Diagnosis and diagnosability of discrete event systems using Petri nets

Alessandro Giua

DIEE, University of Cagliari & LSIS, Aix-Marseille University

DCDS'13, York, UK, 5 September 2013

Outline

- Background and motivation
- PN state estimation with partial observation
- PN diagnosis
- PN diagnosability
- Conclusions

Outline

- **Background and motivation**
- PN state estimation with partial observation
- PN diagnosis
- PN diagnosability
- Conclusions

The state estimation problem

Definition (State estimation problem)

Reconstruct the current **state** values of a dynamical system from the knowledge of the current and past values of its external measurable outputs and inputs.

If such a problem admits a solution, the system is said to be observable.

We denote:

- w an observation
- $\mathcal{C}(w)$ the set of states **consistent** with observation w, i.e., the possible values of the system's state after w has been observed

Some issues in DES estimation

Choice of suitable **inputs**:

- input events (in I/O automata)
- no inputs in autonomous systems

Choice of suitable **outputs**:

- event labels (e.g., Mealy automaton)
- state labels (e.g., Moore automaton) or measurements (sensors on PN places)
- combination of both

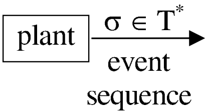
"Events as outputs" is the most popular choice.

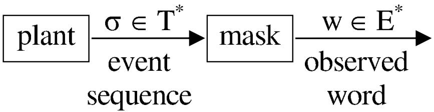
Estimate vs. enumeration:

- TDS: estimate $\chi(t)$ of the actual state $x(t)$
- DES: set of states $\mathcal{C}(w)$ consistent with the observation w

Two main approaches to the state estimation of DES

- **Total observation**: all events are observable (deterministic system) but the initial state is (partially) unknown.
- **Partial observation**: not all events are observable (nondeterministic system) but the initial state is usually known.

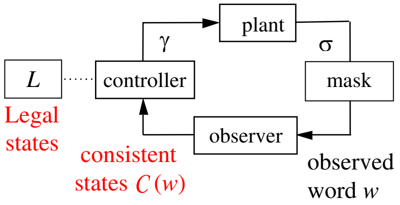


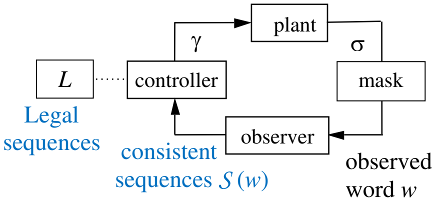


In the second case we may also be interested in reconstructing the set $\mathcal{S}(w)$ of event sequences consistent with observation w (**event estimation**).

Motivation for state estimation: supervisory control

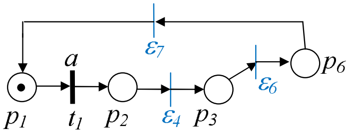
State-feedback or event-feedback control scheme



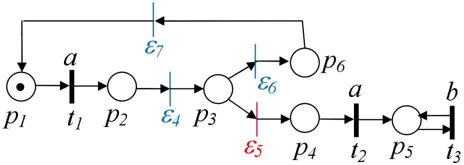


Motivation for state estimation: diagnosis

Given a nominal model and a fault model (with unobservable fault events) determine if a fault has occurred.



Nominal model



Faulty model

Other motivations for state estimation

- Monitoring the evolution of a partially observed system
- Surveillance / intrusion detection
- Testing, e.g., determine final state after a test (synchronizing and homing sequences)
- Opacity: current state is to remain ambiguous

Rest of the talk

1. A Petri net approach for **state estimation with partial observation**

2. A Petri net approach for **diagnosis** and **diagnosability**

Advantages wrt automata based approaches will be pointed out.

Outline

- Background and motivation
- **PN state estimation with partial observation**
- PN diagnosis
- PN diagnosability
- Conclusions

Estimation problem with partial observation

Setting

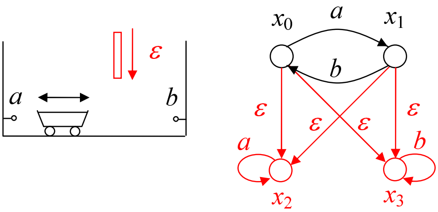
- To each transition is associated a **label** (possibly the empty string ε)
- When a transition fires its label is **observed**
- Events associated to the empty string produce no observation and are called **silent** or **unobservable**
- Events sharing the same label are called **undistinguishable**.

Here we focus on the problem of reconstructing the state consistent with a given observation.

Example

An **AGV with obstacle**.

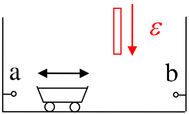
- The AGV moves from left to right and viz. automatically
- Two contacts at end points generate a signal when the AGV touches them
- An obstacle may block the path (there is no sensor to detect this)

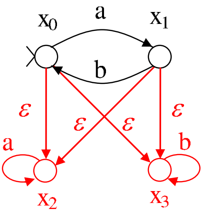


Observer design with automata

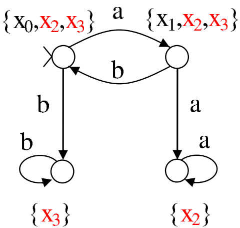
The **observer is constructed by determinization**: NFA $\rightarrow$ DFA.

- Each state of the DFA corresponds to a set of states of the NFA.
- The state reached on the DFA after the word w is observed gives the set of states of the NFA consistent with w.





NFA



DFA

The automata determinization procedure

Advantages

- Generality: works for any NFA: $\mathcal{L}_{NFA} = \mathcal{L}_{DFA} = \mathcal{L}_{regex}$.

Drawbacks

- Each set $\mathcal{C}(w)$ must be exhaustively enumerated
- To compute $\mathcal{C}(w)$ need to compute $\mathcal{C}(w')$ for all prefixes $w' \preceq w$
- If the NFA has n states, the DFA can have up to 2^n states
- Does not allow to reconstruct the set $\mathcal{S}(w)$ of consistent sequences

Can a determinization procedure be applied to Petri nets?

Unfortunately this is not possible in the general case. In fact:

$$\mathcal{L}_{\text{det}} \subsetneq \mathcal{L}_{\lambda}$$

where

- $\mathcal{L}_{\text{det}}$: set of deterministic PN languages.
- $\mathcal{L}_{\lambda}$: set of arbitrary PN languages. Nondeterminism is due both to silent events and to undistinguishable events.

Proposed approach

We propose a different technique:

- At each step the set of consistent markings is represented by the integer solutions of a linear constraint set thus one needs not exhaustively enumerate all consistent markings.
- The linear constraint set depends on some parameters (the so-called basis markings) that can be recursively computed each time a new event is observed.
- We pose some structural constraints but the same procedure works for bounded and unbounded nets.

Net structure

A **Place/Transition net** (P/T net) is a structure $N = (P, T, Pre, Post)$ where:

- P is a set of **places** represented by circles, $|P| = m$;
- T is a set of **transitions** represented by bars, $|T| = n$;
- $Pre : P \times T \to \mathbb{N}$ is the **pre-incidence function** that specifies the arcs directed from places to transitions;
- $Post : P \times T \to \mathbb{N}$ is the **post-incidence function** that specifies the arcs directed from transitions to places.

Notation

- Labeling function $L : T \to E \cup \{\varepsilon\}$ assigns to each transition $t \in T$ either a symbol from a given alphabet E or the empty string ε.

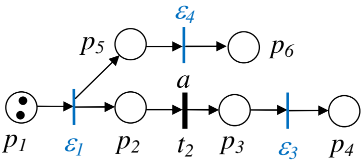
- Set of silent or unobservable transitions: $T_u = \{t \in T \mid L(t) = \varepsilon\}$.

- $\bar{T}$−induced subnet of N: the new net $\bar{N}$ obtained from N removing all transitions in $T \setminus \bar{T}$.

Assumptions:

- the structure of the net N is known;
- the initial marking M_0 is known;
- the net is labeled $\Longrightarrow$ when $\sigma \in T^*$ fires we observe $w = L(\sigma) \in E^*$;
- the T_u-induced subnet is acyclic.

Example



Unobservable transitions are in blue.

Consistent markings/sequences

Definition

Given a word w, the set of **w-consistent markings** is:

$$\mathcal{C}(w) = \{M \in \mathbb{N}^m \mid \ (\exists \sigma \in T^*) \ : M_0[\sigma\rangle M, \ L(\sigma) = w\}.$$

and the set of **w-consistent sequences** is:

$$\mathcal{S}(w) = \{\sigma \in T^* \mid \ M_0[\sigma\rangle, \ L(\sigma) = w\}.$$

Basic notions

The solution we propose is based on the following notions:

- Justifications
- Basis markings

ADVANTAGE: no need to explore all reachability set but only the smaller basis marking set.

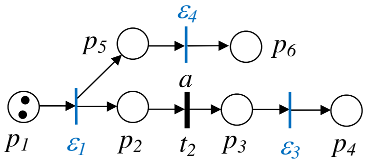
Justifications

Set of justifications of the observed word $w \in L^*$

$$\mathcal{J}(w) = \{(\sigma_o, \sigma_u), (\sigma'_o, \sigma'_u), \ldots\}$$

where in each couple

- sequence $\sigma_o \in T_o^*$ is such that $L(\sigma) = w$
- sequence $\sigma_u \in T_u^*$ (called **justification**) is a sequence of unobservable transitions that must be interleaved with σ_o to produce a firable sequence and whose firing vector $\pi(\sigma_u)$ is minimal.



If a is observed $J(a) = \{(t_2, \varepsilon_1)\}$. Note that also ε_3 and ε_4 may have fired.

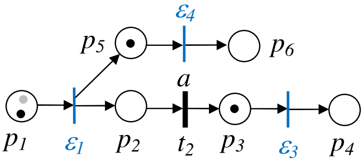
Basis markings

For each couple $(\sigma_o, \sigma_u) \in \mathcal{J}(w)$, the marking

$$M_b = M_0 + C_u \cdot \pi(\sigma_u) + C_o \cdot \pi(\sigma_o)$$

i.e., the marking reached firing σ_o interleaved with the minimal justification σ_u, is called basis marking and the firing vector $\pi(\sigma_u)$ is called its j-vector (or justification-vector).

$\mathcal{M}(w)$ is the set of pairs (basis marking - relative j-vector) that are consistent with $w \in L^*$ and $\mathcal{M}_b(w)$ is the set of basis markings that are consistent with $w \in L^*$.



If a is observed
$\mathcal{M}(a) = \{([1\ 0\ 1\ 0\ 1\ 0]^T, [1\ 0\ 0])\}$
where $j = [\varepsilon_1\ \varepsilon_3\ \varepsilon_4] = [1\ 0\ 0]$.

Computing the set of consistent markings

Theorem

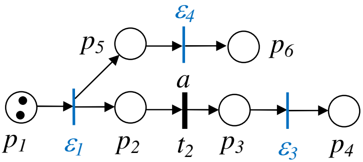
Let us consider a net system $\langle N, M_0 \rangle$ whose unobservable subnet is acyclic. For any $w \in L^*$ it holds that

$$\begin{aligned} \mathcal{C}(w) &= \bigcup_{M^b \in \mathcal{M}_b(w)} R(N_u, M^b) \\ &= \bigcup_{M^b \in \mathcal{M}_b(w)} \{M \in \mathbb{N}^m \mid (\exists y \geq \vec{0})\ M = M_b + C_u \cdot y\}. \end{aligned}$$

Recursive computation of $\mathcal{M}(w)$

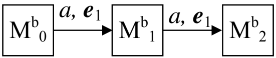
The set $\mathcal{M}(w)$ and $\mathcal{M}_b(w)$ can be recursively computed.

For bounded nets it can be done off-line computing a Basis Reachability Graph (may be nondeterministic)



$M^b_0 = [2\ 0\ 0\ 0\ 0\ 0]^T$
$M^b_1 = [1\ 0\ 1\ 0\ 1\ 0]^T$
$M^b_0 = [0\ 0\ 2\ 0\ 2\ 0]^T$

$\boldsymbol{e}_1 = \pi(\varepsilon_1) = [1\ 0\ 0]$



$$\mathcal{M}(\varepsilon) = \{(M_0^b, 0)\} \quad \mathcal{M}(a) = \{(M_1^b, e_1)\} \quad \mathcal{M}(aa) = \{(M_2^b, e_1 + e_1)\}.$$

Summary

- **Set of consistent markings** needs not be enumerated but is **described by a constraint set** in terms of the basis marking and unobservable subnet reachability.
- The set of basis marking can be easily **recursively computed**.
- In the **worst case** the set of basis markings is equal to the reachability set.
- There are nets where the size of the reachability graph is exponential in some net parameters, while the set of basis marking is constant or grows linearly.

Outline

- Background and motivation
- PN state estimation with total observation
- **PN diagnosis**
- PN diagnosability
- Conclusions

Main idea

We want to use the previous framework of estimation with partial observation to solve a diagnosis problem.

The set of unobservable transitions is partitioned: $T_u = T_f \cup T_{reg}$.

- T_f: set of **fault** transitions
- T_{reg}: set of **regular** transitions (unobservable but not fault)

The set of fault transitions can be partitioned into fault classes

$$T_f = T_f^1 \cup T_f^2 \cup \ldots \cup T_f^r$$

Two problems:

- **Diagnosis**: given an observation w determine if the i-th fault has occurred, i.e., if a transition in T_f^i has fired.
- **Diagnosability**: determine if a given fault can be diagnosed in a fixed number of steps.

Diagnoser

A *diagnoser* is a function $\Delta : L^* \times \{T_f^1, T_f^2, \ldots, T_f^r\} \rightarrow \{0, 1, 2, 3\}$:

$\Delta(w, T_f^i) = 0$ NO FAULT

$\Rightarrow$ The ith fault cannot have occurred because none of the firing sequences consistent with the observation contains transitions in T_f^i.

$\Delta(w, T_f^i) = 1$ POSSIBLE FAULT

$\Rightarrow$ The ith fault may have occurred but never while firing a justification of w.

$\Delta(w, T_f^i) = 2$ POSSIBLE FAULT

$\Rightarrow$ Some (but not all) justification of W contains some transition in T_f^i.

$\Delta(w, T_f^i) = 3$ FAULT DETECTED

$\Rightarrow$ The ith fault has occurred because each justification of w contains at least one transition in T_f^i.

Characterization of diagnosis states

Proposition: Consider an observed word $w \in L^*$.

$\Delta(w, T_f^i) \in \{0,1\}$ iff $\forall\ (M^b, j) \in \mathcal{M}(w)$ and $\forall t_f \in T_f^i$ it holds $j(t_f) = 0$.

$\Delta(w, T_f^i) = 2$ iff $\exists\ (M^b, j) \in \mathcal{M}(w)$ and $(M^{b'}, j') \in \mathcal{M}(w)$ such that:
(i) $\exists t_f \in T_f^i$ such that $j(t_f) > 0$,
(ii) $\forall t_f \in T_f^i$, $j'(t_f) = 0$.

$\Delta(w, T_f^i) = 3$ iff $\forall\ (M^b, j) \in \mathcal{M}(w)\ \exists t_f \in T_f^i$ such that $j(t_f) > 0$.

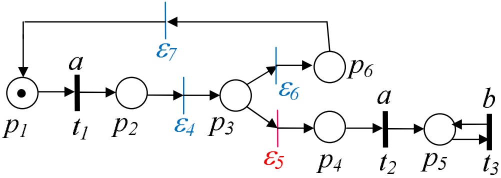
Characterization of diagnosis states

Proposition: For a Petri net whose unobservable subnet is acyclic, let $w \in L^*$ be an observed word : $\forall\ (M^b, j) \in \mathcal{M}(w)$ it holds $j(t_f) = 0$.
Let us consider the constraint set

$$\mathcal{T}(M^b) \quad = \quad \begin{cases} M^b + C_u \cdot z \geq \vec{0}, \\ \sum\limits_{t_f \in T_f^i} z(t_f) > 0, \\ z \in \mathbb{N}^{n_u}. \end{cases}$$

- $\Delta(w, T_f^i) = 0$ if $\forall\ (M^b, j) \in \mathcal{M}(w)$ the constraint set has no admissible solution.
- $\Delta(w, T_f^i) = 1$ if $\exists\ (M^b, j) \in \mathcal{M}(w)$ such that the constraint set has a solution.

Example



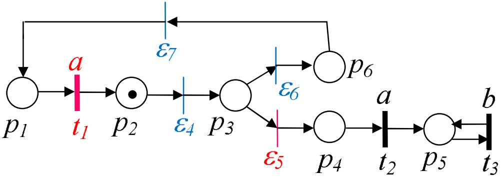
$j = [\ j(\varepsilon_4)\ \ j(\varepsilon_5)\ \ j(\varepsilon_6)\ \ j(\varepsilon_7)\ ]$

$M_0 = [1\ 0\ 0\ 0\ 0\ 0]^T$, $w = \varepsilon$

$\mathcal{J}(w) = \{(\varepsilon, \varepsilon)\}$

$M_0^b = [1\ 0\ 0\ 0\ 0\ 0]^T \quad j = [\ 0\ \ 0\ \ 0\ \ 0\ ] \qquad \Longrightarrow \Delta(T_f, \varepsilon) = 0$

Example



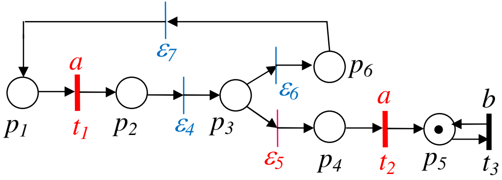
$j = [\ j(\varepsilon_4) \quad j(\varepsilon_5) \quad j(\varepsilon_6) \quad j(\varepsilon_7)\ ]$

$M_0^b = [1\ 0\ 0\ 0\ 0\ 0]^T$, $w = a$

$\mathcal{J}(w) = \{(t_1, \varepsilon)\}$

$M_1^b = [0\ 1\ 0\ 0\ 0\ 0]^T \qquad j = [\ 0 \quad 0 \quad 0 \quad 0\ ] \qquad \Longrightarrow \Delta(T_f, a) = 1$

Example



$j = [\ j(\varepsilon_4)\quad j(\varepsilon_5)\quad j(\varepsilon_6)\quad j(\varepsilon_7)\ ]$

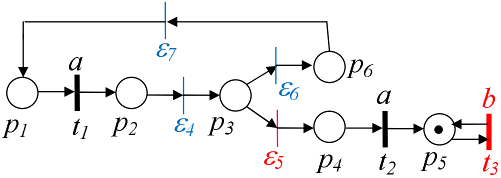
$M_1^b = [0\ 1\ 0\ 0\ 0\ 0]^T,\ w = aa$

$\mathcal{J}(w) = \{(t_1 t_1, \varepsilon_4\varepsilon_6\varepsilon_7), (t_1 t_2, \varepsilon_4\varepsilon_5)\}$

$M_1^b = [0\ 1\ 0\ 0\ 0\ 0]^T \quad j = [\ 1\quad 0\quad 1\quad 1\ ]$
$M_2^b = [0\ 0\ 0\ 0\ 0\ 1]^T \quad j = [\ 1\quad 1\quad 0\quad 0\ ]$

$\Longrightarrow \Delta(T_f, ab) = 2$

Example



$j = [\ j(\varepsilon_4)\quad j(\varepsilon_5)\quad j(\varepsilon_6)\quad j(\varepsilon_7)\ ]$

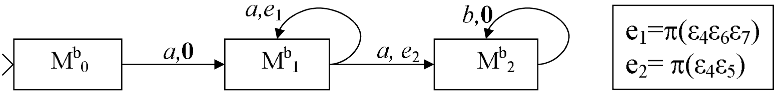
$M_2^b = [0\ 0\ 0\ 0\ 0\ 1]^T$, $w = aab$

$\mathcal{J}(w) = \{(t_1 t_2 t_3, \varepsilon_4 \varepsilon_5)\}$

$M_2^b = [0\ 0\ 0\ 0\ 0\ 1]^T \quad j = [\ 1\quad 1\quad 0\quad 0\ ] \qquad \Longrightarrow \Delta(T_f, aab) = 3$

Bounded net systems

BOUNDED NET SYSTEMS $\Longrightarrow$ BASIS REACHABILITY GRAPH



Basis Reachability Diagnoser

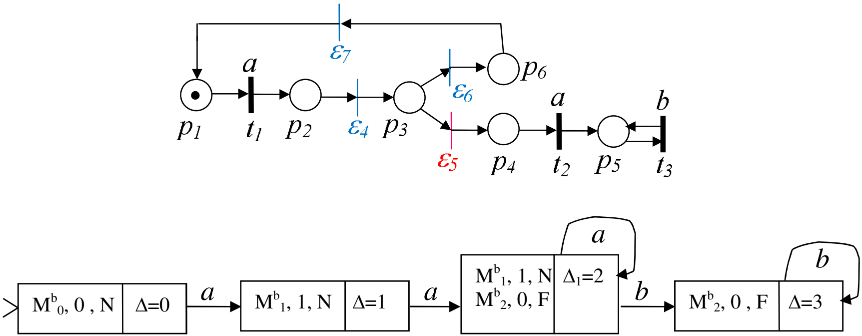
Definition: We define BRD as a deterministic graph where each node is represented by:

- one or more triple (M^b, x, h), where M is a reachable basis marking, $x \in \{0,1\}^{|T_f|}$ is a row vector in which each entry assumes value equal to 0 or 1 if $\mathcal{C}(M^b)$ is feasible or not, respectively, and $h \in \{N, F\}^{|T_f|}$ is a row vector in which each entry is equal to N if reaching M from M_0 the fault has not occurred and equal to F otherwise;
- one tag Δ_i that represents the diagnosis state of the node with respect to the fault class i.

and each arc is labeled with a label $l \in L$.

Example

The BRD can easily be built starting from BRG.



Final comments

- The **state estimation** approach previously proposed can be naturally used as a building block for diagnosis
- Just a **partial enumeration** of the state space (basis markings) is necessary
- The technique can be used **on-line** for bounded or unbounded nets constructing the consistent set of basis markings on the fly.
- If the set of bounded markings is finite (this holds for bounded nets) it may be convenient to construct **off-line** the basis reachability diagnoser.

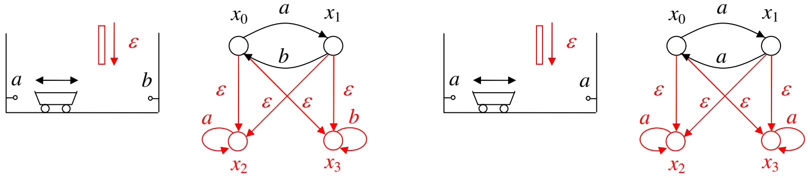
Outline

- Background and motivation
- PN state estimation with partial observation
- PN diagnosis
- **PN diagnosability**
- Conclusions

Problem Statement for Diagnosability

A Petri net system $\langle N, M_0 \rangle$ is diagnosable if when any failure transition occurs, its failure type is detected after the firing of a finite number of transitions from its occurrence.

Example: A diagnosable system (left) and a non diagnosable one (right).



AIM: Given a net system $\langle N, M_0 \rangle$ we want to determine if the system is diagnosable or not.

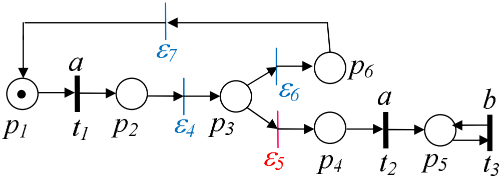
Modified Basis Reachability Graph

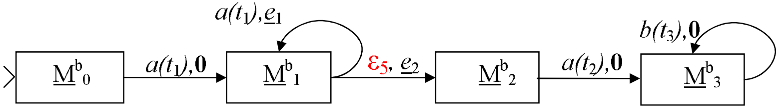
To deal with diagnosability of bounded nets the BRG does not contain enough information.

If we are interested in diagnosability we need to construct a Modified Basis Reachability Graph (MBRG) where fault transitions are treated as observable transitions.

Example

$T_o = \{t_1, t_2, t_3\}$ $\quad$ $T_u = \{\varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7\}$ $\quad$ $T_f = \{\varepsilon_5\}$





$\underline{e}_1 = \pi(\varepsilon_4\varepsilon_6\varepsilon_7)$ $\quad$ $\underline{e}_2 = \pi(\varepsilon_4)$

Features of the modified BRG

- The arcs are labeled either with observable transitions or with fault transitions.
- $|MBRG| \geq |BRG|$.
- We have presented a technique to determine the diagnosability of a PN based on the analysis of the cycles of its **modified basis reachability diagnoser** (MBRD), i.e., the diagnoser obtained by the MBRG.

An interesting general result for bounded systems

Jiroveanu and Boel[1] have proved in a slightly different context a result that also applies to our case.

Theorem

A Petri net is diagnosable if and only if its MBRG is a diagnosable automaton.

Thus one just needs to construct the MBRG and may use automata based approaches to test diagnosability.

[1] G. Jiroveanu, R.K. Boel, "The Diagnosability of Petri Net Models Using Minimal Explanations," IEEE Trans. on Automatic Control, 2010.

Diagnosability of unbounded nets

We have shown that **testing diagnosability of a Petri net is a decidable problem** even if the net is unbounded[2].

The method used to check decidability does not use efficient techniques such as basis markings: it constructs a **verifier net** whose reachability space (quadratic w.r.t. the system's reachability space) must be enumerated.

In the case of unbounded Petri nets in addition to the classical notion of **diagnosability** it is also possible to define the stronger notion of **diagnosability in** k **steps**: both properties are decidable.

[2] M.P. Cabasino, A. Giua, S. Lafortune, C. Seatzu, "A new approach for diagnosability analysis of Petri nets using verifier nets," *IEEE Trans. on Automatic Control*, 2012.

Summary

- The techniques for state estimation with partial observation can be easily extended to solve a problem of diagnosis and of diagnosability.
- We need to extend the set of basis markings considering markings reached by firing a fault transition.
- For bounded nets we have presented an approach for testing diagnosability where the extended set of basis markings needs to be explored instead of the complete reachability set.
- For unbounded nets diagnosability is also decidable but no efficient approach is known.

Outline

- Background and motivation
- PN state estimation with partial observation
- PN diagnosis
- PN diagnosisability
- **Conclusions**

Conclusions

- The notion of **state estimation** and **observer** for DES's is meaningful and has practical motivations
- **Petri nets are a good model for DES** offering several computational advantages wrt automata
- A Petri net approach based on **state estimation under partial observation** founded on the notion of basis marking has been discussed.
- The basis marking approach can be naturally extended to **fault diagnosis**

Relevant literature

PN state estimation with partial observation

- A. Giua, D. Corona, C. Seatzu, "State estimation of λ-free labeled Petri nets with contact-free nondeterministic transitions," *Discrete Event Dynamic Systems*, 15(1), 2005.
- D. Corona, A. Giua, C. Seatzu, "Marking estimation of Petri nets with silent transitions," *IEEE Trans. on Automatic Control*, 52(9), 2007.

PN diagnosis

- M.P. Cabasino, A. Giua, C. Seatzu, "Fault detection for discrete event systems using Petri nets with unobservable transitions," *Automatica*, 46(9), 2010.
- M.P. Cabasino, A. Giua, M. Pocci, C. Seatzu, "Discrete event diagnosis using labeled Petri nets. An application to manufacturing systems," *Control Engineering Practice*, 19(9), 2011.
- M.P. Cabasino, A. Giua, S. Lafortune, C. Seatzu, "A new approach for diagnosability analysis of Petri nets using verifier nets," *IEEE Trans. on Automatic Control*, 2012.
- M.P. Cabasino, A. Giua, C. Seatzu, "Diagnosis using labeled Petri nets with silent or undistinguishable fault events," *IEEE Trans. on Systems Man & Cyb. – A*, 2012.