Automatica 82 (2017) 35–41

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Fault prognosis of timed stochastic discrete event systems with bounded estimation error☆

Rabah Ammour[a], Edouard Leclercq[a], Eric Sanlaville[b], Dimitri Lefebvre[a]

[a] Normandie Univ, UNIHAVRE, GREAH, 76600 Le Havre, France
[b] Normandie Univ, UNIHAVRE, LITIS, 76600 Le Havre, France

ARTICLE INFO

Article history:
Received 27 April 2016
Received in revised form
18 October 2016
Accepted 5 March 2017

Keywords:
Discrete-event systems
Fault prognosis
Stochastic Petri nets

ABSTRACT

This article deals with the problem of fault prognosis in timed stochastic discrete event systems. For that purpose, partially observed stochastic Petri nets are considered to model the system with its sensors. The model represents both healthy and faulty behaviors of the system. Using a timed measurement sequence issued from the sensors, an approach denoted (ρ, δ)*-prognosis* is proposed to estimate the probability of a future fault occurrence. The method is based on two input parameters: the error bound ρ and the prognosis horizon δ. The main contribution is to bound the estimation error by ρ when the prognosis horizon does not exceed δ. An example is presented to illustrate the results.

1. Introduction

Many systems can be seen as Discrete Event Systems (DESs). Their state evolution depends entirely on the occurrence of discrete events (Cassandras & Lafortune, 2008). For such systems, faults prognosis, which consists in predicting a fault event before its occurrence, is a main challenge. The objective is to anticipate a fault occurrence in order to take any corrective actions in advance.

In the literature, fault prognosis of DESs has received considerable attention. In untimed context, Genc and Lafortune (2009) is one of the pioneer works that addresses the problem of event prediction (or prognosis) for partially observed DESs. Finite automata models, where some events are observable, are used to model the system. The notion of *predictability* was formulated as the capability to deduce future faults occurrences based on actual measurement records. Its principle is inspired from *diagnosability* of DESs introduced in Sampath, Sengupta, Lafortune, Sinnamohideen, and Teneketzis (1995). In Jéron, Marchand, Genc, and Lafortune (2008), predictability concerns sequence patterns rather than a single event. In Kumar and Takai (2010), the authors consider fault prognosis in decentralized settings, such that a global prognosis decision is computed from local decisions. Recent works on fault prognosis focus on some properties such as robustness (Takai, 2015), and on guaranteed performance bound of decentralized fault prognosis (Yin & Li, 2016). In timed context, predicting event occurrences for partially observed timed automata has been investigated in Cassez and Grastien (2013). It shows that the explicit consideration of time is crucial for fault prognosis of DESs. Indeed, it gives for example the remaining time before the occurrence of a fault event, in order to stop or to reconfigure the system.

Note that the deterministic methods are rather rigid since they only consider the case where the fault occurs without ambiguity. Recently, to overcome this binary analysis, the authors of Chen and Kumar (2015) and Nouioua, Dague, and Ye (2014) studied stochastic failure prognosability of DESs monitored with probabilistic automaton. In Chen and Kumar (2015) the notion of S_m*-Prognosability* or *m-steps Stochastic Prognosability,* which is the ability to predict a fault at least m-steps prior to its occurrence, is proposed in untimed settings.

In this work, we investigate the problem of faults prognosis in timed stochastic DESs modeled by Partially Observed Stochastic Petri Nets (POSPNs). It consists in computing, at a given time, the probability that the system behavior becomes faulty in the future time. Given two input parameters, an error bound ρ and a prognosis time horizon δ, the objective of the present study is to estimate the probability that a fault will occur within δ, and to show that the estimation error is lower than ρ. This estimation is performed online during the system evolution. To this end, the $(\rho,\ \delta)$*-prognosis* method is introduced.

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Christoforos Hadjicostis under the direction of Editor Christos G. Cassandras.

E-mail addresses: rabah.ammour@univ-lehavre.fr (R. Ammour), edouard.leclercq@univ-lehavre.fr (E. Leclercq), eric.sanlaville@univ-lehavre.fr (E. Sanlaville), dimitri.lefebvre@univ-lehavre.fr (D. Lefebvre).

http://dx.doi.org/10.1016/j.automatica.2017.04.028
0005-1098/

The problem of faults prognosis in stochastic DESs has been introduced in our previous works (Ammour, Leclercq, Sanlaville, & Lefebvre, 2016a; Lefebvre, 2014). Compared to the approach presented here, only the untimed context for the prognosis issue was considered in Lefebvre (2014). Indeed, the method is based on the probability of firing a specific transition before any other transition. In Ammour et al. (2016a), prognosis in timed context has been introduced. But only possible continuations (future behaviors), with limited size l_{max} from the current markings, were considered for fault prognosis regardless the prognosis time horizon. Consequently, no guarantee on the estimation error was provided since the choice of l_{max} is arbitrary and does not depend on the prognosis horizon. Indeed, a small value of l_{max} saves computational efforts but implies a limited prognosis accuracy (with respect to (wrt) the prognosis horizon). On the other hand, a large value of l_{max} leads to high computational efforts without improving necessarily the accuracy of the prognosis. The key idea behind the $(\rho,\ \delta)$-*prognosis* approach introduced here is, from a state estimation consistent with the measurements, to determine the appropriate bounded set of the most probable continuations in order to make possible the estimation of the future fault probability and, subsequently, to show that this estimation fulfills a bounded estimation error. To the best of our knowledge, faults prognosis based on POSPNs was rarely explored despite this formalism considers both partial measurements of markings and events.

The remainder of this paper is structured as follows. In Section 2, POSPNs and timed observation sequences are described. Thereafter, the stochastic fault prognosis is studied in Section 3 where the $(\rho,\ \delta)$-*prognosis* approach is detailed. Section 4 presents some conclusions and perspectives.

2. Preliminaries

2.1. Partially observed stochastic Petri nets

Let $G_s = \langle \boldsymbol{P}, \boldsymbol{T}, W_{PR}, W_{PO}, \mu \rangle$ be a Stochastic Petri Net (SPN) structure, where $\boldsymbol{P} = \{P_1, \ldots, P_n\}$ is a set of n places and $\boldsymbol{T} = \{T_1, \ldots, T_q\}$ is a set of q transitions. $W_{PO} \in (\mathbb{N})^{n\times q}$ and $W_{PR} \in (\mathbb{N})^{n\times q}$ are the post and pre incidence matrices and $W = W_{PO} - W_{PR}$ is the incidence matrix. SPNs are characterized by random firing delays associated with the transitions (Molloy, 1982). $\mu = (\mu_j) \in (\mathbb{R}^+)^q$ is the firing rate vector which characterizes the transition firing periods. $\langle G_s,\ M_I\rangle$ is a SPN model with initial marking M_I and $M \in (\mathbb{N})^n$ represents the SPN marking vector. A transition T_j is enabled at marking M if and only if (iff) $M \geq W_{PR}(:, j)$, where $W_{PR}(:, j)$ is the column j of the pre incidence matrix. One writes $M\,[T_j\rangle$ to denote that the transition T_j may fire from the marking M. For each transition T_j, the firing periods are given, at marking M, by a random variable (rv) with an exponential probability density function (pdf). The parameter of the rv is $n_j(M)\cdot\mu_j$ where $n_j\,(M)$ stands for the enabled degree of transition T_j at marking M. It is given by $n_j\,(M) = \min\left\{\left\lfloor\left(\frac{m_k}{w_{PR}(k,j)}\right)\right\rfloor,\ P_k \in {}^\circ T_j\right\}$ where ${}^\circ T_j$ is the set of T_j upstream places denoted by P_k and m_k their markings. Finally, $\lfloor\cdot\rfloor$ stands for the lower rounded value of $(\cdot)$. When T_j fires once, the marking varies according to $\Delta M = M' - M = W(:, j)$. This is denoted by $M[T_j\rangle\, M'$. An untimed firing sequence σ_U of size $h = |\sigma_U|$ fired at marking $M\,(0)$ is a sequence of h transitions $\sigma_U = T(1)T(2)\ldots T(h)$, with $T\,(j) \in \boldsymbol{T}, j = 1, \ldots, h$ that consecutively fire from $M\,(0)$. This leads to the untimed marking trajectory denoted by $(\sigma_U,\ M\,(0))$. The probability of the untimed trajectory $(\sigma_U, M\,(0))$ is given by:

$$P\,(\sigma_U, M\,(0)) = \prod_{k=1\ldots h}\left(\frac{n_k\,(M\,(k-1))\ \mu_k}{\sum\limits_{T_j\in\boldsymbol{T}} n_j\,(M\,(k-1))\ \mu_j}\right). \tag{1}$$

The integer $x_j(\sigma_U)$ is the number of occurrences of the transition T_j in σ_U, and $X(\sigma_U) = (x_j(\sigma_U)) \in (\mathbb{N})^q$ is the firing count vector of σ_U. When time is considered, a timed firing sequence will simply be denoted by σ. When the sequence is fired at marking $M\,(t_0)$ in time interval $[t_0\ t_h]$, σ is defined in a similar way: $\sigma = T(t_1)T(t_2)\ldots T(t_h)$ where $t_j, j = 1, \ldots, h$ represent the firing dates of the transitions $T(t_j) \in \boldsymbol{T}$ that satisfy $t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_h$. The timed marking trajectory is then denoted by $(\sigma, M\,(t_0))$ and $M\,(t_0)\ [\sigma\rangle$ denotes that σ may fire from the marking $M\,(t_0)$. A labeling function $\mathcal{L} : T \to E \cup \{\varepsilon\}$ is introduced to assign a label to each transition where $E = \{e_1, \ldots, e_{q_o}\}$ is the set of q_o labels assigned to observable transitions and ε is the null label assigned to the silent ones. A marking sensor matrix $H \in (\mathbb{R})^{n_o\times n}$ is also introduced to define the projection of the marking vector M over a subset of measured places of dimension n_o. The measured part of the marking is denoted as $M_o = H \cdot M$ (Lefebvre, 2014). A POSPN with initial marking M_I is defined as $\langle G_s, \mathcal{L}, H, M_I\rangle$ where G_s is the SPN structure, $\mathcal{L}$ is the event sensor function and H the marking sensor matrix. Finally, $\mathcal{F} = \{f_1, \ldots, f_s\}$ is the set of s fault classes and the row vector $F_\alpha = (f_{\alpha j}) \in (\mathbb{N})^{1\times q}$ assigns the fault class f_α to some transitions such that $f_{\alpha j} = 1$ if T_j represents a fault of class f_α else $f_{\alpha j} = 0$. Transitions without any fault class are assumed to be healthy and correspond to the expected behaviors.

2.2. Timed measurement sequences

A measurement function Γ is introduced according to the sensor configuration. It collects, using the function L and the matrix H, the K successive dated marking and event measurements of a timed marking trajectory $(\sigma,\ M(\tau_0))$ over the time horizon $[\tau_0,\ \tau_K]$ (where τ_K is the date of the last measurement), and organizes them in the timed measurement sequence (2):

$$\begin{aligned}\Gamma\,(\sigma,\ M) = {} & (M_{O0}, \tau_0)\, e_O\,(\tau_1)\,(M_{O1}, \tau_1)\\ & \ldots e_O\,(\tau_K)\,(M_{OK}, \tau_K)\end{aligned} \tag{2}$$

where $M_{O0} = H\cdot M(\tau_0), K$ is the length of the measurement sequence and $\tau_j,\ j = 1, \ldots, K$ refer to the dates of the measurements. Given a timed measurement sequence in $[\tau_0,\ \tau_K]$ of form (2) denoted by TR_O, a marking trajectory (σ, M), that satisfies $\Gamma\,(\sigma,\ M) = TR_O$, is said to be consistent with TR_O. The set of all timed consistent trajectories wrt the measurement sequence TR_O is denoted by $\Gamma^{-1}\ (TR_O)$. Assumption A is considered in order to avoid consistent firing sequences with infinite number of events:

Assumption A. The unobservable part of the reachability graph is acyclic.

Assumption A is commonly adopted in the field of fault detection with Petri net models. It can be ensured by the sensor configuration. It guarantees that the size of unobservable firing sequences that do not provide any measurement is bounded.

The set $\Gamma^{-1}\ (TR_O)$ of timed consistent trajectories is obtained in two steps (Lefebvre, 2014); the first one is to get the set of untimed trajectories that are consistent with the measurements by solving linear matrix inequalities (LMI). The second step consists on adding time constraints issued from the dates of the measurements, in order to obtain timed trajectories consistent with the measurements.

2.3. Sum of exponential random variables

Consider a set of n independent rvs $X_i,\ i = 1, \ldots, n$ having exponential pdfs with parameters $\lambda_i, i = 1, \ldots, n$ respectively. Let us denote by S_n the sum of these rvs: $S_n = \sum_{i=1}^n X_i$. In the case where all the parameters λ_i are equal, the rv S_n