2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Marking Diagnosis in Labeled Petri Nets Using Basis Diagnosers

Ziyue Ma[1], Yin Xiang[2], Zhiwu Li[1,3]

Abstract— **This paper studies the marking diagnosis problem in labeled Petri nets, i.e., to determine whether a plant has reached a given set of *faulty markings* or not in the past, from the observation history. We observe that the conventional *basis reachability graphs* cannot be used for marking diagnosis due to the existence of *partially faulty basis markings*. To overcome such a problem, we propose a notion called the *diagnostic basis partition* such that the corresponding basis reachability graphs does not contain any partially-faulty basis markings. By properly selecting a set of explicit transitions, a diagnostic agent is synthesized to perform the online diagnosis.**

I. Introduction

Fault diagnosis [1] in *discrete event systems* (DESs) has drawn considerable attention over the past decades. The aim of fault diagnosis is to determine whether some particular events called *faults* have occurred or not according to the observation history. In recent years, abundant results on fault diagnosis in *Petri nets* have been achieved [2], [3], [4], [5]. In particular, in [6], an automaton-like structure called the *basis reachability graph* was developed for diagnosis with the assumption that the unobservable subnet is acyclic.

The works in the literature consider *faults* as the occurrence of certain events (in automata) and the firings of transitions (in Petri nets). On the other hand, in many practical situations, an operator of a plant may be not interested in tracking the occurrence of events in the system but wants to determine if the plant has reached some certain states (i.e., markings) in the past. For example, in an automated production line, when the content in a buffer exceeds a threshold, an alarm is expected to be issued in a finite number of future steps such that some actions of maintenance are properly taken. This motivates the problem of *marking diagnosis*. In plain words, the goal of marking diagnosis is to determine if the plant has reached a given set of markings of physical importance, namely *faulty markings*, according to the observation history.

Event-based and state-based diagnosis are equivalent in automaton models [7]. However, in Petri nets, methods for transition-based fault diagnosis are not applicable for marking diagnosis. The fact that the firing of a transition at one marking yields a faulty marking does not necessary mean that the firing of such a transition at any marking *always* leads to a faulty marking. Therefore, one cannot simply convert the marking diagnosis problem to a transition diagnosis problem based on the original net. Hence, for Petri nets, it is desirable to investigate the marking diagnosis problem in a more efficient way using structural analysis techniques.

In this paper, we investigate the marking diagnosis problem in plants modeled by *labelled Petri nets*. To efficiently perform marking diagnosis without constructing the entire reachability graph, we develop a method based on the *basis reachability graphs* (BRGs), which have been proved to be an efficient tool for abstracting the state space of a Petri net [6], [8], [9], [10], [11], [12]. However, the conventional BRG-based methods designed for transition diagnosis cannot be used for marking diagnosis due to the existence of *partially-faulty basis markings* (which will be defined in Section III). To overcome such a problem, we propose a notion called the *diagnostic basis partition* such that the corresponding BRG, called the *diagnostic BRG*, does not contain any partially-faulty basis markings. We then develop a structure called the *marking basis diagnoser* for marking diagnosis. Since the number of basis markings is generally much smaller than the number of reachable markings, the proposed method for marking diagnosability verification in LPNs is of efficiency.

The rest of this paper is organized as follows. Basic notions of Petri nets and BRGs are recalled in Section II. The marking diagnosis problem is formulated in Section III. In Section IV, some useful notions for marking diagnosis in the BRGs are introduced. In Section V, diagnostic basis partition and diagnostic BRGs are defined, and a method for marking diagnosis using *basis diagnoser* is developed. Section IV draws the conclusion.

II. Preliminaries

A. Petri Nets

A Petri net is a four-tuple $N = (P, T, Pre, Post)$, where P is a set of m *places* represented by circles; T is a set of n *transitions* represented by bars; $Pre : P \times T \to \mathbb{N}$ and $Post : P \times T \to \mathbb{N}$ are the *pre-* and *post-incidence functions*, respectively, specifying the arcs in the net and can also be represented as matrices in $\mathbb{N}^{m\times n}$ (here $\mathbb{N} = \{0, 1, 2, \ldots\}$). The *incidence matrix* of a net is defined by $C = Post - Pre \in \mathbb{Z}^{m\times n}$ (here $\mathbb{Z} = \{0, \pm1, \pm2, \ldots\}$).

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61472295, 61703321, 61873342, Shaanxi Provincial Natural Science Foundation under Grant No. 2019JQ-022, the Fundamental Research Funds for the Central Universities under Grant JB190407, and the Science and Technology Development Fund, MSAR, under Grant No. 0012/2019/A1.

[1]School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China (e-mail: maziyue@xidian.edu.cn, maziyue@gmail.com).

[2]Department of Automation and the Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, Shanghai 201108, China (e-mail: yinxiang@sjtu.edu.cn).

[3]Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macau (e-mail: zhwli@xidian.edu.cn, systemscontrol@gmail.com).

A *marking* is a function $M : P \to \mathbb{N}$ that assigns each place of a Petri net a non-negative integer number of tokens, represented by black dots; a marking can also be represented as an m-component vector. We denote by $M(p)$ the number of tokens in place p at marking M. A *marked net* $G = \langle N, M_0\rangle$ is a net N with an initial marking M_0. Marking $[x_1, \ldots, x_m]^T$ is also denoted as $x_1p_1 + \cdots + x_mp_m$.

A transition t is *enabled* at a marking M if $M \geq Pre(\cdot, t)$. If t is enabled at M, then it may fire and reach a new marking $M' = M + C(\cdot, t)$. We write $M[t\rangle$ and $M[t\rangle M'$ to denote, respectively, that transition t is enabled at M and its firing yields M'. We denote by T^* the set of all finite sequences of transitions over T. Then, $M[\sigma\rangle M'$ analogously denotes that a sequence $\sigma = t_1t_2 \ldots t_k \in T^*$ is enabled (sequentially) at M and its firing finally yields M'. In this case we say that M' is *reachable* from M. We denote by $R(N, M_0)$ the set of all markings reachable from the initial marking M_0. The *language* of $\langle N, M_0\rangle$ is defined as $L(N, M_0) = \{\sigma \in T^* \mid M_0[\sigma\rangle\}$. For any sequence $\sigma \in T^*$, $\mathbf{y}_\sigma$ denotes its *firing vector*, i.e., $y_\sigma(t) = k$ if transition t occurs k times in σ.

Given a sequence $\sigma \in T^*$, the *prefix-closure* of σ is defined as $Pr(\sigma) = \{\sigma' \in T^* \mid (\exists \sigma'' \in T^*)\sigma = \sigma'\sigma''\}$.

Given a Petri net $N = (P, T, Pre, Post)$, net $\hat{N} = (\hat{P}, \hat{T}, \hat{P}re, \hat{P}ost)$ is a *subnet* of N if $\hat{P} \subseteq P$, $\hat{T} \subseteq T$, and $\hat{P}re, \hat{P}ost$ are the restriction of $Pre, Post$ to $\hat{P} \times \hat{T}$, respectively (i.e., only rows and columns associated with $\hat{P}$ and $\hat{T}$ are kept). In particular, $\hat{N}$ is called the *$\hat{T}$-induced subnet* if $\hat{N} = (P, \hat{T}, \hat{P}re, \hat{P}ost)$.

B. Labeled Petri Nets

A *labeled Petri net* (LPN) is a 4-tuple $G = (N, M_0, E, \ell)$, where $\langle N, M_0\rangle$ is a marked net, E is the *alphabet* (a set of labels), and $\ell : T \to E \cup \{\varepsilon\}$ is the *labeling function* that assigns each transition $t \in T$ either a symbol from E or the *silent* label ε. This naturally leads to a partition of the transition set as $T = T_o \dot\cup T_{uo}$, where $T_o = \{t \in T \mid \ell(t) \in E\}$ is the set of *observable* transitions and $T_{uo} = T \setminus T_o = \{t \in T \mid \ell(t) = \varepsilon\}$ is the set of *unobservable* transitions.

The labeling function can be extended to $\ell : T^* \to E^*$ recursively by: (i) $\ell(\varepsilon) = \varepsilon$; and (ii) $\ell(\sigma t) = \ell(\sigma)\ell(t)$ with $\sigma \in T^*$ and $t \in T$. The *inverse projection* of an observation $w \in E^*$ with respect to $G = (N, M_0, E, \ell)$ is defined as: $\ell^{-1}(w) = \{\sigma \in L(N, M_0) \mid \ell(\sigma) = w\}$. The *language* of an LPN $G = (N, M_0, E, \ell)$ is defined as $L(G) = \{\ell(\sigma) \mid \sigma \in L(N, M_0)\}$. The *set of consistent markings* of an observation $w \in L(G)$ is defined as $\mathcal{C}(w) = \{M \in R(N, M_0) \mid (\exists \sigma : \ell(\sigma) = w)\ M_0[\sigma\rangle M\}$.

C. Basis Reachability Graph

Definition 1: [8] Given a Petri net $N = (P, T, Pre, Post)$, a pair $\pi = (T_E, T_I)$ is called a *basis partition* of T if (1) $T_I \subseteq T$, $T_E = T \setminus T_I$; and (2) the T_I-induced subnet is acyclic. The sets T_E and T_I are called the set of *explicit transitions* and the set of *implicit transitions*, respectively. □

Definition 2: Given a Petri net $N = (P, T, Pre, Post)$, a basis partition $\pi = (T_E, T_I)$, a marking M, and a transition $t \in T_E$, we define:

- $\Sigma(M, t) = \{\sigma \in T_I^* \mid M[\sigma\rangle M', M' \geq Pre(\cdot, t)\}$ as the set of *explanations* of transition t at marking M;
- $Y(M, t) = \{\mathbf{y}_\sigma \in \mathbb{N}^{|T|} \mid \sigma \in \Sigma(M, t)\}$ as the set of *explanation vectors* of transition t at marking M;
- $Y_{min}(M, t)$ denotes the set of all minimal elements of $Y(M, t)$, i.e., the *minimal explanation vectors*.

□

Definition 3: Given a Petri net $N = (P, T, Pre, Post)$ with an initial marking M_0 and a basis partition $\pi = (T_E, T_I)$, its set of *basis markings* $\mathcal{M}$ is defined as follows:

- $M_0 \in \mathcal{M}$;
- If $M \in \mathcal{M}$, then $\forall t \in T_E$, $\forall \mathbf{y} \in Y_{min}(M, t)$,

$$(M' = M + C \cdot \mathbf{y} + C(\cdot, t)) \Rightarrow (M' \in \mathcal{M}).$$

The *basis reachability graph* (BRG) is a non-deterministic finite state automaton $\mathcal{B}$ output by an algorithm presented in [8]. The BRG $\mathcal{B}$ is a quadruple $(\mathcal{M}, Tr_B, \Delta_B, M_0)$, where:

- the state set $\mathcal{M}$ is the set of *basis markings*;
- the event set Tr_B is the set of pairs $(t, \mathbf{y}) \in T_E \times \mathbb{N}^{|T|}$;
- the transition relation Δ_B is:

$$\Delta_B = \{(M, (t,\mathbf{y}), M') \mid \mathbf{y} \in Y_{min}(M, t), \\ t \in T_E, M' = M + C \cdot \mathbf{y} + C(\cdot, t)\}$$

- initial marking $M_0 \in \mathcal{M}$ is the initial state.

□

For the convenience of presentation, in the sequel of this paper, we use $\phi = (t_{i_1}, \mathbf{y}_{i_1})(t_{i_2}, \mathbf{y}_{i_2}) \cdots (t_{i_n}, \mathbf{y}_{i_n})$ to denote a sequence of labels of arcs in a BRG, i.e., a path $M_{b,1} \to M_{b,2} \to \cdots \to M_{b,n}$, where the arcs on this path are sequentially labeled by $(t_{i_1}, \mathbf{y}_{i_1}), (t_{i_2}, \mathbf{y}_{i_2}), \cdots, (t_{i_n}, \mathbf{y}_{i_n})$. Let $\ell(\phi) = \ell(t_{i_1}t_{i_2} \cdots t_{i_n})$ and $\phi_{\uparrow T_E} = (t_{i_1}t_{i_2} \cdots t_{i_n})_{\uparrow T_E}$ ($\uparrow$ is the *natural projection* operator).

Definition 4: [8] Given a net $G = \langle N, M_0\rangle$ with $\pi = (T_E, T_I)$, the *implicit reach* of a marking M is a set of markings: $R_I(M) = \{M' \mid M[\sigma\rangle M', \sigma \in T_I^*\}$. □

Proposition 1: [8] Given a Petri net $N = (P, T, Pre, Post)$, let $\mathcal{B} = (\mathcal{M}, Tr, \Delta, M_0)$ be the BRG with respect to $\pi = (T_E, T_I)$. The following two statements are equivalent:

1) there exist a marking M and a firing sequence in the form of $\sigma = \sigma_1 t_{i_1} \cdots \sigma_n t_{i_n} \sigma_{n+1}$, where $\sigma_j \in T_I^*$, $t_{i_j} \in T_E$ for all $j \in \{1, \ldots, n+1\}$, such that $M_0[\sigma\rangle M$;
2) there is a following path in the BRG $\mathcal{B}$

$$M_0 \xrightarrow{(t_{i_1}, \mathbf{y}_1)} M_{b,1} \xrightarrow{(t_{i_2}, \mathbf{y}_2)} \cdots \xrightarrow{(t_{i_n}, \mathbf{y}_n)} M_{b,n}$$

such that $M \in R_I(M_{b,n})$;

III. Marking Diagnosability Problem Formulation and Basis Reachability Graphs

The goal of marking diagnosis is to determine if the plant has reached a given set of markings of physical importance, namely *faulty markings* $F \subseteq \mathbb{N}^{|P|}$, according to the observation history. As we have mentioned in the introductory section, the fact that the firing of a transition at one marking yields a faulty marking does not necessary mean that the firing of such a transition at any marking *always* leads to a

faulty marking. Therefore, one cannot simply transform the marking diagnosis problem to a transition diagnosis problem by manipulating the original net. Such a transformation is only possible if the entire reachability graph of the system is constructed (which is quite exhaustive) and being treated as an automaton.

To simplify the presentation, in the sequel of this paper, we assume that a plant LPN $G = (N, M_0, E, \ell)$ with $N = (P, T, Pre, Post)$ satisfies the following assumptions that are widely used in the literatures of fault diagnosis [6], [13], [10]:

- A1: G is deadlock-free;
- A2: G is bounded;
- A3: the T_{uo}-induced subnet is acyclic;

A. Marking Diagnosability Problem Formulation

In this paper, the set of faulty markings F we consider are defined by a *generalized mutual exclusion constraint* (GMEC).

Definition 5: [14] A *generalized mutual exclusion constraint* is a pair $(\mathbf{w}, k)$, where $\mathbf{w} \in \mathbb{Z}^m$ and $k \in \mathbb{Z}$, that defines a set of markings:

$$\mathcal{L}_{(\mathbf{w},k)} = \{M \in \mathbb{N}^m \mid \mathbf{w}^T \cdot M \leq k\}.$$

The *token count* of GMEC $(\mathbf{w}, k)$ at a marking M is defined as the quantity of $\mathbf{w}^T \cdot M$. □

Definition 6: Given an LPN $G = (N, M_0, E, \ell)$ and a set of faulty markings $F \subseteq \mathbb{N}^{|P|}$, we define the *fault language* of a marking M with respect to F as $L_{M,F}$:

$$L_{M,F} = \{\sigma \in L(N, M) \mid \exists \bar{\sigma} \in Pr(\sigma), M[\bar{\sigma}\rangle M' \in F\}.$$

□

In plain words, fault language of $L_{M,F}$ consists of all such sequences σ's that have at least one prefix that reaches set F. Note that this definition does not mean that the final marking reached by the firing $\sigma \in L_{M,F}$ is in F.

Similar to the set-ups of transition-based diagnosis, in the marking diagnosis, a *diagnostic agent* that runs in parallel with the plant and reports if the plant has reached some marking in F in the past. A diagnostic agent can be considered as a function $\mathcal{A} : L(G) \to \{0, 1, 2\}$. Given an observation $w \in L(G)$, the corresponding *diagnostic state* $\mathcal{A}(w)$ is in either of the following three cases:

1) $\mathcal{A}(w) = 0$: the plant must have not passed any faulty markings, i.e., $\ell^{-1}(w) \cap L_{M_0,F} = \emptyset$;
2) $\mathcal{A}(w) = 1$: the plant may have or have not passed some faulty marking, i.e., $\ell^{-1}(w) \cap L_{M_0,F} \neq \emptyset$ and $\ell^{-1}(w) \cap (L(G) \setminus L_{M_0,F}) \neq \emptyset$;
3) $\mathcal{A}(w) = 2$: the plant must have passed some faulty markings, i.e., $\ell^{-1}(w) \subseteq L_{M_0,F}$;

Notice that given an observation $w \in L(G)$, to determine the value of $\mathcal{A}(w)$ it is sufficient to determine if $\ell^{-1}(w) \cap L_{M_0,F} = \emptyset$ and if $\ell^{-1}(w) \subseteq L_{M_0,F}$. Now we can introduce the problem of marking diagnosis in LPNs as the following.

Problem 1 (Marking Diagnosis): Given a plant LPN $G = (N, E, \ell, M_0)$ with $N = (P, T, Pre, Post)$, a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, and an observation $w \in L(G)$, determine (1) if $\ell^{-1}(w) \cap L_{M_0,F} = \emptyset$, and (2) if $\ell^{-1}(w) \subseteq L_{M_0,F}$. □

Note the aim of marking diagnosis is not to determine if the plant is currently at a faulty marking but to determine if the plant *was* at a fault marking some time ago.

Before we proceed, let us briefly recall the result on marking estimations using BRGs.

Definition 7: Given an LPN $G = (N, M_0, E, \ell)$ with $N = (P, T, Pre, Post)$, let $\mathcal{B} = (\mathcal{M}, Tr, \Delta, M_0)$ be the BRG with respect to $\pi = (T_E, T_I)$. The set of *consistent basis markings* of an observation $w \in L(G)$, denoted as $\mathcal{M}(w)$, is the set of basis markings M_i's such that there exists a path ϕ from M_0 to M_i in the BRG such that $\ell(\phi) = w$. □

Proposition 2: [15] Given an LPN $G = (N, M_0, E, \ell)$ with $N = (P, T, Pre, Post)$, let the BRG with respect to $\pi = (T_o, T_{uo})$ be $\mathcal{B} = (\mathcal{M}, Tr, \Delta, M_0)$. For an observation $w \in L(G)$, it holds:

$$\begin{aligned} \mathcal{C}(w) &= \bigcup_{M_b \in \mathcal{M}(w)} R_{uo}(M_b) \\ &= \bigcup_{M_b \in \mathcal{M}(w)} \{M \mid M = M_b + C_{uo} \cdot \mathbf{y}\} \end{aligned} \tag{1}$$

where C_{uo} is the incidence matrix of the unobservable subnet.

IV. Encoding the Information of Faults into BRGs

In this section we show that the conventional BRGs cannot be used for marking diagnosis. To see this, the following definition is useful.

Definition 8: Given an LPN $G = (N, M_0, E, \ell)$, a BRG $\mathcal{B} = (M, Tr, \Delta, M_0)$ with respect to $T_E = T_o$, and a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$:

- the set of *faulty basis markings* is defined as: $\mathcal{F} = \{M_b \in \mathcal{M} \mid R_{uo}(M_b) \subseteq F\}$;
- the set of *partially-faulty basis markings* is defined as: $\mathcal{P} = \{M_b \in \mathcal{M} \mid R_{uo}(M_b) \nsubseteq F \wedge R_{uo}(M_b) \cap F \neq \emptyset\}$;
- a basis marking $M_b \in \mathcal{M} \setminus (\mathcal{F} \cup \mathcal{P})$ is called a *non-faulty basis marking*.

In short words, a basis marking is faulty ($M_b \in \mathcal{F}$) if all markings in its unobservable reach belong to F, and a basis marking is partially- faulty ($M_b \in \mathcal{P}$) if not all but at least one marking in its unobservable reach belongs to F. Both sets $\mathcal{F}$ and $\mathcal{P}$ can be computed using the following proposition.

Proposition 3: Given an LPN $G = (N, M_0, E, \ell)$, a BRG $\mathcal{B} = (\mathcal{M}, Tr, \Delta, M_0)$ with respect to $\pi = (T_o, T_{uo})$, and a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$,

1) a basis marking $M_b \in \mathcal{F}$ if and only if the following integer linear constraint Eq. (2) is *not* feasible:

$$\begin{cases} M_b + C_{uo} \cdot \mathbf{y} \geq \mathbf{0} \\ \mathbf{w}^T \cdot (M_b + C_{uo} \cdot \mathbf{y}) \geq k + 1 \\ \mathbf{y} \geq \mathbf{0}. \end{cases} \tag{2}$$

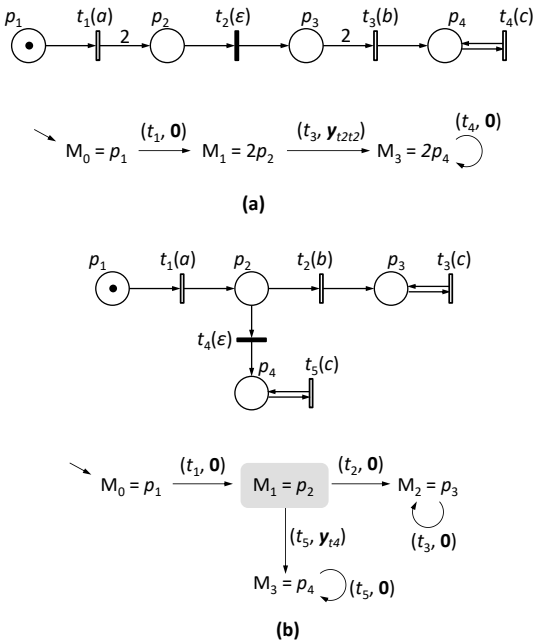

Fig. 1: The LPNs used in Example 1.

2) a basis marking $M_b \in \mathcal{P}$ if and only if integer constraint Eq. (2) is feasible and the following integer constraint Eq. (3) is also feasible:

$$\begin{cases} M_b + C_{uo} \cdot \mathbf{y}' \geq \mathbf{0} \\ \mathbf{w}^T \cdot (M_b + C_{uo} \cdot \mathbf{y}') \leq k \\ \mathbf{y}' \geq \mathbf{0}. \end{cases} \tag{3}$$

Proof: The acyclicity of the unobservable subnet ensures that the state equation gives a sufficient and necessary condition for reachability. On one hand, a basis marking $M_b \in \mathcal{F}$ if and only if Eq. (2) is not feasible, i.e., there is no marking $M \in R_{uo}(M_b)$ such that $M \notin F$. On the other hand, $M_b \in \mathcal{P}$ if and only if Eqs. (2) and (3) are both feasible, i.e., there exists two markings $M, M' \in R_{uo}(M_b)$ such that $M \in F$ and $M' \notin F$. □

One may intuitively conjecture that marking diagnosis can be done by inspecting if the consistent paths in the BRG passes faulty and/or partially-faulty basis markings, which is proved to be an efficient technique for transition-based diagnosis. Precisely speaking, given an observation w, one may think that:

1) "$\mathcal{A}(w) = 1$ if there exists a consistent path of w in the BRG which passes $\mathcal{F} \cup \mathcal{P}$";
2) "$\mathcal{A}(w) = 2$ if and only if all consistent paths of w in the BRG pass $\mathcal{F}$".

Unfortunately, the following example shows that both conjectures are false.

Example 1: Consider the LPN G with a set of faulty markings $F = \{M \mid M(p_3) \geq 2\}$ shown in Figure 1 (a). Its BRG with respect to $T_E = \{t_1, t_3, t_4\}$, shown in the same figure, has three basis markings such that $\mathcal{F} = \emptyset$ and $\mathcal{P} = \{M_1\}$ where $M_1 = 2p_2$. The consistent path of observation $w = ab$ is

$$M_0 \xrightarrow{(t_1, \mathbf{0})} M_1 \xrightarrow{(t_3, \mathbf{y}_{t_2 t_2})} M_2.$$

One can readily verify that all trajectories coinciding with ab necessarily pass set F: to generate event b, transition t_3 must fire, which implies that place p_3 necessarily holds two tokens before firing t_3.

As another example, consider the LPN G with a set of faulty markings $F = \{M \mid M(p_4) \geq 1\}$ shown in Figure 1 (b). Its BRG with respect to $T_E = \{t_1, t_2, t_3, t_5\}$, shown in the same figure, has four basis markings such that $\mathcal{F} = \{M_3\}$ and $\mathcal{P} = \{M_1\}$ where $M_1 = p_2$. The consistent path of observation $w = ab$ is

$$M_0 \xrightarrow{(t_1, \mathbf{0})} M_1 \xrightarrow{(t_2, \mathbf{0})} M_2.$$

However, all trajectories coinciding with ab are $t_1 t_2 t_3^n$ and hence necessarily do not pass set F. □

Example 1 shows that, given a path in a BRG that passes a partially-faulty basis marking, there may or may not exist a consistent trajectory that pass F, and there may or may not exist a consistent trajectory that does not pass F. Hence a consistent path in a BRG passes $\mathcal{P}$ does not provide us any information on the passage of set F. The existence of trajectories passing or not passing F can only be verified by exploring all unobservable trajectories from these partially-faulty basis markings. However, to fully explore the unobservable reach of a basis marking is quite exhaustive. As a result, in the next section we propose a notion of basis partition called the *diagnostic basis partition*, and we prove that a BRG with respect to a diagnosis basis partition does not contain any partially-faulty basis markings.

V. Marking Diagnosis Using Diagnostic BRGs and Basis Diagnosers

A. Diagnostic Basis Partition

Definition 9: Given an LPN in which the set of transitions is $T = T_o \cup T_{uo}$ and a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, the basis partition $\pi_d = (T_{E,d}, T_{I,d})$ is called a *diagnostic basis partition* if

$$T_{E,d} = T_o \cup \{t \in T \mid \mathbf{w}^T \cdot C(\cdot, t) \neq 0\}. \tag{4}$$

The BRG with respect to the diagnostic basis partition π_d is called the *diagnostic BRG* and is denoted as $\mathcal{B}_d$. □

In plain words, in the diagnostic basis partition, the set of explicit transitions consists of all observable transitions and all transitions whose influence on $(\mathbf{w}, k)$ are non-zero. The following proposition shows that in the corresponding diagnostic BRG there is no partially-faulty basis markings.

Proposition 4: Given an LPN $G = (N, M_0, E, \ell)$ and a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, the corresponding diagnostic BRG $\mathcal{B}_d$ does not contain any partially-faulty basis markings, i.e., $\mathcal{P} = \emptyset$.

Proof: Let M_b be an arbitrary basis marking in the diagnostic BRG. Since by Eq. (4), for all transitions $t \in T_I$, $\mathbf{w}^T \cdot C(\cdot, t) = 0$ holds, we have $\mathbf{w}^T \cdot M = \mathbf{w}^T \cdot M'$ for all $M, M' \in R_I(M_b)$. Hence $M_b \in F$ if and only if all marking

$M \in R_I(M_b)$ belongs to $M \in F$ holds, which concludes the proof. □

Since $T_{E,d}$ is a (possibly proper) superset of T_o, the size of a diagnostic BRG is in general larger than the conventional BRG with respect to partition (T_o, T_{uo}). However, the following proposition shows that, the passage of faulty markings can be identified by examining the consistent paths in the diagnostic BRG, since the interfere of partially-faulty basis markings is excluded.

Proposition 5: Given an LPN $G = (N, M_0, E, \ell)$, a set of faulty markings F, and the diagnostic BRG $\mathcal{B}_d$,

- (i) there exists a sequence $\sigma \in L_{M_0,F}$ if and only if there exists a path $M_0 \xrightarrow{\phi}$ in $\mathcal{B}$ such that $\ell(\sigma) = \ell(\phi)$ and passes at least one faulty basis marking in $\mathcal{F}$;
- (ii) there exists a sequence $\sigma \notin L_{M_0,F}$ if and only if there exists a path $M_0 \xrightarrow{\phi}$ in $\mathcal{B}$ such that $\ell(\sigma) = \ell(\phi)$ and does not pass any faulty basis marking in $\mathcal{F}$.

Proof: By Proposition 1, the "⇐" for (i) and (ii) are both trivial. Now we prove "⇒".

For (i), let $\sigma = \sigma_1 t_{i_1} \cdots \sigma_n t_{i_n} \sigma_{n+1}$ be a sequence in $L_{M_0,F}$, i.e., $M_0[\sigma\rangle M$ passes some faulty marking $M_F \in \mathcal{F}$. By the argument of [6], there exists another sequence $\sigma' = \sigma'_1 t_{i_1} \cdots \sigma'_n t_{i_n} \sigma'_{n+1}$ that also yields M where $\sigma'_j \in T_I^*$ such that each σ'_j is a minimal explanation of t_{i_j}, i.e.,

$$M_0[\sigma'_1 t_{i_1}\rangle M_{b_1} \cdots [\sigma'_n t_{i_n}\rangle M_{b_n}[\sigma'_{n+1}\rangle M$$

and all M_{b_j}'s are basis markings. By Proposition 1, there necessarily exists a basis marking M_{b_j} such that $M_F \in R_I(M_{b_j})$. Then, by Proposition 5 $M_{b_j} \in \mathcal{F}$, i.e., M_{b_j} is a fully-faulty basis marking. Hence, there exists a path in the BRG that is consistent with σ and passes $\mathcal{F}$.

For (ii), Let $\sigma = \sigma_1 t_{i_1} \cdots \sigma_n t_{i_n} \sigma_{n+1}$ where $\sigma_j \in T_I^*$ and $t_{i_j} \in T_E$. Thus we have the following trajectory:

$$M_0[\sigma_1 t_{i_1}\rangle M_1 \cdots [\sigma_n t_{i_n}\rangle M_n[\sigma_{n+1}\rangle M$$

By the argument of [6], there exists another sequence $\sigma' = \sigma'_1 t_{i_1} \cdots \sigma'_n t_{i_n} \sigma'_{n+1}$ that also yields M where $\sigma'_j \in T_I^*$ such that σ'_j is a minimal explanation of t_{i_j}, i.e.,

$$M_0[\sigma'_1 t_{i_1}\rangle M_{b_1} \cdots [\sigma'_n t_{i_n}\rangle M_{b_n}[\sigma'_{n+1}\rangle M$$

where all M_{b_j}'s are basis markings. Since all transitions $t \in T_I$ has zero-influence on $(\mathbf{w}, k)$, such a rearrangement of transitions does not modify the token count of the trajectories after the firing of each explicit transitions. Hence, all $M_{b_j} \notin \mathcal{F}$. As a result, there exists a path in the BRG that is consistent with σ and does not pass $\mathcal{F}$. □

Proposition 5 shows that any faulty (resp., non-faulty) trajectory of a plant can be associated to a path in the diagnostic BRG that passes at least one faulty basis marking (resp., does not pass any faulty basis marking). Therefore, a diagnostic agent can keep track of the consistent basis markings to determine the passage of set F.

B. Diagnostic Agent Design

Based on Proposition 5, it is not difficult to understand that a correct diagnostic agent $\mathcal{A} : L(G) \to \{0, 1, 2\}$ can be presented as the following:

$$\mathcal{A}(w) = \begin{cases} 0, & \text{if all consistent paths of } w \text{ do not pass } \mathcal{F} \\ 2, & \text{if all consistent paths of } w \text{ pass } \mathcal{F} \\ 1, & \text{otherwise} \end{cases}$$

Such an agent can be presented as a closed-form structure called the *basis diagnoser* as follows. The construction of a basis diagnoser can be done according to the design of the conventional *diagnoser automata* in [1] which is omitted here due to the limit of space.

Definition 10: Given a plant LPN $G = (N, M_0, E, \ell)$, a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, and the diagnostic BRG $\mathcal{B}_d = (\mathcal{M}, Tr, \Delta, M_0)$, let $(M_b, \gamma) \in \mathcal{M} \times \{0, 2\}$ be a pair. The *next* function $Next : \mathcal{M} \times \{0, 2\} \times E \to 2^{\mathcal{M} \times \{0,2\}}$ is defined as a set $Next((M_b, \gamma), e)$ such that: for all arcs in the BRG

$$M_b \xrightarrow{(t,\mathbf{y})} M'_b$$

with $\ell(t) = e$, $(M'_b, \gamma') \in Next((M_b, \gamma), e)$, where $\gamma' = 0$ if $M_b \notin \mathcal{F} \wedge M'_b \notin \mathcal{F}$, otherwise $\gamma' = 2$. Specifically, we define $Next((M_b, \gamma), \varepsilon) = \{(M_b, \gamma)\}$. □

.

Definition 11: Given a plant LPN $G = (N, M_0, E, \ell)$, a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, and the diagnostic BRG $\mathcal{B}_d = (\mathcal{M}, Tr, \Delta, M_0)$, the corresponding *basis diagnoser* is a deterministic finite automaton $\mathcal{D} = (D, E, \delta, d_0)$ where (1) $D \subseteq \mathcal{M} \times \{0, 2\}$ is a set of states, (2) E is the set of events, (3) $d_0 = Next((M_0, 0), \varepsilon)$ is the initial state, and (4) $\delta : D \times E \to D$ is the transition function: $\delta(d, e) = \bigcup_{(M_b,\gamma) \in d} Next((M_b, \gamma), e)$. □

An algorithm to compute the basis diagnoser from a diagnostic BRG can be designed according to Definition 10. In a basis diagnoser, we say state $d \in D$ is

- *normal*, if for all $(M, \gamma) \in d$, $\gamma = 0$;
- *faulty*, if for all $(M, \gamma) \in d$, $\gamma = 2$;
- *ambiguous*, if there exist $(M, \gamma), (M', \gamma') \in d$ with $\gamma = 0$ and $\gamma' = 2$, respectively.

The following theorem shows that a basis diagnoser can be used to solve the marking diagnosis problem.

Theorem 1: Given a plant LPN $G = (N, M_0, E, \ell)$, a set of faulty markings $F = \mathcal{L}_{(\mathbf{w},k)}$, and the basis diagnoser $\mathcal{D} = (D, E, \delta, d_0)$, the following diagnostic function is correct:

$$\mathcal{A}(w) = \begin{cases} 0, & \text{if } d \text{ is normal;} \\ 1, & \text{if } d \text{ is ambiguous;} \\ 2, & \text{if } d \text{ is faulty.} \end{cases}$$

where $d = \delta^*(d_0, w)$ is the consistent state in the basis diagnoser.

Proof: Straightforward from Proposition 5. □

C. Illustrative Example

Consider the LPN in Figure 2 (a) with a set of faulty markings $F = \{M \mid M(p_2) + 2M(p_3) \geq 2\}$. Although $\pi = (T_o, T_{uo})$ is a valid basis partition, the corresponding BRG has 6 basis markings (not drawn), two of which are partially-faulty. As a result, such a BRG cannot be used for

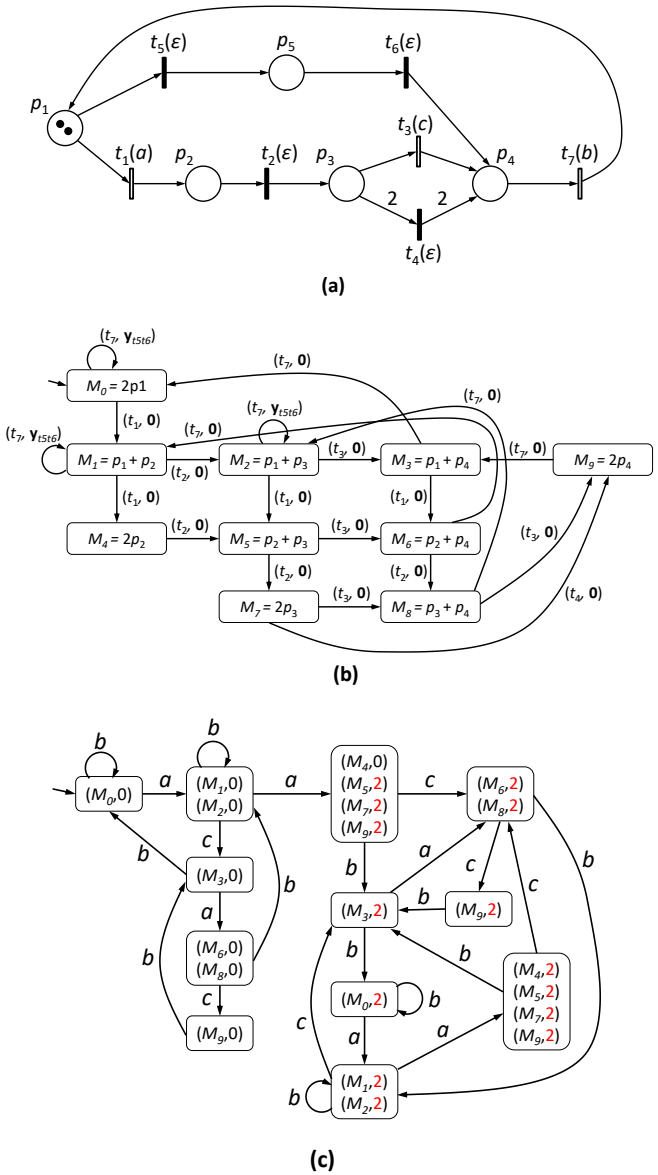

Fig. 2: (a) the LPN used in Section V-c, (b) its diagnostic BRG, (c) its basis diagnoser.

marking diagnosis in this case. On the other hand, since $T_o = \{t_1, t_3, t_7\}$ and the transitions with non-zero influence are t_2 and t_4, the diagnostic basis partition is $\pi_d = (T_E, T_I)$ with $T_E = \{t_1, t_2, t_3, t_4, t_7\}$. The corresponding diagnostic BRG and the basis diagnoser are depicted in Figure 2 (b) and (c), respectively.

Consider observation aa. Since in the basis diagnoser it holds $\delta^*(d_0, aa) = \{(M_4, 0), (M_5, 2), (M_7, 2), (M_9, 2)\}$, the output of the diagnostic agent is $\mathcal{A}(aa) = 1$, which means that the plant may or may have not passed some faulty markings. In fact, two trajectories $M_0[t_1 t_1\rangle$ and $M_0[t_1 t_1 t_2\rangle$ coincide with observation aa while the former does not pass F and the latter passes F. On the other hand, consider observation aac such that $\delta^*(d_0, aac) = \{(M_3, 2)\}$, the diagnostic output is $\mathcal{A}(aac) = 2$, which means that the plant must have passed some faulty markings. In fact, the possible firing sequences that coincide with observation aac are: $t_1 t_1 t_2 t_3$, $t_1 t_2 t_1 t_3$, $t_1 t_2 t_1 t_2 t_3$, and $t_1 t_1 t_2 t_2 t_3$, all of whose trajectories pass F.

The structure complexity of a basis diagnoser is $2^{|\mathcal{M}|}$ where $\mathcal{M}$ is the set of basis markings in the diagnostic BRG. Since the number of basis markings is generally much smaller than the number of reachable markings, the proposed method for marking diagnosability verification in LPNs is of efficiency.

VI. Conclusion

We proposed a method for marking diagnosis in labeled Petri nets using basis reachability graphs. By properly selecting a set of explicit transitions, a particular BRG called the *diagnostic BRG* is computed. Then we have develop an algorithm based on the *basis diagnoser*.

References

[1] S. Lafortune, F. Lin, and C. Hadjicostis, "On the history of diagnosability and opacity in discrete event systems," *Annual Reviews in Control*, vol. 45, pp. 257–266, 2018.

[2] D. Lefebvre, "Fault diagnosis and prognosis with partially observed Petri nets," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 10, pp. 1413–1424, 2014.

[3] X. Yin and S. Lafortune, "Synthesis of maximally-permissive supervisors for the range control problem," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3914–3929, 2017.

[4] F. Basile, M. P. Cabasino, and C. Seatzu, "State estimation and fault diagnosis of labeled time Petri net systems with unobservable transitions," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 997–1009, 2015.

[5] N. Ran, A. Giua, and C. Seatzu, "Enforcement of diagnosability in labeled Petri nets via optimal sensor selection," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2997–3004, July 2019.

[6] M. Cabasino, A. Giua, and C. Seatzu, "Fault detection for discrete event systems using Petri nets with unobservable transitions," *Automatica*, vol. 46, no. 9, pp. 1531–1539, 2010.

[7] R. Kumar and S. Takai, "Decentralized prognosis of failures in discrete event systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 48–59, 2010.

[8] Z. Y. Ma, Y. Tong, Z. W. Li, and A. Giua, "Basis marking representation of Petri net reachability spaces and its application to the reachability problem," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1078–1093, 2017.

[9] Y. Tong, Z. W. Li, C. Seatzu, and A. Giua, "Verification of state-based opacity using Petri nets," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2823–2837, 2017.

[10] N. Ran, H. Su, A. Giua, and C. Seatzu, "Codiagnosability analysis of bounded Petri nets," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 1192–1199, 2018.

[11] Z. Ma, G. Zhu, and Z. Li, "Marking estimation in petri nets using hierarchical basis reachability graphs," *IEEE Transactions on Automatic Control*, p. Early Access. DOI: 10.1109/TAC.2020.2983088, 2021.

[12] H. Lan, Y. Tong, J. Guo, and C. Seatzu, "Verification of C-detectability using Petri nets," *Information Sciences*, vol. 528, pp. 294–310, 2020.

[13] M. P. Cabasino, A. Giua, and C. Seatzu, "Diagnosability of discrete-event systems using labeled Petri nets," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 1, pp. 144–153, Jan 2014.

[14] A. Giua, F. DiCesare, and M. Silva, "Generalized mutual exclusion constraints for Petri nets with uncontrollable transitions," in *Proceedings of the IEEE Int. Conf. on Systems, Man, and Cybernetics*, Chicago, USA, 1992, pp. 947–949.

[15] M. Cabasino, A. Giua, M. Pocci, and C. Seatzu, "Discrete event diagnosis using labeled Petri nets: an application to manufacturing systems," *Control Engineering Practice*, vol. 19, no. 9, pp. 989–1001, 2011.