBS}$ (and even $\mathcal{WOBS}$) is *equivalent* with (11). In fact, every $\kappa \in max_{ON}$ contains γ-labeled events, so the implications $(\gamma \in O) \Rightarrow \mathcal{OBS} \Rightarrow \mathcal{WOBS}$ are immediate. On the other hand, suppose $\gamma \notin O$; then we deduce from the configuration κ on shaded background in the figure that $\beta \in O$ (otherwise κ and two of its prefixes yield witnesses of non-diagnosability). Inspecting the other non-dead configurations of max_γ in a similar way, we see that $\alpha \notin O$ entails $(\delta \in O) \vee (\zeta \in O)$; we deduce that (11) is necessary for (both weak and strong) observability, and thus for (both weak and strong) diagnosability. Now, let us check sufficiency, i.e. whether (11) makes $\mathcal{N}$ *diagnosable*. For this, let us consider the cases $\phi = \eta$ and $\phi = \beta$. Since we have to respect $\phi \notin O$, (11) is refined in the second case to

$$[\gamma \in O]\,. \tag{12}$$

Consider the set max_η of configurations from max_{ON} that contain an η-event. Inspection of Fig. 5 shows that for $\kappa_\eta \in max_\eta$ and any extension κ'_η of κ_η satisfying either $\|\kappa'\| > \|\kappa\| + 1$ or $\langle\langle\kappa'\rangle\rangle > \langle\langle\kappa\rangle\rangle + 2$, contains a γ-instance. For the other fault label, β, one has that the conjunction of (i) $\phi^{-1}(\beta) \cap \kappa \neq \emptyset$ and (ii) $\|\kappa'\| > \|\kappa\| + 1$ or $\langle\langle\kappa'\rangle\rangle > \langle\langle\kappa\rangle\rangle + 1$, implies $\phi^{-1}(\gamma) \cap \kappa \neq \emptyset$. Thus we conclude that $\gamma \in O$ is necessary and sufficient for $\mathcal{OBS}$, $\mathcal{WOBS}$, $\mathbb{D}$, and $\mathbb{W}$.

2. The net in Fig. 3 is *weakly* diagnosable iff $|O| \geqslant 1$, and *strongly* diagnosable iff $(O \cap \{\alpha, \beta\} \neq \emptyset) \wedge (O \cap \{\gamma, \delta\} \neq \emptyset)$.

3. Fig. 5 shows that the upper bounds on the size of the complete prefix are far from sharp; γ can be chosen moderate if there is a high degree of parallelism in $\mathcal{N}$ and no excessive branching. The efficiency of diagnosability checking thus requires a careful choice of prefixes; see [6, 18].

4. For the net in Fig. 4, the complete prefix allows to detect dead configurations; this is not possible using invariants alone, compare the discussion above.

We will now take a closer look at the relational structure of occurrence nets.

6 The *Reveals* Relation

6.1 Definitions

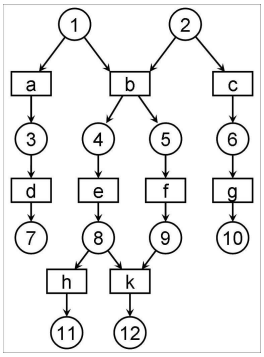

Figure 6: On the relation $\triangleright$

In the above discussion, we use implicitly reasonings of the form 'if x occurs, then y has already occured, or will occur eventually', in the sense that any infinite run that contains x also contains y. Under progress assumption (see above), this means that y *is inevitable given* x. In the context of the occurrence net in Fig. 6, for any run ω,

$$k \in \omega \quad \Rightarrow \quad e \in \omega \Rightarrow b \in \omega; \tag{13}$$

in fact, (13) reflects the inheritance of # under $<$. But one also obtains the following facts in Fig. 6:

$$a \in \omega \iff \neg(b \in \omega) \Longleftrightarrow c \in \omega \tag{14}$$

$$e \in \omega \iff f \in \omega; \tag{15}$$

the reader is invited to check that (14) and (15) follow from the maximality of runs. Now, the inheritance of conflict along causality relations is not sufficient to derive (14) and (15); so how can one formalize the reasoning that leads to them ? One might suspect that, to derive (14 and 15) from the relational structure, one would have to explore the entire set of configurations. We will show here that it suffices to consider an auxiliary relation, computable from the # relation in a finite bounded prefix $\mathcal{R}$ of the unfolding. Let us start formalizing things.

Definition 8 *For a node $x \in (B \cup E)$, the **conflict set** of x is defined as $\#[x] \triangleq \{x' \mid x\#x'\}$. The* root conflict set *is given by $\#_\mu[x] \triangleq \{y \mid x\#y \wedge \forall\, z : z < y \Rightarrow \neg(z\#x)\}$; the symbol $\#_\mu[\bullet]$ is borrowed from [1] where it denotes immediate conflict in event structures. Node x* reveals *y, written $x \triangleright y$, iff $\#[x] \supseteq \#[y]$. Define the* revealed range *of node x as $\triangleright[x] \triangleq \{y \mid x \triangleright y\}$.*

One immediately checks that $\triangleright$ is reflexive and transitive. Moreover, we have:

Lemma 4 ([14]) *$x \triangleright y$ holds iff for all runs ω,*

$$x \in \omega \quad \Rightarrow \quad y \in \omega \tag{16}$$

Proof: If $x \in \omega$ and $y \notin \omega$, there exists a node $z \in \#[y] \cap \omega$; in fact, otherwise $\omega \cup \lceil y \rceil$ would be a configuration, and ω could not be maximal. If $x \triangleright y$, then $z \in \#[x] \cap \omega$, which is impossible, so we must have $\neg(x \triangleright y)$. Conversely, suppose that (16) holds for every ω; then there exists z such that $z\#y$ and $\neg(z\#x)$. But then there exists a run ω_z such that $x, z \in \omega_z$, but by assumption $y \notin \omega_z$, hence (16) is violated for ω_z. □

Relation $\triangleright$ is asymmetric: in fact, in Fig. 8 (left) we have $h \triangleright f$ but $\neg(f \triangleright h)$. On the other hand, $\triangleright$ is not a partial order: consider $e \triangleright f$ and $f \triangleright e$. This is a crucial fact behind the definition of *facets* below. However, the following holds:

Lemma 5 *$x < y$ implies that $y \triangleright x$.*

Proof: By inheritance of #, $x < y$ implies $\#[x] \subseteq \#[y]$. □

As a consequence, we have:

Lemma 6 *$\triangleright[x]$ is a configuration.*

Proof: Since $\lceil x \rceil \subseteq \triangleright[x]$ by Lemma 4, we have $\mathbf{c}_0 \subseteq \triangleright[x]$; thus Lemma 5 implies the result. □

In Fig. 8, we have the following revealed ranges:

$$\triangleright[b] = \triangleright[e] = \triangleright[f] = \{b, e, f\}; \qquad \triangleright[h] = \{b, e, f, h\},\ \triangleright[k] = \{b, e, f, k\};$$
$$\triangleright[a] = \triangleright[d] = \triangleright[c] = \triangleright[g] \quad = \quad \{a, d, c, g\}.$$

The following result is crucial for the feasibility of our approach: it shows that in order to decide whether $x \triangleright y$, it suffices to know $\#_\mu[x]$ and $\#_\mu[y]$:

Theorem 3 *The set $\#[x]$ is generated by $\#_\mu[x]$ through inheritance:*

$$\#[x] \quad = \quad \{z \mid \exists\ y \in \#_\mu[x] :\ y \leqslant z\}\,. \tag{17}$$

As a consequence, $x_1 \triangleright x_2$ iff $\#_\mu[x_1] \supseteq \#_\mu[x_2]$.

Proof: The inclusion$\#[x] \supseteq \{z \mid \exists\ y \in \#_\mu[x] :\ y \leqslant z\}$ being obvious, it remains to show

$$\#[x] \quad \subseteq \quad \{z \mid \exists\ y \in \#_\mu[x] :\ y \leqslant z\}\,. \tag{18}$$

Take any $y \in \#[x]\backslash\#_\mu[x]$. Since $x\#y$, there exist a condition b_1 and events x_1, y_1 such that *(i)* $x_1 \neq y_1$; *(ii)* $b_1 \in {}^\bullet x_1 \cap {}^\bullet y_1$; and *(iii)* $x_1 \leqslant x$ and $y_1 \leqslant y$. Let $n \geq 1$. If $y_n \in \#_\mu[x]$, we are done; otherwise there exist a condition b_{n+1} and events x_{n+1}, y_{n+1} such that *(a)* $x_{n+1} \neq y_{n+1}$; *(b)* $b_{n+1} \in {}^\bullet x_{n+1} \cap {}^\bullet y_{n+1}$; *(c)* $x_{n+1} \leqslant x$ and $y_{n+1} < y_n$. If we find recursively infinitely many such $y_1, y_2, \ldots$, this contradicts property 3) of Definition 1, since $y \geqslant y_1 > y_2 > \ldots$. We conclude that there exists $n \in \mathbb{N}$ such that $y_n \in \#_\mu[x]$, and this proves (18). □

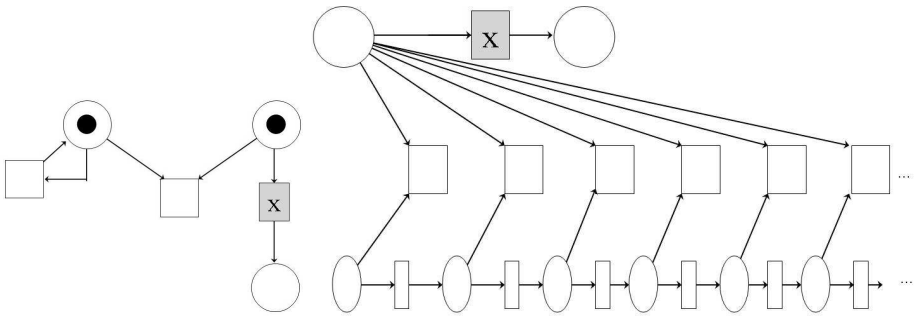

Figure 7: Left: a safe Petri net; right: its unfolding, exhibiting an infinite root conflict set.

So far we were able to reduce the computation of the *reveals* relation to comparison of root conflict sets. These sets can be infinite, as the example in Fig. 6.1 shows: $\#_\mu[x]$ consists of all the events in the central horizontal axis of the figure. However, the relation $\triangleright$ can be effectively computed, by virtue of the following result:

Theorem 4 *Denote as round(x) the smallest n such that x belongs to $\mathcal{U}_n(\mathcal{N})$, and as $\mathbf{K} \triangleq |\mathbf{R}M_0|$ the number of reachable markings of $\mathcal{N} = (P, T, F, M_0)$. Then for any two nodes x, y such that $\neg(x \triangleright y)$, there exists a $\triangleright$-*witness *in $\mathcal{U}_{m+\mathbf{K}-1}$, i.e. a node z such that, with $m \triangleq \max(round(x), round(y))$,*

$$z \ \#\ y \quad \textit{and} \quad \neg\,(z \ \#\ x)\,. \tag{19}$$

Proof: By the assumption $\neg(x \triangleright y)$, some node z satisfying (19) exists; it remains to show that z can be chosen in $\mathcal{U}_{n+\mathbf{K}-1}$. If $x\#y$, we are done immediately, taking x as witness. Thus, assume $\kappa_{xy} \triangleq \lceil x \rceil \cup \lceil y \rceil$ is a configuration, and let M_{xy} be the marking generated by κ_{xy}. Choose $z \in E$ such

that (19) holds (with x, y fixed), and such that no $z' < z$ has that property. Then there exists $u \in E\backslash\{z\}$ that satisfies *1)* ${}^\bullet u \cap {}^\bullet z \neq \emptyset$ and *2)* $u \leqslant x$. Let $\kappa_z \triangleq \lceil {}^\bullet z \rceil \cup \lceil {}^\bullet u \rceil$ (which is a configuration by the way we chose u), and M_z the associated marking. Then by construction,

$$M \xrightarrow{\pi(z)} \quad and \quad M \xrightarrow{\pi(u)}, \tag{20}$$

$$and \qquad \kappa_{xy} \quad \sqsubseteq \quad \kappa_z. \tag{21}$$

If $\max(round(u, z)) > n + \mathbf{K} - 1$, the pigeonhole principle implies that there are two distinct configurations κ_1, κ_2 of $\mathcal{U}$ such that *(1)* $\kappa_{xy} \sqsubseteq \kappa_1 \sqsubseteq \kappa_2 \sqsubseteq \kappa_z$, and *(2)*$\kappa_1 \equiv_M \kappa_2$. We can then replace κ_z by a different configuration κ' that satisfies (20) and (21), obtained by 'removing' the section $\kappa_2\backslash\kappa_1$; in fact, κ' shares κ_1 with κ_z but follows the suffix of κ_1 isomorphic to the suffix of κ_2 in κ_z. Repeat this until no such configurations κ_1, κ_2 can be found; the resulting configuration κ' lies entirely withing $\mathcal{U}_{n+\mathbf{K}-1}$. From (20), we also obtain the existence of an event e such that

$$\pi(e) = \pi(z) \quad \text{and} \quad \kappa' = \lceil e \rceil \cup \lceil u \rceil,$$

since u lies in κ_{xy} and $M' \triangleq M(\kappa')$ satisfies $M' \xrightarrow{z}$ and $M' \xrightarrow{u}$. It follows from the construction that $e \# x$. We claim that

$${}^\bullet e \cap {}^\bullet u \quad = \quad {}^\bullet z \cap {}^\bullet u. \tag{22}$$

In fact, suppose there exists $b \in ({}^\bullet e\backslash{}^\bullet z) \cap {}^\bullet u$. By property 2 of homomorphisms (Def. 2), there must exist $b' \in {}^\bullet z \cap {}^\bullet u$ such that $\pi(b') = \pi(b)$. Then either *(i)* $b' \# b$, *(ii)* $b' < b$, *(iii)* $b < b'$ or *(iv)* b' **co** b. But *(i)* implies $b' \# b$; under *(ii)*, there must exist an event e_0 such that $b' < e_0 < b$, which also implies $b' \# b$; symmetrically, *(iii)* also leads to $b' \# b$; and *(iv)* contradicts Lemma 1.

Consider now the different possibilities for y; we have that:

- if $y \# u$ we are done;
- if $y < u$, then $y < x$, contradicting our assumption;
- Finally, if $u < y$, then we obtain $z \# y$, another contradiction.

Therefore y **co** u must hold. If we assume now that $e \# y$, there must exist an event $v \leqslant y$ such that ${}^\bullet e \cap {}^\bullet v \neq \emptyset$. By reasoning along the same lines as for (22) above, we obtain

$${}^\bullet e \cap {}^\bullet v \quad = \quad {}^\bullet z \cap {}^\bullet v; \tag{23}$$

as a consequence, $z \# y$, contradicting our assumptions. Therefore $\neg(e \# y)$, and we are done. □

In the light of Theorem 4, any safe net allows to compute *reveals* relations for pairs (x, y) of nodes recursively on finite prefixes whose depth grows linearly with $\max(round(x), round(y))$.

6.2 The Reveals-Relation and Weak Diagnosability

Consider again Fig. 5. Every occurrence of δ is detected by a prior occurrence of α, and by a subsequent occurrence of ζ. That is, if δ is a fault event, then it suffices for $\mathcal{N}$ is δ-diagnosable if either δ or α are observable. This can be formalized as a *lifting* of $\triangleright$ to the level of $\mathcal{N}$:

Definition 9 *In $\mathcal{N}$, transition $t_1 \in T$* reveals *$t_2 \in T$, written $t_1 \triangleright_{\mathcal{N}} t_2$, iff for all $e_2 \in \pi^{-1}(t_2)$ there exists $e_1 \in \pi^{-1}(t_1)$ such that $e_1 \triangleright e_2$, where $\triangleright$ is the* reveals *relation in $\mathcal{U}(\mathcal{N})$.*

We have the following obvious result:

Lemma 7 *Let O be as above, and $\phi \in T \backslash O$.*

- *If there exists $t \in O$ such that $t \triangleright_{\mathcal{N}} \phi$, then $\mathcal{N}$ is ϕ-diagnosable, and*
- *if for all $t \in T \backslash O$, there exists $t' \in O$ such that $t' \triangleright_{\mathcal{N}} t$, then $\mathcal{N}$ is observable.*

However, the converse is not true. In fact, consider again Fig. 5. We obtain the following table for $\triangleright_{\mathcal{N}}$ ('+' at (x, y) means that $x \triangleright_{\mathcal{N}} y$, and '-' means $x \not\triangleright_{\mathcal{N}} y$):

$\triangleright_{\mathcal{N}}$	α	β	γ	δ	η	ζ
α	+	−	+	−	−	−
β	−	+	+	−	−	−
γ	−	−	+	−	−	−
δ	+	−	−	+	−	+
η	−	−	−	−	+	−
ζ	+	−	+	+	−	+

Now, let γ be the fault transition, and let α and β be observable. Then $\mathcal{N}$ is clearly β-diagnosable, yet γ is not $\triangleright_{\mathcal{N}}$-revealed by either α or β. We see that $\triangleright_{\mathcal{N}}$ gives sufficient criteria for observability and diagnosability, and allows quick verification of both, if $\triangleright_{\mathcal{N}}$ has been precomputed offline; on the other hand, it has in general to be checked on a prefix of the unfolding (rather than $\mathcal{N}$) whether a particular occurrence of a transition t is revealed by some observable event. Typically, occurrence will not be revealed by occurrences of the same observable transition in all circumstances; however, different occurrences may each be revealed by occurrences of different observable transitions. Smart use of small prefixes and of $\triangleright$ in the unfolding is required for fast verification of observability and diagnosability, and for identification of sufficient observable sets.

7 Facets and $\mathcal{Q}$-Diagnosability

By considering equivalence classes w.r.t. $\triangleright$, an occurrence net ON can be decomposed into subnets that we call *facets*; see Fig. 8. As we will see below, the set of ON's facets is an event structure with the quotient relations induced from ON, and can be represented by an occurrence net. Since also the maximal runs are preserved under the quotient operation (see Theorem 5), many analyses, can be equivalently carried out on the quotient structure.

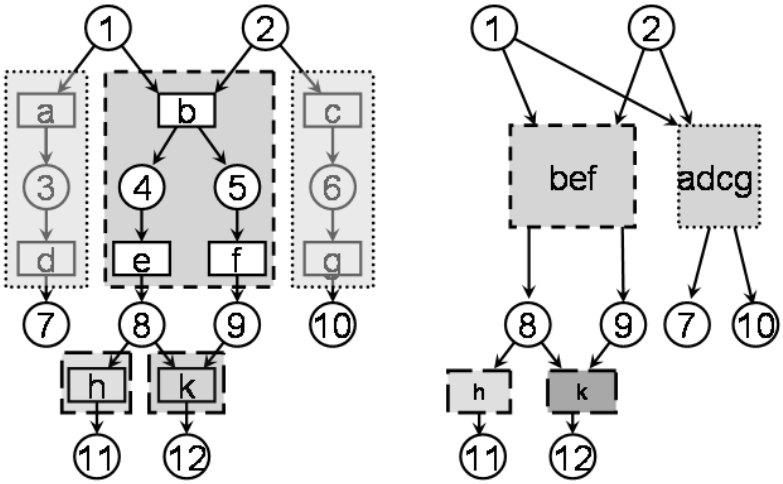

Figure 8: Left: the example from Fig. 6 with facets highlighted; right: the occurrence net obtained from the example through facet abstraction

Definition 10 *A **facet** of ON is a strongly connected component of $\triangleright$, i.e. a maximal set $\psi \subseteq (E \cup B)$ such that for any $x, y \in \psi$, one has $x \triangleright y$ and $y \triangleright x$. Denote as $\psi(x)$ the unique facet that contains x.*

In Fig. 8, the facets are $\{a, d, c, g\}$, $\{b, e, f\}$, $\{h\}$, $\{k\}$; the right hand side shows the occurrence net obtained by abstracting every facet into a single event. Concerning the shape of facets, we obtain easily:

Lemma 8 *Facets are conflict-free.*

Proof: Suppose $e_1 \# e_2$; then $\#[e_1] \backslash \#[e_2] = \{e_1\}$ and $\#[e_2] \backslash \#[e_1] = \{e_2\}$, so neither $e_1 \triangleright e_2$ nor $e_2 \triangleright e_1$; e_1 and e_2 cannot belong to the same facet. □

More is true:

Lemma 9 *Facets are convex, i.e. $x, y \in \psi$ and $x < y < z$ together imply $z \in \psi$.*

Proof: Lemma 5 implies $\#[x] \subseteq \#[z] \subseteq \#[y]$; by assumption, $\#[x] = \#[y]$, hence $\psi(x) = \psi(y) = \psi(z)$. □

Lemma 10 *For any condition b such that $b^\bullet \cap \psi(b) \neq \emptyset$, we have $|b^\bullet| = 1$.*

Proof: Suppose $|b^\bullet| \geqslant 1$. If $b^\bullet \cap \psi(b) \neq \emptyset$, then $\psi(b)$ contains a conflict pair, contradicting Lemma 9. So assume $e_1 \in b^\bullet \cap \psi(b)$ and $e_2 \in b^\bullet \backslash \psi(b)$. But then $e_2 \in \#[e_1] \backslash \#[(]b)$, hence $\psi(b) \neq \psi(e_1)$, contradicting the assumption. □

A consequence of Lemma 10 is that maximal nodes in a facet are conditions.

Facets are Abstractions We first observe that facets carry an induced event structure. To make this precise, let x_i be a node of ON, let $\psi_i \triangleq \psi(x_i)$, and set

$$\psi_1 \prec_\Psi \psi_2 \iff \left\{ \begin{array}{ll} & \psi_1 \neq \psi_2 \\ \exists & y_1 \in \psi_1, y_2 \in \psi_2 : \\ & y_1 < y_2 \end{array} \right. \tag{24}$$

$$\psi_1 \#_\Psi \psi_2 \iff [\exists\ y_1 \in \psi_1, y_2 \in \psi_2 :\ y_1 \# y_2] \tag{25}$$

Relation $\prec_\Psi$ from Definition (24) is a partial order by Lemma 9; $\#_\Psi$ is well-defined since $y_1 \# y_2$ implies $z_1 \# z_2$ for all z_1 from ψ_1 and z_2 from ψ_2. Facets are conflict-free by Lemma 8.

One checks easily that $\psi_1 \# \psi_2 \prec_\Psi \psi_3$ implies $\psi_1 \# \psi_3$, and finds that $\mathcal{F} = (\Psi, \prec_\Psi, \#_\Psi)$ is an event structure in the sense of Definition 2. In fact, contracting every facet ψ into single events e_ψ whose output conditions are the maximal conditions of ψ, and whose input conditions are given by the pre-conditions of the minimal events in ψ, we obtain a reduced occurrence net $ON_{/_\Psi}$, see Fig. 8; below we will see that this abstraction operation preserves and respects runs. We denote as $\lceil \psi \rceil$ the set of facets

$$\lceil \psi \rceil \triangleq \{\psi' \mid \psi' \prec_\Psi \psi\}.$$

By Lemma 9, the set union of all facets in $\lceil \psi \rceil$ spans a configuration of ON; we denote it by

$$\kappa(\psi). \tag{26}$$

Theorem 5 *$\omega_\Psi \subseteq \Psi$ is a run of $\mathcal{F} = (\Psi, \prec_\Psi, \#_\Psi)$ iff*

$$\omega_{\omega_\Psi} \triangleq \bigcup_{\psi \in \omega_\Psi} \psi \tag{27}$$

is a run of $\mathcal{E} = (E, \leqslant, \#)$.

Proof: First, assume ω_Ψ is a run of $\mathcal{F} = (\Psi, \prec_\Psi, \#_\Psi)$; then ω_{ω_Ψ} according to (27) is a configuration of $\mathcal{E} = (E, \leqslant, \#)$ by the above. Suppose ω_{ω_Ψ} is not maximal, and let $e \notin \omega_{\omega_\Psi}$ be such that $\lceil e \rceil \cup \omega_{\omega_\Psi}$ is a configuration. Then, by Lemma 4, the same is true for all $e' \in \psi(e)$, which contradicts maximality of ω_Ψ. Conversely, if ω_{ω_Ψ} is a run of $\mathcal{E} = (E, \leqslant, \#)$, assume there exists $\psi \notin \omega_\Psi$ such that $\lceil \psi \rceil \cup \omega_\Psi$ is a configuration in $\mathcal{F} = (\Psi, \prec_\Psi, \#_\Psi)$; then for any $e \in \psi$, we have that $\lceil e \rceil \cup \omega_{\omega_\Psi}$ is a configuration, contradicting the assumption. □

Theorem 6 *Let $ON = (B, E, G, \mathbf{c}_0)$ be an occurrence net, and Ψ its set of facets. Set*

$$E_{/_\Psi} \triangleq \Psi,\ B_{/_\Psi} \triangleq \mathbf{c}_0 \cup \{b \mid b^\bullet \cap \psi(b) = \emptyset\}$$

$$G_{/_\Psi} \triangleq \big\{(b, \psi) \in B_{/_\Psi} \times E_{/_\Psi} \mid b^\bullet \cap \psi(b) \neq \emptyset\big\} \cup \big\{(\psi, e) \in B_{/_\Psi} \times E_{/_\Psi} \mid b^\bullet \cap \psi(b) = \emptyset\big\};$$

Then $ON_{/_\Psi} = (B_{/_\Psi}, E_{/_\Psi}, G_{/_\Psi}, \mathbf{c}_{0/_\Psi})$ with $\mathbf{c}_{0/_\Psi} \triangleq \mathbf{c}_0$ is an occurrence net.

Proof: Note that the relation $(G_{/_\Psi})^2$ coincides with the immediate successor relation of $\mathcal{F}$. In view of the previous results and the construction of $ON_{/_\Psi}$, it therefore remains to show that

1. $ON_{/_\Psi}$ is a net, and
2. there is no backward branching;

once both are established, the induced relations on $E_{/_\Psi}$ can be easily seen to agree with those in $\mathcal{F}$, and we are done. For 1), just note disjointness and non-emptyness of $E_{/_\Psi}$ and $b_{/_\Psi}$ are immediate, and that by construction, $G_{/_\Psi} \subseteq (b_{/_\Psi} \times E_{/_\Psi}) \cup (E_{/_\Psi} \times b_{/_\Psi})$. To see 2), assume $G_{/_\Psi}$ contains two arcs $(e_{1/_\Psi}, b_{/_\Psi})$ and $(e_{2/_\Psi}, b_{/_\Psi})$ such that $e_{1/_\Psi} \neq e_{2/_\Psi}$. By that assumption, there must exist (in ON) e_1' in the facet of e_1 and e_2' in the facet of e_1 such that $b \in e_1'^\bullet \cap e_2{}^\bullet$, and moreover $e_1' \neq e_2'$ since facets are pairwise disjoint by construction; but then ON contains already a backward branching, which is impossible. □

$\mathcal{Q}$-Diagnosability With the same setting and notations, define the *pro-cone* of a node $x \in E \cup B$ as

$$\lceil\lceil x \rceil\rceil \triangleq \kappa(\psi(x)); \tag{28}$$

the *closure* of a configuration κ is defined as

$$\lceil\lceil \kappa \rceil\rceil \triangleq \bigcup_{x \in \kappa} \lceil\lceil x \rceil\rceil. \tag{29}$$

Configuration κ is *closed* iff $\lceil\lceil \kappa \rceil\rceil = \kappa$. Notice that $\lceil\lceil \kappa \rceil\rceil$ coincides with the configuration obtained by intersecting all runs that extend κ; this makes closed configurations key entities for asynchronous diagnosis. They are obtained as the configurations of the *facet* event structure $(\Psi, \prec_\Psi, \#_\Psi)$; in fact:

Lemma 11 *The configurations of $ON_{/_\Psi}$ correspond one-to-one to the closed configurations of ON.*

We are now ready to give the definition of *$\mathcal{Q}$-diagnosability*:

Definition 11 *If ON satisfies $\mathcal{WOBS}$ w.r.t. E_O, then is* $\mathcal{Q}$-diagnosable *w.r.t. ϕ iff for configurations κ, κ',*

$$[\lceil\lceil \kappa \rceil\rceil \sim_O \lceil\lceil \kappa' \rceil\rceil \ \wedge \ \lceil\lceil \kappa \rceil\rceil \equiv_M \lceil\lceil \kappa' \rceil\rceil] \ \Rightarrow \ \lceil\lceil \kappa \rceil\rceil \equiv_\Phi \lceil\lceil \kappa' \rceil\rceil. \tag{30}$$

In words, ON is $\mathcal{Q}$-diagnosable iff for any two configurations κ, κ' the following holds: if the *inevitable common parts* $\lceil\lceil \kappa \rceil\rceil / \lceil\lceil \kappa' \rceil\rceil$ of all runs that extend κ/κ', respectively, produce the same observations and the same marking, they have to be also fault equivalent. Note that this definition is less restrictive than the one from [14] since it only applies to marking equivalent pairs. We observe that $\mathcal{Q}$-diagnosability includes both diagnosis of the past as 'prediction' of concurrent or future events. This notion of diagnosis is thus well adapted to asynchronous systems where the the precise interleaving of events is not available; concurrent events will occur and go unnoticed *unless* they change future branchings.

Verification of $\mathcal{Q}$-diagnosability for ON reduces - under some simplifying assumptions - to verification of weak diagnosability for $ON_{/_\Psi}$:

Theorem 7 *Assume that ON and E_O are such that for every facet ψ of ON, $|\psi \cap E_O| \in \{0,1\}$, and that $\psi \cap E_\phi \psi \cap E_O = \emptyset$. Define $\lambda_{/_\Psi} : \Psi \to \mathfrak{A}$ by setting*

$$\lambda_{/_\Psi}(\psi) \triangleq \begin{cases} \lambda(\pi(e)) & : \quad \psi \cap E_O = \{e\} \\ \varepsilon & : \quad \psi \cap E_O = \emptyset \end{cases} .$$

Further, let $\Psi_\phi \triangleq \{\psi \in \Psi(ON) \mid E_\phi \cap \psi \neq \emptyset\}$. Then ON is $\mathcal{Q}$-diagnosable for ϕ iff $ON_{/_\Psi}$ is weakly diagnosable for ϕ.

Proof: Suppose first that ON is $\mathcal{Q}$-diagnosable for E_O and $\Phi_{/_\Psi}$, and that $ON_{/_\Psi}$ is not weakly diagnosable. Then by Theorem 1, there exist configurations $\kappa_1, \kappa_2, \kappa_1', \kappa_2'$ of $\mathcal{L}_{\mathbf{prog}}(ON_{/_\Psi})$ such that *(1)* $\kappa_1 \neq \kappa_1'$ and $\kappa_1 \sqsubseteq \kappa_1'$, $\kappa_2 \sqsubseteq \kappa_2'$, *(2)* $\kappa_1 \sim_O \kappa_2$ and $\kappa_1' \sim_O \kappa_2'$, *(3)* $\kappa_1 \equiv_M \kappa_1'$ and $\kappa_2 \equiv_M \kappa_2'$, but *(4)* κ_1' contains ϕ while κ_2' does not. But then the configurations $\kappa(\kappa_1), \kappa(\kappa_1'), \kappa(\kappa_2), \kappa(\kappa_2') \in \mathbf{Con}(ON_{/_\Psi})$ obtained by (26) and Lemma 11 constitute a counterexample to $\mathcal{Q}$-diagnosability of ON. The converse implication is obtained directly from the correspondence of configurations in $\mathbf{Con}(ON_{/_\Psi})$ and closed configurations in $\kappa(ON)$, by Lemma 11. □

Note that the assumption of only one observable event per facet is made here only to make the presentation simpler; in the general case, a more sophisticated labelling must be devised so that a generalization of Theorem 7 can hold. Efficient techniques for this would be an interesting field of future work.

Depending on the particular net under study, the facet net can be considerably smaller than the original unfolding; in some cases, it might be efficient to synthesize a generating Petri net from the quotient unfolding, and perform the diagnosis (or other analysis) on that net rather the original one. We think the tradeoff between this offline effort and the online complexity should be weighed carefully, as there is no general result for its effectiveness: some nets will allow great reductions and speedup by quotienting, while for others there is no gain at all.

8 Conclusion

We have shown how the problem of diagnosability splits into several variants in the context of true concurrency in asynchronous systems. Characterizations of weak and strong diagnosability have been given. For *verification* of both types of properties, we have shown methods based on net invariants and complete prefixes. The discriminating power of unfolding prefixes is unmatched by that of invariants, or structural methods in general. Care must, however, be taken to control the size of the prefix required in any analysis.

Investigating the relational structure further - for the purpose of finer analysis of observability and weak diagnosability - leads to the *reveals*-relation $\triangleright$ and the associated decomposition of occurrence nets into facets. We have seen that $\triangleright$ can be effectively computed on sufficiently large prefixes, and that facets are adequate abstractions for preserving maximal nonsequential behaviour. The analysis of the nets obtained by facet abstraction, and their properties in terms of diagnosis, is an interesting new field. As noted above, knowledge of facets allows for *prediction* into the future. Obviously, the prognostic capacity of diagnosis using $ON_{/_\Psi}$ depends directly on the size of ON's facets: the gain will

thus be strongest in systems with a *high* degree of concurrency and a low to moderate degree of branching.

Generally speaking, strong diagnosability is a notion inherited from sequential systems, while weak diagnosability and $\mathcal{Q}$-diagnosability are genuinely asynchonous properties with no sequential equivalent. (1) The explicit link between weak and $\mathcal{Q}$-diagnosabilities is given by Theorem 7.

It remains to optimize the exploration of the data structures of $\mathcal{U}(\mathcal{N})$ and Ψ for a most efficient verification of diagnosability. Note in particular that computing the $\triangleright$-relation is polynomial in the size of $\mathcal{U}_2(\mathcal{N})$; on the other hand, the worst case size of $\mathcal{U}_2(\mathcal{N})$ is exponential in the sized of P. However, many systems for which modeling with Petri nets is well suitable - namely highly distributed and asynchronous systems -, generally yield an order 2 unfolding of reasonable size.

References

[1] S. Abbes and A. Benveniste. True-concurrency probabilistic models: Branching cells and distributed probabilities for event structures. *Information and Computation* **204** (2) pp. 231–274, 2006.

[2] C.G. Cassandras and S. Lafortune. Introduction to Discrete Event Systems. Kluwer Academic Publishers, Boston etc, 1999.

[3] J. Desel and J. Esparza. Free Choice Petri Nets. *Cambridge Tracts in Theoretical Computer Science* vol. **40**, Cambridge University Press, 1995.

[4] V. Diekert and G. Rozenberg, eds. *The Book of Traces* . World Scientific, 1995.

[5] J. Engelfriet. *Branching Processes of Petri Nets.* Acta Informatica **28**:575–591, 1991.

[6] J. Esparza, S. Römer, and W. Vogler. An improvement of McMillan's unfolding algorithm. *Formal Methods in System Design* **20**(3):285-310, 2002.

[7] E. Fabre and A. Benveniste. Partial Order Techniques for Distributed Discrete Event Systems: why you can't avoid using them. *INRIA Research report* **5916**, Feb. 2007; http://hal.inria.fr/inria-00130025. Extended version of a plenary Address at WODES 2006.

[8] E. Fabre, A. Benveniste, C. Jard, and S. Haar. Diagnosis of Asynchronous Discrete Event Systems, a Net Unfolding Approach. *IEEE Trans. Aut. Control* **48**(5)Â·714–727, May 2003.

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

[10] Infinite traces. Chapter in [4].

[11] A. Giua and C. Seatzu. Observability of Place/Transition Nets. *IEEE Trans. Aut. Control* **47**(9):1424–1437, 2002.

[12] A. Giua and C. Xie. Control of safe ordinary Petri nets using unfolding. *Discrete Event Dynamic Systems* **15**(4):349–373, Dec. 2005.

[13] S. Haar, A. Benveniste, E. Fabre, and C. Jard. Partial Order Diagnosability of Discrete Event Systems Using Petri Net Unfoldings. In: *Proceedings of 42nd IEEE Conference on Decision and Control (CDC),* 2003.

[14] S. Haar. Unfold and Cover: Qualitative Diagnosability for Petri Nets. Proc. 46th IEEE Conference on Decision and Control, 2007.

[15] S. Haar. Diagnosability and Branching Process Semantics. In: *Object Petri Nets, Processes, and Object Calculi*. Festschrift for R. Valk. Bericht (tech. report) **265**, pp.13–34, FB Informatik, University of Hamburg.

[16] L.E. Holloway, B.H. Krogh and A. Giua. A Survey of Petri Net Methods for Controlled Discrete event systems. *Discrete Event Dynamic Systems: Theory and Applications* **7**:151–190, 1997.

[17] S. Jiang, Z. Huang, V. Chandra and R. Kumar. A Polynomial Time Algorithm for Diagnosability of Discrete Event Systems.

[18] V. Khomenko, M. Koutny, and W. Vogler. Canonical Prefixes of Petri Net Unfoldings. *Acta Informatica* **40**:95–118, 2003. Preliminary Version in: D. Brinskma and K.G. Larsen (eds.), *Proc. CAV 2002*, LNCS **2404**:582–595. , Springer Verlag 2002.

[19] R. Kumar and M.A. Shayman. Formulae relating Controllability, Observability, and Co-Observability. *Automatica* vol. 34 no.2, March 1998, pp 211–215.

[20] R. Kummetz and D. Kuske. The topology of Mazurkiewicz Traces. *Theor. Comp. Sci.* **305**:237–258, 2003.

[21] M.Z. Kwiatkowska. A Metric for Traces. *Information Processing Letters* **35**:129–135.

[22] F. Lin. Diagnosability of discrete event systems and its applications. *Discrete Event Dynamic Systems*, 4(1), 1994, pp. 197-212.

[23] K. McMillan. Using Unfoldings to avoid the state explosion problem in the verification of asynchronous circuits. *4th Workshop on Computer Aided Verification* 164–174, 1992.

[24] T. Murata. Petri Nets: Properties, Analysis and Applications. *Proc. of the IEEE* vol. **77** no 4, April 1989.

[25] M. Nielsen, G. Plotkin G. Winskel. Petri nets, event structures, and domains, Part I. *TCS* **13**:85–108, 1981.

[26] J.L. Peterson. *Petri Net Theory and the Modeling of Systems.* Prentice-Hall, 1981.

[27] W. Reisig. *Petri nets.* Springer Verlag, 1985.

[28] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *IEEE Trans. Aut. Control* 40(9), 1555-1575, 1995.

[29] G. Winskel. Event structures. *Advances in Petri nets,* LNCS **255**: 325–392, Springer Verlag, 1987.

[30] T. Yoo and S. Lafortune. Polynomial-Time Verification of Diagnosability of Partially-Observed Discrete-Event Systems. *IEEE Trans. Aut. Control* **47**(9):1491-1495 , 2002.

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399