Diagnosability and fault detection in PNs - Part II: algebraic approaches for bounded systems

From observability to privacy and security in discrete event systems

Prof. Gianmaria De Tommasi
Email: detommas@unina.it

December 2020

DIETI. UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELLE TECNOLOGIE DELL'INFORMAZIONE

Course syllabus

1. Discrete Event Systems (DES), Languages and Automata
2. Petri nets (PNs) and their twofold representation to model DES
3. MILP and ILP formulations: logical conditions, binary variables “do everything", and variable connecting
4. Adding uncertainty: unobservable events and observers for finite state automata and PNs
5. Augmenting the observers: diagnosability of prefix-closed languages, diagnosers and the fault detection for finite state automata
6. Diagnosability and fault detection in PNs - Part I: graph-based approaches
7. **Diagnosability and fault detection in PNs - Part II: algebraic approaches for bounded systems**
8. Security issues in DES: non-interference and opacity
9. Non-interference and opacity enforcement
10. Open issues

1. Diagnosability and K–diagnosability

2. Algebraic characterization of the K–diagnosability problem
   - The unlabeled case
   - The labeled case
   - A benchmark

3. A fault detection approach based on solution of ILP problems

Graph-based vs algebraic approaches for diagnosability

- The notion of diagnosability in DES has been originally introduced by Sampath et al. in the 90s

  M. Sampath et al.
  Diagnosability of Discrete-Event Systems
  IEEE Transactions on Automatic Control, 1995

- The first (and most common) approaches exploit the concept of an *augmented observer* (e.g., *diagnoser*, *basis reachability graph*) (see Prof. Basile's lectures)

  M. P. Cabasino et al.
  Diagnosis Using Labeled Petri Nets With Silent or Undistinguishable Fault Events
  IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2012

- The algebraic representation of Petri net systems enables the development of techniques based on the solution of optimization problems, typically **ILP problems**
- These algebraic approaches permit to avoid to explicitly estimate of the state space → **there is no need to build a diagnoser**

- Given the computational complexity of ILP problems, **the diagnosability conditions provided by *optimization-based* algorithms require the solution of NP-hard problems**
- **ILP programming is a standard optimization tool**
  - **it is possible to rely on efficient off-the-shelf optimization software tools**
  - CPLEX®
  - FICO™ Xpress
- **Despite their computational complexity, the *optimization-based* approaches can be practically more convenient when compared with the *graph-based* ones, which usually require *ad hoc* algorithms**

Labeled and unlabeled case

- We will mainly tackle the unlabeled case
- The uncertainty is due *only* to the presence of unobservable transitions, among which there are the fault transitions
- We will consider net systems with $T = T_{uo} \cup T_o$, $T_{uo} \cap T_o = \emptyset$, and $T_f \subseteq T_{uo}$
- The algebraic approaches can be extended to the labeled case
- In the labeled case a further source of uncertainty is added $\rightarrow$ multiple observable transitions map on the same event

Usefull notation

- Given a **firing count vector** $\boldsymbol{\sigma} \in \mathbb{N}^n$, we would like to consider only the firings of either observable or unobservable transitions
- The following notation is introduced:

$$\boldsymbol{\sigma}_{|T_o} \in \mathbb{N}^n\,, \text{ with } \boldsymbol{\sigma}_{|T_o}(t) = \begin{cases} \boldsymbol{\sigma}(t) & \text{if } t \in T_o \\ 0 & \text{if } t \notin T_o \end{cases}$$

$$\boldsymbol{\sigma}_{|T_{uo}} \in \mathbb{N}^n\,, \text{ with } \boldsymbol{\sigma}_{|T_{uo}}(t) = \begin{cases} \boldsymbol{\sigma}(t) & \text{if } t \in T_{uo} \\ 0 & \text{if } t \notin T_{uo} \end{cases}$$

Main assumptions

Liveness

The net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ does not enter a deadlock after firing any fault transition

- This assumption assures that after a fault occurrence the system does not enter a deadlock
- If this would be the case, the fault detection can be detected by means of a **watchdog timer**

Boundedness

A net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ is *bounded*

- The reachability set is **finite**
- This does not represent a limitation for application in the field of automation systems

Post-language of a sequence σ

Let L be the *live* and *prefix-closed* language generated by a system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$. The **post-language** of L after the occurrence of a sequence σ is

$$L/\sigma = \left\{ v \in T^* \text{ s.t. } \sigma v \in L \right\}$$

A sequence $v \in L/\sigma$ is called *continuation* of σ

Projection and inverse projection

- $Pr : T^* \mapsto T_o^*$ is the natural projection which *erases* the unobservable transitions in a sequence σ
- Inverse projection (extended to the language L)

$$Pr_L^{-1}(r) = \{ \sigma \in L \text{ s.t. } Pr(\sigma) = r \}$$

- A fault transition $t_f \in T_f$ is said to be **diagnosable** if

$$\exists\ h \in \mathbb{N}\ \textit{such that}$$

$$\forall\ \sigma = ut_f \textit{ with } t_f \notin u, \textit{and } \forall\ v \in L/\sigma \textit{ with } |v| \geq h\,,$$

it is

$$r \in Pr_L^{-1}\big(Pr(\sigma v)\big) \Rightarrow t_f \in r\,.$$

- If we consider **any** sequence $\sigma = ut_f$ that *ends* with a failure t_f, and if v is any **sufficiently long** continuation of σ, then diagnosability of t_f implies that along every continuation v of σ it is possible to detect the occurrence of the fault with a **finite delay**

K-diagnosable fault

- Given a fault transition $t_f \in T_f$ and a **positive integer K**, t_f is said to be **K–diagnosable** if

$$\forall\ \sigma = ut_f \ \textit{with}\ t_f \notin u \ \textit{and}\ \forall\ v \in L/\sigma \ \textit{such that}\ |v| \geq K\,,$$

it is

$$r \in Pr_L^{-1}\left(Pr(\sigma v)\right) \Rightarrow t_f \in r\,.$$

- If we consider **any** sequence $\sigma = ut_f$ that *ends* with a failure t_f, then K–diagnosability of t_f implies that it is possible to detect its occurrence within a **finite delay equal to K for all the continuations v of σ**
- K–diagnosability specifies an upper bound for the number of events that are needed to detect a fault
- For a given K, K–diagnosability of a fault always implies its diagnosability, while the converse is not necessarily true
- By definition, it follows that if a fault transition is diagnosable then it exists an integer $\bar{K}$ such that it also $\bar{K}$–diagnosable.

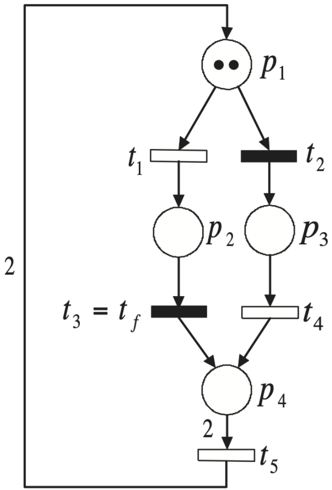
Example

- $t_3 \in T_f$
- $\sigma = t_1 t_3$ is a sequence that ends with t_3
- t_3 is not 2-diagnosable
  - $v = t_2 t_4$ belongs to L/σ with

$$Pr\left(t_1 t_3 t_2 t_4\right) = t_1 t_4\,,$$

and $t_1 t_2 t_4 \in Pr_L^{-1}\left(Pr\left(\sigma v\right)\right)$,
with $t_3 \notin t_1 t_2 t_4$

- Exploiting similar arguments and by exhaustively searching for all possibilities, it follows that t_3 is 3-diagnosable



Linear characterization of enabled sequences

- The fulfilling of the state equation

$$\boldsymbol{m} = \boldsymbol{m}_0 + \boldsymbol{C} \cdot \boldsymbol{\sigma}$$

  is only necessary to determine if $\boldsymbol{m}$ is reachable from $\boldsymbol{m}_0$ after the firing of a sequence σ s.t. $\boldsymbol{\sigma} = \pi(\boldsymbol{\sigma})$, i.e., to check if $\boldsymbol{m} \in R(N, \boldsymbol{m}_0)$
- The fact that $\boldsymbol{\sigma} = \pi(\sigma)$ satisfies the state equation gives only a necessary condition to establish if σ is an enabled sequence

Necessary and sufficient condition to check if $\boldsymbol{m}_0\left[\sigma\right\rangle$

There exists a set of ρ integer vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_\rho$ with $\rho \leq |\sigma|$ such that the following linear constraints are fulfilled

$$\begin{cases} \boldsymbol{m} \geq \mathbf{Pre} \cdot \boldsymbol{s}_1 \\ \boldsymbol{m} + \boldsymbol{C} \cdot \boldsymbol{s}_1 \geq \mathbf{Pre} \cdot \boldsymbol{s}_2 \\ \ldots \\ \boldsymbol{m} + \boldsymbol{C} \cdot \sum_{i=1}^{\rho-1} \boldsymbol{s}_i \geq \mathbf{Pre} \cdot \boldsymbol{s}_\rho \\ \sum_{i=1}^{\rho} \boldsymbol{s}_i = \pi(\sigma) \end{cases}$$

iff there exists at least one sequence σ, which is enabled under the marking $\boldsymbol{m}$ and such that $\pi(\sigma) = \boldsymbol{\sigma}$

F. Garcia Vallés
Contributions to the structural and symbolic analysis of place/transition nets with applications to flexible manufacturing systems ans asynchronous circuits
Ph.D. dissertation, Universidad de Zaragoza, 1999

Reachable markings by means faulty sequences prefixes

- In order to check either the diagnosability or the K–diagnosability of the fault transition t_f, we first need to characterize all markings reachable the prefixes of *faulty sequences*, i.e. marking at which t_f is enabled

$$\mathcal{M}(t_f) = \left\{ \boldsymbol{m} \in \mathbb{N}^m \mid \left[\boldsymbol{m}_0 \left[u \right\rangle \boldsymbol{m} \right] \bigwedge \left[t_f \notin u \right] \bigwedge \left[\boldsymbol{m} \left[t_f \right\rangle \right] \right\} .$$

- Given a marking $\boldsymbol{m} \in \mathcal{M}(t_f)$, in order to check K–diagnosability of t_f we need to characterize all the possible suffixes v of ut_f whose length is $|v| \geq K$

$$\mathcal{S}(t_f\,, K) = \left\{ \sigma \in T^* \mid \left[\sigma = ut_f v \right] \bigwedge \left[\boldsymbol{m}_0 \left[\sigma \right\rangle \right] \right.$$
$$\left. \bigwedge \left[\boldsymbol{m}_0 \left[u \right\rangle \boldsymbol{m} \right] \bigwedge \left[\boldsymbol{m} \in \mathcal{M}(t_f) \right] \bigwedge \left[|v| \geq K \right] \right\} .$$

- Once we have characterized the set $\mathcal{S}(t_f\,, K)$, if and only if t_f belongs to all the **unobservable explanations** of sequences in $\mathcal{S}(t_f\,, K)$, then t_f is K–diagnosable

Consider a net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ a transition $t \in T$ and a positive integer K. There exists at least one sequence $\sigma = utv$, with $\boldsymbol{v} = \pi(v)$, such that

$$\boldsymbol{m}_0 \left[\sigma\right\rangle \tag{1a}$$

$$t \notin u \tag{1b}$$

$$\|\boldsymbol{v}\|_1 \geq K \tag{1c}$$

if and only if there exist an integer J and $J + K$ vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_J, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K \in \mathbb{N}^n$ that fulfill the following constraints – denoted by $\mathcal{F}(\boldsymbol{m}_0, t, J, K)$

$$\left\{\begin{array}{ll}
\boldsymbol{m}_0 \geq \mathbf{Pre} \cdot \boldsymbol{u}_1 & \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \boldsymbol{u}_1 \geq \mathbf{Pre} \cdot \boldsymbol{u}_2 & \\
\ldots & \text{(2a)} \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \sum_{i=1}^{J-1} \boldsymbol{u}_i \geq \mathbf{Pre} \cdot \boldsymbol{u}_J & \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \sum_{i=1}^{J} \boldsymbol{u}_i \geq \mathbf{Pre}(\cdot, t) & \text{(2b)} \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \sum_{i=1}^{J} \boldsymbol{u}_i + \boldsymbol{C}(\cdot, t) \geq \mathbf{Pre} \cdot \boldsymbol{v}_1 & \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \sum_{i=1}^{J} \boldsymbol{u}_i + \boldsymbol{C}(\cdot, t) + \boldsymbol{C} \cdot \boldsymbol{v}_1 \geq \mathbf{Pre} \cdot \boldsymbol{v}_2 & \\
\ldots & \text{(2c)} \\
\boldsymbol{m}_0 + \boldsymbol{C} \cdot \sum_{i=1}^{J} \boldsymbol{u}_i + \boldsymbol{C}(\cdot, t) + \boldsymbol{C} \cdot \sum_{j=1}^{K-1} \boldsymbol{v}_j \geq \mathbf{Pre} \cdot \boldsymbol{v}_K & \\
\sum_{i=1}^{J} \boldsymbol{u}(t) = 0 & \text{(2d)} \\
\left\| \sum_{j=1}^{K} \boldsymbol{v}_j \right\|_1 \geq K & \text{(2e)}
\end{array}\right.$$

- The constraints $\mathcal{F}\left(\boldsymbol{m}_0\,, t_f\,, J\,, K\right)$ depend on the integer J that implicitly defines the maximum length of the sequence u
- For a given integer J, there may exists at least one marking $\widetilde{\boldsymbol{m}} \in \mathcal{M}(t_f)$ that does not satisfy (2), because is reached by a sequence that is **too long**
- $\widetilde{\boldsymbol{m}}$ could enable t_f, and starting from $\widetilde{\boldsymbol{m}}$ the fault may be undiagnosable in K steps!
- Therefore It is important to estimate the minimum value J_{min} that permits to fully describe the set $\mathcal{M}(t_f)$
- For unbounded net systems J_{min} could not exist $\rightarrow$ boundedness assumption
- In general the computation of J_{min} is not an easy task even in the case of bounded net systems.
- In the worst case an overestimation of J_{min} is given by $\text{card}\Big(R\big(N\,, \boldsymbol{m}_0\big)\Big) - 1$
- An estimate of J_{min} can be done exploiting the T-invariants

Consider a net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ and a sequence σ enabled under the initial marking $\boldsymbol{m}_0$. The sequence σ is such that

$$\pi\Big(Pr(\sigma)\Big) = \boldsymbol{b}\,,$$

if and only if there exist 2ρ vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_\rho$, $\epsilon_1, \ldots, \epsilon_\rho$, with $\rho \leq |\sigma|$, that fulfill the following set of constraints – denoted by $\mathcal{E}\big(\boldsymbol{m}_0\,, \boldsymbol{b}\big)$

$$\left\{\begin{array}{ll}
\boldsymbol{m}_0 + \boldsymbol{C}\cdot\epsilon_{1|T_{uo}} \geq \mathbf{Pre}\cdot\boldsymbol{s}_{1|T_o} & \\
\boldsymbol{m}_0 + \boldsymbol{C}\cdot\sum_{i=1}^{2}\epsilon_{i|T_{uo}} + \boldsymbol{C}\cdot\boldsymbol{s}_{1|T_o} \geq \mathbf{Pre}\cdot\boldsymbol{s}_{2|T_o} & \\
\ldots & (3a)\\
\boldsymbol{m}_0 + \boldsymbol{C}\cdot\sum_{i=1}^{\rho}\epsilon_{i|T_{uo}} + \boldsymbol{C}\cdot\sum_{j=1}^{\rho-1}\boldsymbol{s}_{j|T_o} \geq \mathbf{Pre}\cdot\boldsymbol{s}_{\rho|T_o} & \\
\boldsymbol{m}_0 \geq \mathbf{Pre}\cdot\epsilon_{1|T_{uo}} & \\
\boldsymbol{m}_0 + \boldsymbol{C}\cdot\left(\epsilon_{1|T_{uo}} + \boldsymbol{s}_{1|T_o}\right) \geq \mathbf{Pre}\cdot\epsilon_{2|T_{uo}} & \\
\ldots & (3b)\\
\boldsymbol{m}_0 + \boldsymbol{C}\cdot\sum_{i=1}^{\rho-1}\left(\epsilon_{i|T_{uo}} + \boldsymbol{s}_{i|T_o}\right) \geq \mathbf{Pre}\cdot\epsilon_{\rho|T_{uo}} & \\
\sum_{i=1}^{\rho}\boldsymbol{s}_{i|T_o} = \boldsymbol{b} & (3c)
\end{array}\right.$$

Given a net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ (not necessarily bounded) a fault transition t_f, and a positive integer K, if there exist at least one $J \in \mathbb{N}\,, J > 0$ and $3(J+K)$ vectors $\boldsymbol{u}_1\,, \ldots\,, \boldsymbol{u}_J\,, \boldsymbol{v}_1\,, \ldots\,, \boldsymbol{v}_K\,,$ $\boldsymbol{\epsilon}_1\,, \ldots\,, \boldsymbol{\epsilon}_{J+K}\,, \boldsymbol{s}_1\,, \ldots\,, \boldsymbol{s}_{J+K} \in \mathbb{N}^n$ such that

$$\min_{\text{s.t. } \mathcal{D}(\boldsymbol{m}_0\,,t_f\,,J\,,K)} \sum_{r=1}^{J+K} \epsilon_r(t_f) = 0\,,$$

where the set of constraints $\mathcal{D}\big(\boldsymbol{m}_0\,, t_f\,, J\,, K\big)$ is equal to

$$\mathcal{D}\left(\boldsymbol{m}_0\,, t_f\,, J\,, K\right) : \begin{cases} \mathcal{F}\left(\boldsymbol{m}_0\,, t_f\,, J\,, K\right) & \text{(4a)} \\ \mathcal{E}\left(\boldsymbol{m}_0\,, \sum_{i=1}^{J} \boldsymbol{u}_{i_{|T_o}} + \sum_{j=1}^{K} \boldsymbol{v}_{j_{|T_o}}\right) & \text{(4b)} \\ \boldsymbol{s}_{1_{|T_o}} = \boldsymbol{u}_{1_{|T_o}} & \\ \ldots & \\ \boldsymbol{s}_{J_{|T_o}} = \boldsymbol{u}_{J_{|T_o}} & \\ \boldsymbol{s}_{J+1_{|T_o}} = \boldsymbol{v}_{1_{|T_o}} & \text{(4c)} \\ \ldots & \\ \boldsymbol{s}_{J+K_{|T_o}} = \boldsymbol{v}_{K_{|T_o}} & \end{cases}$$

then t_f is K–undiagnosable.

K–diagnosability

Consider a bounded net system $\mathcal{S} = \langle N, \boldsymbol{m}_0 \rangle$ and a fault transition t_f, let J be a positive integer such that $J \geq J_{\min}$. Given a positive integer K, t_f is K–diagnosable **if and only if** there exist $3(J + K)$ vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_J, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$, $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{J+K}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{J+K} \in \mathbb{N}^n$ such that

$$\min_{\text{s.t. } \mathcal{D}\left(\boldsymbol{m}_0, t_f, J, K\right)} \sum_{r=1}^{J+K} \epsilon_r(t_f) \neq 0 \,,$$

Labeled systems

- $\mathcal{S}_L = \langle N, \boldsymbol{m}_0, \ell \rangle$ is a *labeled* Petri net (LPN) system
- $\ell : T \mapsto E \cup \{\varepsilon\}$ is the *labeling function*
  - $\ell(\cdot)$ assigns to each transition $t \in T$ either an event in E or the *silent event* ε
  - $\ell(t) = \varepsilon$ if $t \in T_{uo}$, while $\ell(t) \neq \varepsilon$ otherwise
- We denote with

$$T^\alpha = \left\{ t \in T \mid \ell(t) = \alpha \right\},$$

  the set of transitions associated with the same event $\alpha \in E$.
- w denotes a word of events associated with a sequence σ such that $w = \ell(\sigma)$
- $|w|$ denotes the length of w, while $|w|_\alpha$ denotes the number of occurrences of the event α in w

$\mathcal{K}$-diagnosability for labeled systems

Given a positive integer K, t_f is K–diagnosable **if and only if** there exist $3 \cdot (J+K)$ vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_J, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{J+K}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{J+K} \in \mathbb{N}^n$ such that

$$\min_{\text{s.t. } \mathcal{LD}(\boldsymbol{m}_0, t_f, J, K)} \sum_{r=1}^{J+K} \epsilon_r(t_f) \neq 0,$$

where the set $\mathcal{LD}(\boldsymbol{m}_0, t_f, J, K)$ is equal to

$$\begin{cases}
\mathcal{F}(\boldsymbol{m}_0, t_f, J, K) & \text{(5a)}\\
\mathcal{LE}\Big(\boldsymbol{m}_0, \sum_{t_l \in T^{\alpha_1}} \Big(\sum_{i=1}^{J} \boldsymbol{u}_i(t_l) + \sum_{j=1}^{K} \boldsymbol{v}_j(t_l)\Big), \ldots, \sum_{t_l \in T^{\alpha_e}} \Big(\sum_{i=1}^{J} \boldsymbol{u}_i(t_l) + \sum_{j=1}^{K} \boldsymbol{v}_j(t_l)\Big)\Big) & \text{(5b)}\\
\sum_{t_j \in T^{\alpha_l}} \boldsymbol{s}_1(t_j) = \sum_{t_j \in T^{\alpha_l}} \boldsymbol{u}_1(t_j), \quad l = 1, \ldots, e & \\
\ldots & \\
\sum_{t_j \in T^{\alpha_l}} \boldsymbol{s}_J(t_j) = \sum_{t_j \in T^{\alpha_l}} \boldsymbol{u}_J(t_j), \quad l = 1, \ldots, e & \\
\sum_{t_j \in T^{\alpha_l}} \boldsymbol{s}_{J+1}(t_j) = \sum_{t_j \in T^{\alpha_l}} \boldsymbol{v}_1(t_j), \quad l = 1, \ldots, e & \text{(5c)}\\
\ldots & \\
\sum_{t_j \in T^{\alpha_l}} \boldsymbol{s}_{J+K}(t_j) = \sum_{t_j \in T^{\alpha_l}} \boldsymbol{v}_K(t_j), \quad l = 1, \ldots, e &
\end{cases}$$

A benchmark to compare graph-based vs optimization-based approaches

1. Optimization-based approach → K–diagnosability via solution of ILP problems
2. Graph-based approach → the **semi-symbolic diagnoser (SSD)**

A. Boussif, M. Ghazel, K. Klai
Combining enumerative and symbolic techniques for diagnosis of discrete-event systems
Workshop on Verification and Evaluation of Computer and Communication Systems, 2015

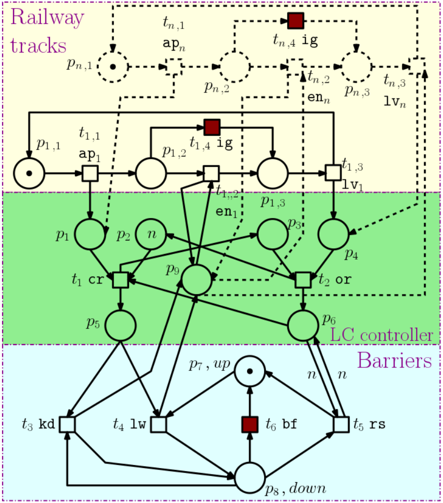
- The comparison is carried out using the *modular* railway benchmark presented in

A. Boussif, B. Liu, M. Ghazel
An experimental comparison of three diagnosis techniques for discrete event systems
International Workshop on Principles of Diagnosis, 2017

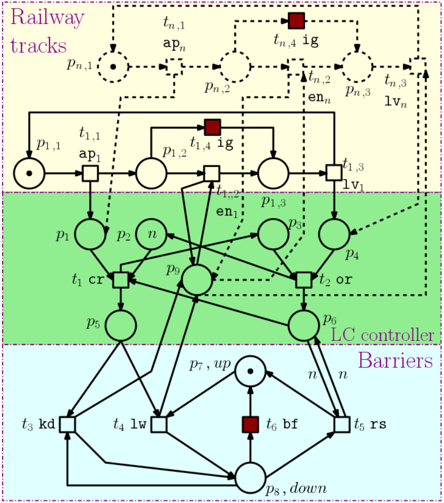
- **Objective: efficiency assessment of the *optimization-based* algorithm 2 → The *graph-based* approach 1 was choosen since it outperforms other approaches on the considered benchmark**

M. Ghazel and B. Liu
A customizable railway benchmark to deal with fault diagnosis issues in DES
Workshop on Discrete Event Systems, 2016

- **Modular** PN model of a railway system that includes
  - *n* tracks
  - level crossing (LC) controller
  - the barriers
- Two *classes* of fault events are modeled by unobservable transitions
  - the *i*-th transition ($t_{i,4}$, `ig`) indicates that the *i*-th train enters the LC zone before the controller lowers the barriers;
  - the transition (t_6, `bf`) indicates a defect in the barriers that results in a premature raising.



- The proposed *optimization-based* approach cannot be used to assess non-diagnosability
- The fault $(t_{i,4}, \texttt{ig})$ is not diagnosable when $n > 1$.
- **Only $(t_6, \texttt{bf})$ will be considered for the comparison**



Experimental setup

- In order to apply the chosen *optimization-based* approach, a Matlab® script that calls the FICO™ Xpress API to solve the ILP problem was used **(off-the-shelf software)**
- The SSD approach is implemented by the DPN-SOG tool **(ad hoc software tool)**
- The hardware platform was a 64-bit PC equipped with CPU Intel® Core™ i3-6100U, at 2.30 GHz with 4GB of RAM

Comments

- The current implementation of the SSD approach within DPN-SOG permits to assess diagnosability but not K-diagnosability
- The considered *ILP-based* approach cannot be used to assess non-diagnosability
- **The comparison is made only on fault** $(t_6, \mathtt{bf})$

Experimental results

n	Petri net features				Diagnosability via SSD				$\mathcal{K}$-diagnosability via ILP				
	$\|P\|$	$\|T\|$	$\|\mathcal{N}\|$	$\|\mathcal{A}\|$	$\|\mathcal{D}_S\|$	$\|\mathcal{D}_T\|$	$\mathcal{D}_e$ (s)	$\mathcal{D}_m$ (kB)	$\mathcal{K}$	Last_ILP$_e$ (s)	Total_ILP$_e$ (s)	#constr. (origin / Xpress)	#unkow. (origin / Xpress)
1	12	10	20	43	10	14	**0**	44	7	0.3	**9**	721 / 225	228 / 180
2	15	14	142	500	83	205	**0**	1056	13	0.6	**26**	1171 / 467	425 / 380
3	18	18	832	4085	483	1745	**1**	8696	19	0.7	**56**	1729 / 798	682 / 639
4	21	22	4314	27142	2434	11774	**2**	80400	25	1.1	**108**	2395 / 1237	999 / 923
5	24	26	20556	157551	11304	69112	**30**	430456	31	4	**194**	3169 / 1764	1376 / 1294
6	27	30	92070	831384	56136	414299	**458**	2155100	37	2.7	**326**	4051 / 2386	1813 / 1725
7	30	34	393336	4086585	261262	2282890	**7836**	10167015	43	5.6	**507**	5041 / 3110	2310 / 2197
8	33	38	1618866	19013130	*	*	o.t.	*	49	6.4	**767**	6139 / 3940	2867 / 2743
9	36	42	*	*	*	*	o.t.	*	55	8.8	**1079**	7345 / 5006	3484 / 3351
10	39	46	*	*	*	*	o.t.	*	61	11.8	**1514**	8659 / 6013	5822 / 4017
11	42	50	*	*	*	*	o.t.	*	64	20	**1874**	12519 / 6686	11400 / 4555
12	45	54	*	*	*	*	o.t.	*	67	32	**3630**	13962 / 7688	12798 / 5125

*: No result obtained in 4 hours. o.t.: Out of time (more than 4 hours).

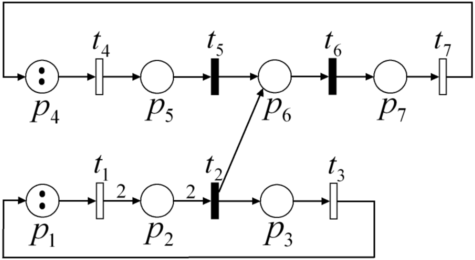
- n: the number of tracks;
- $|P|$ and $|T|$: the number of places and transitions in the PN models, respectively;
- $|\mathcal{N}|$ and $|\mathcal{A}|$: the number of nodes and arcs in the reachability graph, respectively;
- $|\mathcal{D}_S|$ and $|\mathcal{D}_T|$: the numbers of nodes and arcs in the SSD, respectively;
- $\mathcal{D}_e$ and $\mathcal{D}_m$: the required time and memory to generate perform the verification respectively;
- $\mathcal{K}$: number of events needed to detect the fault;
- Last_ILP$_e$: the time taken by Xpress to solve the ILP problem that satisfies Theorem 1;
- Total_ILP$_e$: the time taken by Xpress to solve the $\mathcal{K}$ ILP problems needed to assess $\mathcal{K}$-diagnosability;
- #const.: the number of constraints in the ILP problem that satisfies Theorem 1 before and after Xpress presolver, respectively;
- #unkow.: the number of unknowns in the ILP problem that satisfies Theorem 1 before and after Xpress presolver, respectively.

Results

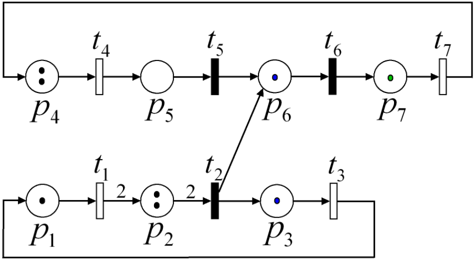
- The proposed *optimization-based* approach requires to solve a number of ILP problems equal to K to assess K-diagnosability
- As soon as the size of the model becomes relatively large (in our case, as soon as $n > 6$), the time needed to perform the analysis becomes way lower than the one required by the *graph-based* SSD approach
  - The SSD algorithm has been directly implemented in C++
  - The *ILP-based* approach has been deployed in the Matlab® environment and relies on the FICO™ Xpress API
  - There is a time overhead for the latter approach that is bigger than for the former, and this fact may have a non negligible impact when the size of the problem is relatively small
- Given the exponential explosion of the state space, the *graph-based* approach becomes practically unfeasible for $n > 7$, not terminating within the 4 hours timeout that was considered for the adopted platform

Results

- Since it does not require the explicit computation of the reachability set, the *ILP-based* is particularly well suited for LPN models with a high level of *parallelism*
  - An additional track has a significant impact on the size of the model state space, but it does not affect too much the efficiency of *ILP-based* approach
  - This result is achieved thanks to the fact that the algebraic formulation enables to exploit the parallelism in the dynamic evolution of each track, and that the tracks evolve in parallel
- The *ILP-based* approach exploits commercial tools for the solution of the ILP problems
  - This permits to takes advantage of all the preprocessing processes of these commercial tools
  - In the considered case, the number of constraints and unknowns after the run of the Xpress presolver is always smaller than the one of the original ILP problem, and this has a positive impact on the time needed to solve the problem



$\boldsymbol{m}_0 = \begin{bmatrix} 2 & 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$ - t_1 fires.



$\boldsymbol{m}_1 = \begin{bmatrix} 1 & 2 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$

$\boldsymbol{m}_2 = \begin{bmatrix} 1 & 0 & 1 & 2 & 0 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$ - if t_2 has fired

$\boldsymbol{m}_3 = \begin{bmatrix} 1 & 0 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$ - if t_2 and t_6 have fired

Fault detection via solution of optimization problems

In order to cope with the state space estimation explosion

- fault detection can be performed by means of the on-line solution of ILP problems
- the concept of *generalized marking* is introduced
- at each step the estimated generalized marking **is always unique**

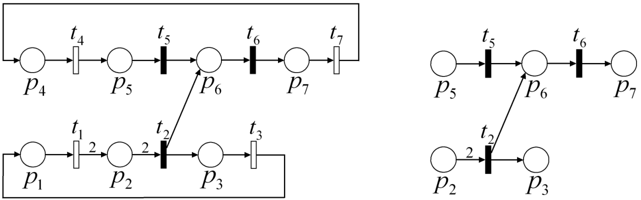
Induced subnets

T'-Induced subnet

Given a net $N = (P, T, \mathbf{Pre}, \mathbf{Post})$, and a subset $T' \subseteq T$, the T'-induced subnet on N, denoted with $N' \prec_{T'} N$, is the 4-tuple $N' = (P', T', \mathbf{Pre}', \mathbf{Post}')$, where $P' = ^{\bullet} T' \cup T'^{\bullet}$, while $\mathbf{Pre}'$ and $\mathbf{Post}'$ are the restrictions of $\mathbf{Pre}$ and $\mathbf{Post}$ to P' and T'.

The subnet $N' \prec_{T'} N$ can be obtained from N by removing all the places which are not connected to any transition in T', and all the transitions in $T \setminus T'$

Induced subnets - Example



(a) A net N.

(b) The $N_{uo} \prec_{T_{uo}} N$ subnet.

Figure: Example of induced subnet

- Unlabeled systems
- $N_{uo} \prec_{T_{uo}} N$ is *acyclic*
- **The state equation of the $N_{uo} \prec_{T_{uo}} N$ subnet does not admit spurious solutions**
- **The fulfilment of the state equation is a necessary and sufficient condition for reachability**

Generalized marking $\boldsymbol{\mu}$

A *generalized marking* is a function

$$\boldsymbol{\mu} : P \rightarrow \mathbb{Z}$$

A transition t is enabled at $\boldsymbol{\mu}$ *if and only if*

ia) $t \in T_o$,

iia) $t \in T_{uo}$ and $\exists\ \sigma \in T_{uo}^*$ s.t. $\boldsymbol{\mu}' = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{\sigma} \geq \mathbf{0}$, $t \in \sigma$, with $\boldsymbol{\sigma} = \pi(\sigma)$

The notation $\boldsymbol{\mu}[t\rangle$ denotes that t is enabled at $\boldsymbol{\mu}$

A transition t may fire if

ib) $t \in T_o$ is enabled and its firing has been observed

iib) $t \in T_{uo}$ is enabled

When a transition t fires, it yields the generalized marking $\boldsymbol{\mu}' = \boldsymbol{\mu} + \boldsymbol{C}(\cdot, t)$, this is denoted as $\boldsymbol{\mu}[t\rangle\boldsymbol{\mu}'$

Negative markings

- The negative components of μ represent the tokens that are needed to explain
  - the firing of an observed transition
  - the firing of an unobservable transition that must have fired
- As far as the fault diagnosis is concerned, μ allows to store in a compact way all the needed information about the state estimate

Unobservable explanations

Given a generalized marking $\boldsymbol{\mu} \in \mathbb{Z}^m$

$$\Sigma(N, \boldsymbol{\mu}) = \{\sigma \in T_{uo}^* \mid \boldsymbol{\mu}[\sigma\rangle\boldsymbol{\mu}' \text{ s.t. } \boldsymbol{\mu}' \geq 0\}$$

is the set of all the unobservable explanations enabled at $\boldsymbol{\mu}$ and

$$\Sigma_f(N, \boldsymbol{\mu}, t_f) = \{\sigma \in T_{uo}^* \mid \boldsymbol{\mu}[\sigma\rangle\boldsymbol{\mu}' \text{ s.t. } \boldsymbol{\mu}' \geq 0 \text{and } \boldsymbol{\sigma}(t_f) \neq 0\,,\ \text{with } \boldsymbol{\sigma} = \pi(\sigma)\}$$

is the set of all the faulty unobservable explanations which includes the fault t_f enabled at $\boldsymbol{\mu}$

The sets

$$\boldsymbol{\Sigma}(N, \boldsymbol{\mu}) = \{\boldsymbol{\sigma} \in \mathbb{N}^n \mid \exists\, \sigma \in \Sigma(N, \boldsymbol{\mu}) \text{ s.t. } \pi(\sigma) = \boldsymbol{\sigma}\}$$

and

$$\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f) = \{\boldsymbol{\sigma} \in \mathbb{N}^n \mid \exists\, \sigma \in \Sigma_f(N, \boldsymbol{\mu}, t_f) \text{s.t. } \pi(\sigma) = \boldsymbol{\sigma}\}$$

are the corresponding set of firing count vectors

Theorem 1

Given a net N with $T = T_o \cup T_{uo}$. Let μ be a generalized marking, $t_f \in T_f \subseteq T_{uo}$ a fault transition, then

$$|\boldsymbol{\Sigma}(N,\mu)| = |\boldsymbol{\Sigma}_f(N,\mu,t_f)| \iff \min_{\sigma \in \boldsymbol{\Sigma}(N,\mu)} \sigma(t_f) \neq 0\,.$$

Corollary 1

Given a net N with $T = T_o \cup T_{uo}$. Let μ be a generalized marking, $t_f \in T_f \subseteq T_{uo}$ a fault transition, then

$$|\boldsymbol{\Sigma}(N,\mu)| = |\boldsymbol{\Sigma}_f(N,\mu,t_f)| \iff \begin{array}{l} \forall\, \sigma \in \boldsymbol{\Sigma}(N,\mu)\,, \\ \sigma(t_f) \neq 0\,. \end{array}$$

Theorem 2

Given a net N with $T = T_o \cup T_{uo}$. Let μ be a generalized marking, $t_f \in T_f \subseteq T_{uo}$ a fault transition, then

$$|\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| \neq 0 \iff \max_{\boldsymbol{\sigma} \in \boldsymbol{\Sigma}(N, \boldsymbol{\mu})} \boldsymbol{\sigma}(t_f) \neq 0 \,.$$

Corollary 2

Given a net N with $T = T_o \cup T_{uo}$. Let μ be a generalized marking, $t_f \in T_f \subseteq T_{uo}$ a fault transition, then

$$|\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| \neq 0 \iff \exists\, \boldsymbol{\sigma} \in \boldsymbol{\Sigma}(N, \boldsymbol{\mu}) \,,\, \boldsymbol{\sigma}(t_f) \neq 0 \,,$$

and

$$|\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| = 0 \iff \forall\, \boldsymbol{\sigma} \in \boldsymbol{\Sigma}(N, \boldsymbol{\mu}) \,,\, \boldsymbol{\sigma}(t_f) = 0 \,.$$

Compute min and max

Since $N_{uo} \prec_{T_{uo}} N$ is *acyclic*, then

$$\mathbf{\Sigma}(N,\boldsymbol{\mu}) = \left\{\boldsymbol{\sigma} \in \mathbb{N}^n \mid \boldsymbol{C}_{uo}\boldsymbol{\sigma}_{|T_{uo}} \geq -\boldsymbol{\mu}_{|P_{uo}} \text{ and } \boldsymbol{\sigma}_{|T_o} = \mathbf{0}\right\},$$

thus $\min_{\boldsymbol{\sigma}\in\mathbf{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\sigma}(t_f)$ and $\max_{\boldsymbol{\sigma}\in\mathbf{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\sigma}(t_f)$ can be computed by solving an Integer Linear Programming (ILP) problem

- Let $t_f \in T_f \subseteq T_{uo}$ and $\boldsymbol{\mu}$ a generalized markings.
- As far as the detection of t_f is concerned, the following three conditions have to be checked

1a) $\boldsymbol{\mu} \ngeq \mathbf{0}$ and $|\boldsymbol{\Sigma}(N, \boldsymbol{\mu})| = |\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| > 0 \Rightarrow$ t_f has occurred

2a) $|\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| = 0 \iff$ t_f has not occurred

3a) $|\boldsymbol{\Sigma}_f(N, \boldsymbol{\mu}, t_f)| \neq 0 \iff$ t_f may have occurred

The three conditions listed above are equivalent to

1b) $\boldsymbol{\mu} \ngeq \mathbf{0}$ and $\min_{\epsilon \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \epsilon(t_f) \neq 0 \Rightarrow$ t_f has occurred

2b) $\max_{\epsilon \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \epsilon(t_f) = 0 \iff$ t_f has not occurred

3b) $\max_{\epsilon \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \epsilon(t_f) \neq 0 \iff$ t_f may have occurred

Require: $\boldsymbol{C}, \boldsymbol{m}_0, T_o, T_{uo}, T_f$

1 $\boldsymbol{\mu} = \boldsymbol{\mu}_0 = \boldsymbol{m}_0$ (* Initialization *)

2 **for all** $t_{f_i} \in T_f$ **do**

  2.1 **if** $\boldsymbol{\mu} \ngeq \mathbf{0}$,
  then (* if the g-marking has at least one negative component *)

    2.1.1 **if** $\min_{\boldsymbol{\epsilon} \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\epsilon}(t_{f_i}) = F \neq 0$,
    then (* t_{f_i} has occurred F times *)

      2.1.1.1 report that t_{f_i} **has occurred**
      2.1.1.2 $\boldsymbol{\mu}_{|P_{uo}} = \boldsymbol{\mu}_{|P_{uo}} + \boldsymbol{C}_{uo}(\cdot, t_{f_i})F$ (* Update $\boldsymbol{\mu}$ *)
      2.1.1.3 **go to** Step **2** (* Restart the **for** cycle *)

  2.2 **if** $\max_{\boldsymbol{\epsilon} \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\epsilon}(t_{f_i}) = G \neq 0$,
  then report that t_{f_i} **may have occurred**
  (* t_{f_i} may be occurred G times *)

  2.3 **else** report that t_{f_i} **has not occurred yet**

3 **end for**

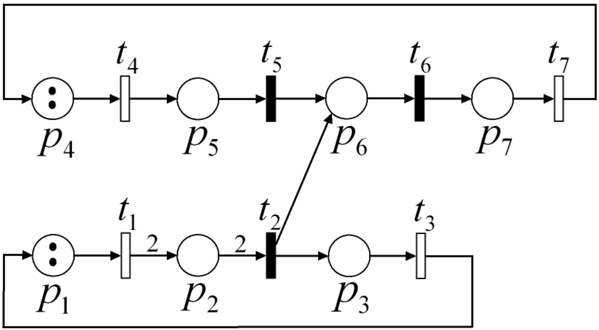
4 **if** $\boldsymbol{C}_{uo}\boldsymbol{\epsilon}_{|T_{uo}} \geq -\boldsymbol{\mu}_{|P_{uo}}$ admits only one solution $\boldsymbol{\epsilon}^*$ s.t. $\boldsymbol{\epsilon}^*_{|T_o} = \mathbf{0}$,
then $\boldsymbol{\mu}_{|P_{uo}} = \boldsymbol{\mu}_{|P_{uo}} + \boldsymbol{C}_{uo}\boldsymbol{\epsilon}^*_{|T_{uo}}$ (* Update $\boldsymbol{\mu}$ *)

5 **wait for** a new observed transition $\bar{t} \in T_o$

6 $\boldsymbol{\mu} = \boldsymbol{\mu} + \boldsymbol{C}(\cdot, \bar{t})$ (* Update $\boldsymbol{\mu}$ *)

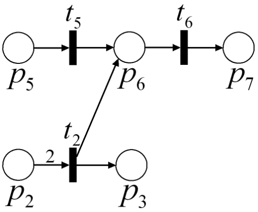
7 **go to** Step **2**

Example



Let $\mu_0 = \left[2\ 0\ 0\ 2\ 0\ 0\ 0\right]^{\mathrm{T}}$, and $T_f = \{t_5\}$.

Example



The $N_{uo} \prec_{T_{uo}} N$ subnet is TS2, thus the ILP problems $\min_{\sigma \in \Sigma(N,\mu)} \sigma(t_5)$ and $\max_{\sigma \in \Sigma(N,\mu)} \sigma(t_5)$ admit the following *closed - form* solutions:

$$\min_{\sigma \in \Sigma(N,\mu)} \sigma(t_5) = \max\left(-\mu_{|p_6} - \mu_{|p_7} - \left\lfloor \frac{\mu_{|p_2}}{2} \right\rfloor, 0\right)$$

$$\max_{\sigma \in \Sigma(N,\mu)} \sigma(t_5) = \mu_{|p_5}$$

Y. Li, W. M. Wonham

Control of vector discrete-event systems II – Controller synthesis

IEEE Transactions on Automatic Control, 1994

Example

Action	$\boldsymbol{\mu}$	$\min_{\boldsymbol{\sigma} \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\sigma}(t_5)$	$\max_{\boldsymbol{\sigma} \in \boldsymbol{\Sigma}(N,\boldsymbol{\mu})} \boldsymbol{\sigma}(t_5)$
Initialization	$\begin{bmatrix} 2 & 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}^T$	0	0
t_1 fires	$\begin{bmatrix} 1 & 2 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}^T$	0	0
t_4 fires	$\begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}^T$	0	1
t_7 fires	$\begin{bmatrix} 1 & 2 & 0 & 2 & 1 & 0 & -1 \end{bmatrix}^T$	0	1
t_7 fires	$\begin{bmatrix} 1 & 2 & 0 & 3 & 1 & 0 & -2 \end{bmatrix}^T$	1	1
Update $\boldsymbol{\mu}$ (Step **2.1.2**)	$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 1 & -2 \end{bmatrix}^T$	0	0
Update $\boldsymbol{\mu}$ (Step **4**)	$\begin{bmatrix} 1 & 0 & 1 & 3 & 0 & 0 & 0 \end{bmatrix}^T$	0	0

References

F. Basile, P. Chiacchio, G. De Tommasi
On $\mathcal{K}-$ *diagnosability* of Petri nets via integer linear programming
Automatica, 2012

F. Basile, P. Chiacchio, G. De Tommasi
An Efficient Approach for Online Diagnosis of Discrete Event Systems
IEEE Transactions on Automatic Control, 2009

Diagnosability and fault detection in PNs - Part II: algebraic approaches for bounded systems

From observability to privacy and security in discrete event systems

Prof. Gianmaria De Tommasi
Email: detommas@unina.it

December 2020

DIETI. UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELLE TECNOLOGIE DELL'INFORMAZIONE