Stochastic Failure Prognosis of Discrete Event Systems

Jun Chen, *Senior Member, IEEE*, and Ratnesh Kumar, *Fellow, IEEE*

***Abstract*—This article studies the prognosis of failure, i.e., its prediction prior to its occurrence, in stochastic discrete event systems. Prior work has focused on the definition and offline verification of m-steps stochastic-prognosability, or S_m-prognosability, which allows the prediction of a fault at least m-steps in advance. This article complements the existing work by proposing an algorithm for the computation of online failure prognoser. The proposed algorithm reduces the condition for issuing an affirmative prognostic decision to verification condition of a safety property of a Markov chain. We discuss how such a verification condition can be computed using a finitely terminating algorithm.**

***Index Terms*—Discrete event systems (DESs), failure prognosis, fixed-point computation, state distribution, stochastic systems.**

I. Introduction

The problem of predicting a fault prior to its occurrence is a well-researched area (see for example [1]). The failure prognosis problem has been widely studied in the context of discrete event systems (DESs) [2]–[11]). The notion of uniformly bounded prognosability of fault was formulated in [2] for logical DESs, requiring each fault trace possesses a nonfault prefix (termed an *indicator*) such that for all indistinguishable traces, a future fault is inevitable within a bounded delay that is uniform across all fault traces.

Kumar and Takai [4] removed the requirement of the existence of a uniform bound, provided a computable online prognoser, and further established that the notion of prognosability is equivalent to the existence of a prognoser with no false alarm (FA) and no missed detection (MD). The issue of prognosability under a general decentralized inferencing mechanism was proposed in [5], where a prognostic decision involved inferencing among a group of local prognosers over their local decisions and their ambiguity levels, and the notion of inference-prognosability and its verification was introduced to capture the necessity and sufficiency of inferencing-based decentralized prognosis. The problem of distributed prognosability under bounded-delay communications among the local prognosers was studied in [8], where the notion of joint-prognosability and its verification was proposed. To account for model uncertainty, robust prognosability with respect to a set of system models was studied in [6]. The notion of failure prognosis was later extended to labeled Petri net in [12]–[16]. Finally, Orchard and Vachtsevanos [17] discuss the potential application of particle filtering in continuous systems.

Failure prognosis has also been studied for stochastic DESs [10], [18]. In [10], the notion of m-steps stochastic-prognosability, or simply S_m-prognosability, was introduced, which requires for any tolerance level ρ and error bound τ, there exists a reaction bound $k \geq m$, such that the set of fault traces for which a fault cannot be predicted k steps in advance with tolerance level ρ occurs with probability smaller than τ. In [10], we further showed that S_m-prognosability is a necessary and sufficient condition for the existence of a prognoser with reaction bound at least m (i.e., prediction at least m-steps prior to the occurrence of a fault) that can achieve any specified FA[1] and MD rate requirements.

The prognoser formalized in [10] requires the calculation of the least probability of no-fault over all finite future steps, and this calculation involves matrix multiplication for arbitrary number of times. Hence, the prognoser of Chen and Kumar [10] cannot be implemented. In this article, we address this limitation by proposing an algorithm for the computation of online failure prognoser, which reduces the online prognosis problem to a control problem of Markov chain under linear safety constraint [19]–[21]. In particular, for a given threshold ρ, the algorithm computes the set of state distributions from which the probability of future fault is less than ρ, denoted as maximal initial nonfault set (MINS). The algorithm starts by setting MINS as the set of stationary state distributions that a fault cannot be detected, and iteratively sets MINS to the set of state distributions π so that either π is already in MINS in previous iteration or its one step successor falls in MINS of previous step. The algorithm repeats until a fixed point is reached, and such termination is proven to be guaranteed. Once MINS is computed, given an observation, the online failure prognosis problem reduces to checking whether the conditional state distribution resulting from that observation is in MINS or not. The online prognoser issues a failure prediction decision "F" if the conditional state distribution is not a member of MINS, and "no-decision" otherwise.

The rest of this article is organized as follows. The notations and some preliminaries are presented in Section II, followed by a brief review of S_m-prognosability and stochastic prognoser in Section III. Section IV gives the main result, an algorithm for the implementation of the prognoser for online prognosis, whose termination is guaranteed. Finally, Section V concludes this article.

II. Notations and Preliminaries

For an event set Σ, define $\overline{\Sigma} := \Sigma \cup \{\epsilon\}$, where ϵ denotes "no-event." The set of all finite-length event sequences over Σ, including ϵ, is denoted as Σ^*. A *trace* is a member of Σ^* and a *language* is a subset of Σ^*. We use $s \leq t$ to denote that $s \in \Sigma^*$ is a prefix of $t \in \Sigma^*$, $pr(s)$ to denote the set of all prefixes of s, and $|s|$ to denote the length of s or the number of events in s. For $n \in \mathbb{N}$ where $\mathbb{N}$ denotes the set of all nonnegative integers, define $\Sigma^{<n} := \{s \in \Sigma^* : |s| < n\}$. Similarly define $\Sigma^{\leq n}$, $\Sigma^{>n}$, $\Sigma^{\geq n}$, and $\Sigma^{=n}$ according to $\Sigma^{\sim n} := \{s \in \Sigma^* : |s| \sim n\}$ where $\sim$ can be one of $\sim \in \{<, \leq, >, \geq, =\}$. Note that $\Sigma^{=n}$ is also denoted as Σ^n for simplicity. For $L \subseteq \Sigma^*$, its prefix-closure is defined

[1] FA may also be termed as Type I error, whereas MD is Type II error when the null hypothesis is "there is no future fault."

Manuscript received 9 November 2016; revised 19 August 2021; accepted 3 October 2021. Date of publication 7 October 2021; date of current version 27 September 2022. This work was supported in part by the National Science Foundation under Grant ECCS-1509420 and Grant CSSI-2004766. Recommended by Associate Editor Carla Seatzu. *(Corresponding author: Jun Chen.)*

Jun Chen is with the Department of Electrical and Computer Engineering, Oakland University, Rochester, MI 48309 USA (e-mail: junchen@oakland.edu).

Ratnesh Kumar is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: rkumar@iastate.edu).

Digital Object Identifier 10.1109/TAC.2021.3118670

as $pr(L) := \bigcup_{s\in L} pr(s)$, and L is said to be prefix-closed (or simply closed) if $pr(L) = L$. Given two languages L_1 and L_2, their *concatenation* is defined as $L_1L_2 := \{st : s \in L_1, t \in L_2\}$, the set of traces in L_1 *after* L_2 is defined as $L_1\backslash L_2 := \{t \in \Sigma^* : \exists s \in L_2, st \in L_1\}$, and the set of traces in L_1 *quotient* L_2 is defined as $L_1/L_2 := \{s \in pr(L_1) : \exists t \in L_2, st \in L_1\}$.

A stochastic DES can be modeled by a *stochastic automaton* $G = (X, \Sigma, \alpha, x_0)$, where X is the set of states, Σ is the set of events, $x_0 \in X$ is the initial state, and $\alpha : X \times \Sigma \times X \rightarrow [0, 1]$ is the transition probability function [22] satisfying $\forall x \in X, \sum_{\sigma\in\Sigma}\sum_{x'\in X} \alpha(x, \sigma, x') = 1$, i.e., there is no "termination" at any of the states (note there is no loss of generality in assuming no termination, since otherwise, one can augment the model with a newly introduced "termination-state," and transitions from each state to the termination state on a newly introduced "termination-event" that is unobservable and whose occurrence probability equals the probability of termination of the said state). G is nonstochastic if $\alpha : X \times \Sigma \times X \rightarrow \{0, 1\}$, and a nonstochastic DES is deterministic if $\forall x \in X, \sigma \in \Sigma, \sum_{x'\in X} \alpha(x, \sigma, x') \in \{0, 1\}$, i.e., each state has at most one transition on each event. The transition probability function α can be generalized to $\alpha : X \times \Sigma^* \times X$ in a natural way by multiplying the probabilities of the individual transitions. Define the language generated by G as $L(G) := \{s \in \Sigma^* : \exists x \in X, \alpha(x_0, s, x) > 0\}$. A *component* $C = (X_C, \alpha_C)$ of G is a "subgraph" of G, i.e., $X_C \subseteq X$ and $\forall x, x' \in X_C$ and $\sigma \in \Sigma$, $\alpha_C(x, \sigma, x') := \alpha(x, \sigma, x')$, whenever the latter is defined. C is said to be a *strongly connected component* (SCC) or *irreducible* if $\forall x, x' \in X_C$, $\exists s \in \Sigma^*$ such that $\alpha_C(x, s, x') > 0$. An SCC C is said to be *closed* if for each $x \in X_C$, $\sum_{\sigma\in\Sigma}\sum_{x'\in X_C} \alpha_C(x, \sigma, x') = 1$.

To represent the limited sensing capabilities of a prognoser, we introduce an event observation mask, $M : \overline{\Sigma} \rightarrow \overline{\Delta}$, where Δ is the set of observed symbols and $M(\epsilon) = \epsilon$. An event σ is *unobservable* if $M(\sigma) = \epsilon$. The set of unobservable events is denoted as Σ_{uo}, and so the set of observable events is given by $\Sigma - \Sigma_{uo}$. The observation mask can be generalized to $M : 2^{\Sigma^*} \rightarrow 2^{\Delta^*}$ in a natural way: $\forall s \in \Sigma^*, \sigma \in \overline{\Sigma}, L \subseteq \Sigma^*, M(\epsilon) = \epsilon, M(s\sigma) = M(s)M(\sigma)$, and $M(L) = \{M(s) : s \in L\}$.

For a stochastic automaton $G = (X, \Sigma, \alpha, x_0)$ with generated language $L(G) = L$, let $K \subseteq L$ be a nonempty closed sublanguage representing a nonfault specification for G, i.e., $L - K$ consists of behaviors that execute some fault. Then, the task of prognosis is to predict with sufficient confidence the execution of any trace in $L - K$ prior to its execution. Let $K \subseteq L$ be generated by a *deterministic* automaton $R = (Q, \Sigma, \beta, q)$ such that $L(R) = K$ (from now on we interchangeably use K and R to refer to the "nonfault specification"). Then, the refinement of the plant with respect to the specification, denoted as G^R, can be used to capture the fault traces in the form of the reachability of a fault state carrying the label F in G^R, which is given by $G^R := (X \times \overline{Q}, \Sigma, \gamma, (x_0, q_0))$, where $\overline{Q} = Q \cup \{F\}$, and $\forall (x, \overline{q}), (x,' \overline{q}') \in X \times \overline{Q}, \sigma \in \Sigma, \gamma((x, \overline{q}), \sigma, (x,' \overline{q}')) = \alpha(x, \sigma, x')$ if the following holds:

$$(\overline{q}, \overline{q}' \in Q \wedge \beta(\overline{q}, \sigma, \overline{q}') > 0) \vee (\overline{q} = \overline{q}' = F) \vee \left(\overline{q}' = F \wedge \sum_{q\in Q} \beta(\overline{q}, \sigma, q) = 0\right)$$

and otherwise $\gamma((x, \overline{q}), \sigma, (x,' \overline{q}')) = 0$. Then, it can be seen that the refined plant G^R has the following properties: (1) $L(G^R) = L(G) = L$; (2) any fault trace $s \in L - K$ transitions the refinement G^R to a fault state (a state containing F as its second coordinate); and (3) the occurrence probability of each trace in G^R is the same as that in G, i.e., $\sum_{x\in X} \alpha(x_0, s, x) = \sum_{(x,\overline{q})\in X\times\overline{Q}} \gamma((x_0, q_0), s, (x, \overline{q}))$.

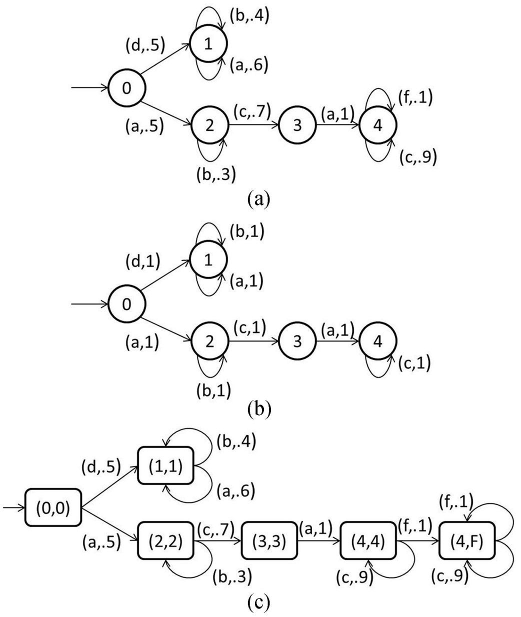

Fig. 1. (a) Stochastic automaton G. (b) Nonfault specification R. (c) Refinement G^R.

Example 1: Fig. 1(a) is an example of a stochastic automaton G. The set of states is $X = \{0, 1, 2, 3, 4\}$ with initial state $x_0 = 0$, and event set $\Sigma = \{a, b, c, d, f\}$. A state is depicted as a node, whereas a transition is depicted as an edge between its origin and termination states, with its event name and probability value labeled on the edge. The observation mask M is such that $M(\{d, f\}) = \{\epsilon\}$ and $M(\sigma) = \sigma$ for $\sigma \in \Sigma - \{d, f\}$. The nonfault specification is given in Fig. 1(b). Therefore, $L - K = \{ab^*cac^*f\}\Sigma^* \cap L$ and the refinement G^R is shown in Fig. 1(c). As can be seen, all traces in $L - K$ transitions G^R to the only fault state $(4, F)$. In G^R, there are two closed SCCs, one is formed by the nonfault state (1,1) and its self-loop transitions, whereas the other is formed by the fault state $(4, F)$ and its selfloop transitions. ■

For $x_i, x_j \in X$ and $\sigma \in \Sigma - \Sigma_{uo}$, define the set of traces originating at x_i, terminating at x_j, and executing a sequence of unobservable events followed by a single observable event σ as $L_G(x_i, \sigma, x_j) := \{s \in \Sigma^* : s = u\sigma, M(u) = \epsilon, \alpha(x_i, s, x_j) > 0\}$. Define $\alpha(L_G(x_i, \sigma, x_j)) := \sum_{s\in L_G(x_i,\sigma,x_j)} \alpha(x_i, s, x_j)$ as the occurrence probability of traces in $L_G(x_i, \sigma, x_j)$ and denote it as $\mu_{i,\sigma,j}$ for short. Also define $\lambda_{ij} = \sum_{\sigma\in\Sigma_{uo}} \alpha(x_i, \sigma, x_j)$ as the probability of transitioning from x_i to x_j while executing a single unobservable event. Then, it can be seen that $\mu_{i,\sigma,j} = \sum_k \lambda_{ik}\mu_{k,\sigma,j} + \alpha(x_i, \sigma, x_j)$, where the first term on the right-hand side (RHS) involves transitioning in at least two steps via some intermediate state, whereas the second RHS term involves transitioning directly in exactly one step. Thus, for each $\sigma \in \Sigma - \Sigma_{uo}$, given the values $\{\lambda_{ij} | i, j \in X\}$ and $\{\alpha(x_i, \sigma, x_j) | i, j \in X\}$, all the probabilities $\{\mu_{i,\sigma,j} | i, j \in X, \sigma \in \Sigma - \Sigma_{uo}\}$ can be found by solving the following matrix equation (see for example Wang and Ray [23] for a similar matrix equation):

$$\boldsymbol{\mu}(\sigma) = \boldsymbol{\lambda}\boldsymbol{\mu}(\sigma) + \boldsymbol{\alpha}(\sigma) \qquad (1)$$

where $\boldsymbol{\mu}(\sigma)$, $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}(\sigma)$ are all $|X| \times |X|$ square matrices whose ijth elements are given by $\mu_{i,\sigma,j}$, λ_{ij}, and $\alpha(x_i, \sigma, x_j)$, respectively. In the presence of partial observability, we define $L_G(x_i, M(\sigma), x_j) := \cup_{\sigma'\in\Sigma:M(\sigma')=M(\sigma)} L_G(x_i, \sigma,' x_j)$, i.e., it is the set of all traces originating at x_i, terminating at x_j, and executing a sequence of unobservable events followed by a single observable event that has the same mask value $M(\sigma)$. Then, their occurrence probability is given by $\alpha(L_G(x_i, M(\sigma), x_j)) := \sum_{\sigma' \in \Sigma : M(\sigma')=M(\sigma)} \mu_{i,\sigma,'j}$.

III. Prognosability in Stochastic DES

In this section, we give a brief review of the S_m-*prognosability* and m-*prognoser* for stochastic DESs. These notions are needed to establish the major results of this article. For more details, refer to Chen and Kumar [10].

The following definition introduces the notions of *boundary* fault traces whose all strict prefixes are nonfault, and *persistent* nonfault traces whose all extensions are nonfault.

Definition 1 (see[10]): Given a pair (L, K) of closed languages with $K \subseteq L$, we define the set of the following:

1) *boundary* fault traces as $\partial := \{s \in L - K : pr(s) - \{s\} \subseteq K\}$;
2) *persistent* nonfault traces of K with respect to L as $\aleph := \{s \in K : \forall n \in \mathbb{N}, \{s\}\Sigma^n \cap (L - K) = \emptyset\} = \{s \in K : \forall n \in \mathbb{N}, \Pr(\{s\}\Sigma^n \cap K) = 1\}$.

Define the n-*step prognostic probability of no-fault* following an observation $o \in M(L)$ as

$$P_N^n(o) := \frac{\Pr(\{M^{-1}(o)\}\Sigma^n \cap K)}{\Pr(\{M^{-1}(o)\}\Sigma^n \cap L)} \tag{2}$$

and the *least prognostic probability of no-fault* following $o \in M(L)$ as

$$P_N^*(o) := \min_{n \in \mathbb{N}} P_N^n(o). \tag{3}$$

The following definition of S_m-prognosability is reproduced from Chen and Kumar [10].

Definition 2 (see[10]): A pair (L, K) of closed languages with $K \subseteq L$ is said to be m-*steps stochastically-prognosable*, or simply S_m-*Prognosable*, if

$$(\forall \tau, \rho > 0)(\exists k \geq m)\Pr(s \in \partial : [|s| \leq k] \\ \vee [\forall u \in s/\Sigma^{>k}, P_N^*(M(u)) > \rho]) < \tau \tag{4}$$

where P_N^* is as defined by (2) and (3).

Next theorem from Chen and Kumar [10] shows that S_0-prognosability is stronger than AA-diagnosability that were studied in [24], i.e., whenever it is possible to predict a fault, it is also possible to diagnose it.

Theorem 1: Given a pair (L, K) of closed regular languages with $K \subseteq L$, if (L, K) is S_0-prognosable, then it is AA-diagnosable. However, the converse need not hold.

In order to predict a fault in advance, the prognoser introduced in [10] computes for each $o \in M(L)$, the prognostic probability of no-fault $P_N^*(o)$ as defined by (2)–(3), and compares it with an appropriately chosen threshold ρ. Whenever $P_N^*(o)$ is below this threshold, implying that there is only a small likelihood of no-fault in future, the prognoser issues a fault warning F, predicting/prognosing a future fault, and otherwise it remains silent (issues ϵ). In other words, a prognoser is formally a map $D : M(L) \to \{F, \epsilon\}$ defined as

$$\forall o \in M(L), [D(o) = F] \Leftrightarrow [\exists \bar{o} \leq o : P_N^*(\bar{o}) \leq \rho] \tag{5}$$

where P_N^* is as defined by (2) and (3).

For a prognoser that aims to predict a fault at least m steps before its occurrence, a *miss detection* (MD) occurs when a fault happens while the prognoser fails to issue a warning m steps in advance. On the other hand, an FA occurs when a warning is issued for a trace whose all extensions are nonfault, i.e., a trace in $\aleph$. Therefore, the MD rate P^{md} and the FA rate P^{fa} for a m-prognoser can be defined as

$$P^{\text{md}} = \Pr(s \in \partial : [|s| \leq m] \vee [D(M(s/\Sigma^{m+1})) = \epsilon] \tag{6}$$

$$P^{\text{fa}} = \Pr(s \in \aleph : D(M(s)) = F). \tag{7}$$

Example 2: For the system G^R shown in Fig. 1. Suppose G^R executes $dabbb$ and produces observation $o = abbb$, then $P_N^*(o) = 0.5872$. Hence, for any m-prognoser with threshold $\rho \geq 0.5872$, traces in $\{dabbb\}\Sigma^* \cap L$ will be false alarmed. When G^R executes a trace in $ab^*cac^*f \subseteq \partial$ and produces an observation $o \in ab^*cac^*$, then $P_N^*(\bar{o})$ approaches 0 for any $\bar{o} \in ab^*c$. Therefore, for a one-prognoser with any threshold ρ, all fault traces can be prognosed, and hence no MD. However, for a two-prognoser with $\rho = 0.3$, when G^R executes the fault trace $abcaf$, a prognostic decision can be made only upon observing abc (since for all its prefixes, the threshold remains lower than the prognostic probability of no fault: $P_N^*(\epsilon) = 0.5$, $P_N^*(a) = 0.375$, $P_N^*(ab) = 0.444$, $P_N^*(abc) = 0$), which violates the least reaction bound $m = 2$, and hence $abcaf$ gets missed detected.

Next theorem shows that for regular languages L and K, S_m-prognosability is necessary and sufficient for the existence of a m-prognoser to satisfy any level of FA and MD rates.

Theorem 2 (see[10]): Consider a pair (L, K) of closed regular languages with $K \subseteq L$. Then, for any FA rate $\phi > 0$ and MD rate $\tau > 0$, there exists a m-prognoser (and its associated prognostic decision threshold) defined by (4) such that the MD and FA rates defined by (5)–(6) satisfy $P^{\text{md}} \leq \tau$ and $P^{\text{fa}} \leq \phi$ if and only if (L, K) is S_m-prognosable.

Note that the results presented in Section III are reproduced from Chen and Kumar [10] and serve as preliminary results needed to establish the main results in Section IV. For more details, such as practical examples and intuitive insights, refer to Chen and Kumar [10].

IV. Computation of Statistics for Online Prognosis

Now, we are ready to present our main results.

A. Online Prognosis

The prognoser formalized in (4) requires the calculation of the least prognostic probability of no-fault as defined by (3), which can be further reduced to

$$\begin{aligned} P_N^*(o) &:= \min_{n \in \mathbb{N}} P_N^n(o) \\ &= \min_{n \in \mathbb{N}} \frac{\Pr(\{M^{-1}(o)\}\Sigma^n \cap K)}{\Pr(\{M^{-1}(o)\}\Sigma^n \cap L)} \\ &= \frac{\min_{n \in \mathbb{N}} \Pr(\{M^{-1}(o)\}\Sigma^n \cap K)}{\Pr(\{M^{-1}(o)\} \cap L)}. \end{aligned} \tag{8}$$

Recall that $P_N^n(o)$ is the probability, following the observation o, that the system does not execute a fault in the next n steps; and $P_N^*(o)$ is the least probability, following the observation o, that the system does not execute a fault over all finite-step futures. Note that in the denominator of (2), we used the fact that probability of all extensions of length n, beyond the traces in $M^{-1}(o)$, is the same as the probability of traces in $M^{-1}(o)$, for there is no termination at any of the states. As a result, the denominator is constant with respect to n, and the minimum only applies to the numerator in (3).

The challenge of computing (3) or (7) lies in the fact that one needs to apply the min operator over all finite future steps, which cannot be implemented straightforwardly using (3) or (7). Therefore, this calculation requires a computable approach that remains open. In this section, we present the verification of $P_N^*(o) \leq \rho$ for any observation $o \in M(L)$ by reducing it to a verification problem of Markov chain under a safety constraint [19]–[21]. In particular, for a given threshold ρ, we compute the set of state distributions from which the probability of future fault remains less than ρ, denoted as MINS. Note that this is done offline. Then, the online failure prognosis problem, as formalized in (4), reduces to checking, given an observation, whether the conditional state distribution following that observation belongs to MINS or not. The online prognoser issues a prognosis decision of "F" if the conditional state distribution is not a member of MINS, and "no-decision" otherwise. We will also show that for a given threshold $\rho \in [0, 1]$, the computation of corresponding MINS is guaranteed to terminate.

Given a refined plant model G^R with state space $Y \subseteq X \times \overline{Q}$, its embedded Markov chain can be obtained by reducing the event information associated with the transition, i.e., the Markov chain has state space Y and transition matrix Ω_{G^R}, which is a $|Y| \times |Y|$ square matrix with ijth entry given by $\Omega_{ij} = \sum_{\sigma\in\Sigma} \gamma(y_i, \sigma, y_j)$ (note that the Markov chain contains *at most one* transition between a pair of states *in each direction* and *does not carry* an event label). Let Π denote the set of probability distributions on the state space of G^R, i.e., each element $\pi \in \Pi$ is a vector with $|Y|$ nonnegative elements and $\|\pi\| = 1$, where $\|\cdot\|$ is simply the sum of all vector elements. A state distribution $\pi^* \in \Pi$ is said to be a *stationary state distribution* of Ω_{G^R} if $\pi^*\Omega_{G^R} = \pi^*$ and $\Pi_s \subseteq \Pi$ is said to be a *stationary set of state distributions* of Ω_{G^R} if $\pi \in \Pi_s \Rightarrow \pi\Omega_{G^R} \in \Pi_s$, or equivalently, $\Pi_s\Omega_{G^R} \subseteq \Pi_s$.

Given current observation o, define the *current state distribution mapping* $\pi : M(L) \to \Pi$, which is the state distribution conditioned upon the observation o, and can be recursively computed as [25]: $\pi(\epsilon) = \pi_0$, where π_0 is the initial state distribution, and for any $o \in M(L), \delta \in \Delta$

$$\pi(o\delta) = \frac{\pi(o)\mu(\delta)}{||\pi(o)\mu(\delta)||} \tag{9}$$

where $\mu(\sigma)$ is defined in Section II and can be computed by (1). Define a nonfault indicator binary column vector $I_{nf} \in \{0, 1\}^{|Y|\times 1}$, where an entry 1 indicates a nonfault state. It is easy to see that $P^n_N(o)$, the n-step prognosis probability of no-fault following o in (2), can be computed by

$$P^n_N(o) = \pi(o)\Omega^n_{G^R} I_{nf}$$

and then $P^*_N(o)$, the least prognosis probability of no-fault following o in as defined in (7), can be computed by

$$P^*_N(o) = \min_{n\in\mathbb{N}} P^n_N(o) = \min_{n\in\mathbb{N}} \pi(o)\Omega^n_{G^R} I_{nf}.$$

This requires the computation of $P^n_N(o)$ for all possible $n \in \mathbb{N}$, and hence is intractable. Next, we present a new characterization of the prognoser that converts the online prognosis problem to control problem of Markov chain under linear safety constraint, as presented in [19]–[21]. Define

$$\begin{aligned}\Pi_\rho &:= \{\pi \in \Pi : \pi I_{nf} > \rho\}\\ \mathcal{I}_\rho &:= \{\hat{\Pi} \subseteq \Pi_\rho : \forall \pi \in \hat{\Pi}, \pi\Omega_{G^R} \in \hat{\Pi}\}\\ &= \{\hat{\Pi} \subseteq \Pi_\rho : \forall n \in \mathbb{N}, \pi_0 \in \hat{\Pi} \Rightarrow \pi_0\Omega^n_{G^R} \in \hat{\Pi}\}\end{aligned}$$

i.e., Π_ρ is the set of state distributions such that the nonfault probability is greater than ρ, and $\mathcal{I}_\rho$ is a subset of Π_ρ such that all state distributions $\{\pi_0\Omega^n_{G^R}, n \in \mathbb{N}\}$ are members of $\mathcal{I}_\rho$ whenever π_0 itself is an element of $\mathcal{I}_\rho$. It is obvious that $\mathcal{I}_\rho$ is closed under unions and, hence, possesses a unique maximal element called MINS, denoted by Π^*_ρ, i.e., $\forall\hat{\Pi} \in \mathcal{I}_\rho, \hat{\Pi} \subseteq \Pi^*_\rho$. Then, we have

$$\begin{aligned}(P^*_N(o) \leq \rho) &\Leftrightarrow (\exists n \in \mathbb{N}, \pi(o)\Omega^n_{G^R} I_{nf} \leq \rho)\\ &\Leftrightarrow (\exists n \in \mathbb{N}, \pi(o)\Omega^n_{G^R} \notin \Pi_\rho)\\ &\Leftrightarrow (\pi(o) \notin \Pi^*_\rho).\end{aligned}$$

The next theorem follows directly from the aforementioned analysis, and is given without proof.

Theorem 3: The prognoser of (4) can be equivalently reformulated as

$$\forall o \in M(L), [D(o) = F] \Leftrightarrow [\exists \overline{o} \leq o : \boldsymbol{\pi}(\overline{o}) \notin \Pi^*_\rho]. \tag{10}$$

B. Computation of MINS

Now the issue remains as to compute the MINS Π^*_ρ for a given ρ. Similar to the works in [19]–[21], we have the following results. Theorem 4 provides a criterion for verifying that Π^*_ρ is nonempty.

Theorem 4: Given a Markov chain with transition matrix Ω_{G^R}, let $\Pi^*_\rho \subseteq \Pi_\rho$ be the MINS corresponding to decision threshold ρ. Then, Π^*_ρ is nonempty if and only if Ω_{G^R} has a stationary state distribution that lies in Π_ρ.

Proof: Suppose $\pi^* \in \Pi_\rho$ is a stationary state distribution of Ω_{G^R}. Then, $\{\pi^*\} \in \Pi_\rho$, which implies $\{\pi^*\} \in \Pi^*_\rho$ and, hence, Π^*_ρ is nonempty and sufficiency follows. To show the necessity, suppose $\Pi^*_\rho \neq \emptyset$. There should exist a sequence of state distributions $\pi_1, \pi_2, \ldots$ such that $\pi_k \in \Pi^*_\rho$ and $\pi_{k+1} = \pi_k\Omega_{G^R}$ for all $k \geq 1$. Let $d \geq 1$ be the period of Ω_{G^R}. Then, there exists $m \geq 1$, such that $\pi_m = \pi_{m+d}$. Then, it follows that $\pi = \frac{1}{d}\sum_{i=0}^{d-1} \pi_{m+i}$ is a stationary state distribution of Ω_{G^R}. Moreover, since $\pi_{m+i} \in \Pi^*_\rho$ for all $i = 0, \ldots, d-1$, $\pi I_{nf} = \frac{1}{d}\sum_{i=0}^{d-1} \pi_{m+i} I_{nf} > \frac{1}{d}\sum_{i=0}^{d-1} \rho = \rho$, i.e., $\pi \in \Pi_\rho$, which establishes the necessity.

When all the stationary state distributions of Ω_{G^R} are not in Π_ρ, according to Theorem 4, $\Pi^*_\rho = \emptyset$ and the prognoser issues F for all observations. When there is one stationary distribution of Ω_{G^R} lies in Π_ρ, the following algorithm computes Π^*_ρ. The algorithm starts by initializing Π^*_ρ as the set of stationary state distributions that lies in Π_ρ, and iteratively enlarges Π^*_ρ to the set of state distributions π so that either π is already in Π^*_ρ in previous iteration, or its one step successor $\pi\Omega_{G^R}$ falls in Π^*_ρ. The algorithm repeats until a fixed point is reached.

Algorithm 1: Let $\mathcal{N}_s \subseteq \Pi_\rho$ be the set of stationary state distributions of $\Omega^d_{G^R}$ that lies in Π_ρ and $\mathcal{N}_0 \subseteq \Pi_\rho$ be the set of state distribution π such that $\exists \pi_s \in \mathcal{N}_s, \pi\Omega^{kd}_{G^R} \to \pi_s$ as $k \to \infty$, where d is the period of Ω_{G^R}. For $k = 1, 2, \ldots$, iteratively compute

$$\begin{aligned}\mathcal{N}^{(0)} &:= \mathcal{N}_s\\ \mathcal{N}^{(k)} &:= \{\pi \in \mathcal{N}_0 : \pi\Omega_{G^R} \in \mathcal{N}^{(k-1)}\}.\end{aligned}$$

Terminate the algorithm when $\mathcal{N}^{k+1} = \mathcal{N}^k$.

Note that the iterative update in Algorithm 1 can be rephrased as

$$\mathcal{N}^{(k)} = \{\pi \in \mathcal{N}_0 : \pi\Omega^k_{G^R} \in \mathcal{N}_s \text{ and } \pi\Omega^l_{G^R} \in \mathcal{N}_0 \subseteq \Pi_\rho, 1 \leq l < k\}. \tag{11}$$

The next theorem ensures the finite iteration of Algorithm 1 and is inspired from [20, Theorem 4.3].

Theorem 5: There exists a finite integer $\hat{k} \in \mathbb{N}$ such that $\mathcal{N}^{(\hat{k}+1)} = \mathcal{N}^{(\hat{k})} = \Pi^*_\rho$.

Proof: We first prove the finite termination by contradiction. If the iteration does not terminate in a finite number of steps, then there exists a sequence of $\{\pi^{(k)}\} \subset \mathcal{N}_0$ such that $\pi^{(k)} \in \mathcal{N}^{(k+1)} - \mathcal{N}^{(k)}$ or $\pi^{(k)} \in \mathcal{N}^{(k)} - \mathcal{N}^{(k-1)}$, for all $k \geq 1$. According to (10), $\pi^{(k)}$ satisfies

$$\begin{aligned}&\text{either } [\pi^{(k)}\Omega^k_{G^R} \notin \mathcal{N}_s, \pi^{(k)}\Omega^{k+1}_{G^R} \in \mathcal{N}_s]\\ &\text{or } [\pi^{(k)}\Omega^k_{G^R} \in \mathcal{N}_s, \pi^{(k)}\Omega^{k+1}_{G^R} \notin \mathcal{N}_s]\end{aligned}$$

which implies

$$\left\{\pi^{(k)}\Omega^k_{G^R}, \pi^{(k)}\Omega^{k+1}_{G^R}\right\} \not\subset \mathcal{N}_s \quad \forall k \in \mathbb{N}. \tag{12}$$

Let $\widetilde{\pi}$ be any limit point of the sequence $\{\pi^{(k)}\}$. Since Ω^{d}_{GR} is aperiodic, Ω^{ld}_{GR} tends to a limit as $l \to \infty$. Thus, $\widetilde{\pi}\Omega^{ld}_{GR} \to \pi^*_0$ and $\widetilde{\pi}\Omega^{ld+1}_{GR} \to \pi^*_1$, as $l \to \infty$, for some $\pi^*_0, \pi^*_1 \in \Pi$. However, since $\widetilde{\pi} \in \mathcal{N}_0$, we have $\pi^*_0, \pi^*_1 \in \mathcal{N}_s$. Since $\pi^{(l_m d)} \to \widetilde{\pi}$ along some subsequence $\{l_m\} \subseteq \mathbb{N}$, we can conclude that there exist $l_0, m_0 \in \mathbb{N}$ such that $\{\pi^{(l_m d)}\Omega^{ld}_{GR}, \pi^{(l_m d)}\Omega^{ld+1}_{GR}\} \subset \mathcal{N}_s$, for all $l \geq l_0, m \geq m_0$. Choosing $m' \geq m_0$ such that $l_{m'} \geq l_0$, we have $\{\pi^{(l_{m'} d)}\Omega^{l_{m'} d}_{GR}, \pi^{(l_{m'} d)}\Omega^{l_{m'} d+1}_{GR}\} \subset \mathcal{N}_s$, which contradicts (11) if we set $k := l_{m'} d$. Hence, the iteration terminates at some finite $\hat{k} \in \mathbb{N}$.

Next, we show that $\mathcal{N}^{(\hat{k})}$ is the MINS. It is clear that $\mathcal{N}^{(\hat{k})} \subseteq \Pi^*_\rho$. To show that $\mathcal{N}^{(\hat{k})} \supseteq \Pi^*_\rho$, let $\hat{\pi} \in \Pi^*_\rho$ be arbitrary. Since $\hat{\pi}\Omega^{ld+i}_{GR} \to \pi^*_i$, as $l \to \infty$, for some finite collection $\{\pi^*_i \in \Pi_s, i = 0, \ldots, d-1\} \subseteq \mathcal{N}_s$, we can conclude there exists $k_0 \in \mathbb{N}$ such that $\hat{\pi}\Omega^{k}_{GR} \in \mathcal{N}_s$, for all $k \geq k_0$. Therefore, by (10), $\hat{\pi} \in \mathcal{N}^{(k)}$, for all sufficiently large k. ■

Remark 1: Once the maximum initial nonfault set Π^*_ρ has been computed, then for each observation o, the online prognoser (4) computes the current state distribution $\pi(o)$ according to (8), determines whether $\pi(o)$ belongs to Π^*_ρ or not, and issues a prognostic decision according to (9).

Remark 2: Theorem 5 guarantees that Algorithm 1 terminates within finite number of iterations. Furthermore, it can be seen that the computation of online prognoser, as highlighted in Remark 1, is of polynomial complexity in the number of states, i.e., $|X|$. In the setting when the system consists of multiple local components, $|X|$ is the product of the individual component's states, which is standard for any DES analysis. In such case, a decentralized/distributed framework can be developed as is customary; see for example in [8], [11], and [26].

Example 3: For the refined model G^R in Fig. 1, we relabel the refined states space as $0 := (0,0), 1 := (1,1), 2 := (2,2), 3 := (3,3), 4 := (4,4), 5 := (4,F)$. Then, we have

$$\begin{aligned}
\Pi_s &= \{\pi \in \Pi : \pi_1 + \pi_5 = 1\}\\
\Pi_\rho &= \{\pi \in \Pi : \pi_5 \leq 1 - \rho\}\\
\mathcal{N}_s &= \{\pi \in \Pi : \pi_1 > \rho, \pi_1 + \pi_5 = 1\}\\
\mathcal{N}_0 &= \{\pi \in \Pi : 0.5\pi_0 + \pi_1 > \rho\}.
\end{aligned}$$

Algorithm 1 for this example terminates in the fifth iteration as

$$\begin{aligned}
\mathcal{N}^{(0)} &:= \mathcal{N}_s = \{\pi \in \Pi : \pi_1 > \rho, \pi_1 + \pi_5 = 1\}\\
\mathcal{N}^{(1)} &:= \{\pi \in \Pi : \pi_1 > \rho, \pi_1 + \pi_4 + \pi_5 = 1\}\\
\mathcal{N}^{(2)} &:= \{\pi \in \Pi : \pi_1 > \rho, \pi_1 + \pi_3 + \pi_4 + \pi_5 = 1\}\\
\mathcal{N}^{(3)} &:= \{\pi \in \Pi : \pi_1 > \rho, \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1\}\\
\mathcal{N}^{(4)} &:= \{\pi \in \Pi : 0.5\pi_0 + \pi_1 > \rho\}\\
\mathcal{N}^{(5)} &:= \{\pi \in \Pi : 0.5\pi_0 + \pi_1 > \rho\} = \mathcal{N}^{(4)}.
\end{aligned}$$

Therefore, $\Pi^*_\rho = \mathcal{N}^{(5)} = \{\pi \in \Pi : 0.5\pi_0 + \pi_1 > \rho\}$. Note that in this case, it happens that Π^*_ρ equals $\mathcal{N}_0$.

Example 4: Next, we illustrate the online prognosis for the system in Example 3. When the system executes $s = dabbb$ and produces observation $o = abbb$, the resulting state distributions are

$$\begin{aligned}
\pi(a) &= \begin{bmatrix} 0 & 0.375 & 0.625 & 0 & 0 & 0 \end{bmatrix}\\
\pi(ab) &= \begin{bmatrix} 0 & 0.4444 & 0.5556 & 0 & 0 & 0 \end{bmatrix}\\
\pi(abb) &= \begin{bmatrix} 0 & 0.5161 & 0.4839 & 0 & 0 & 0 \end{bmatrix}\\
\pi(abbb) &= \begin{bmatrix} 0 & 0.5872 & 0.4128 & 0 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Suppose $\rho = 0.3$ in this example. According to Example 3, we have

$$\Pi^*_\rho = \{\pi \in \Pi : 0.5\pi_0 + \pi_1 > 0.3\}.$$

It is then trivial to see that $\pi(\overline{o}) \in \Pi^*_\rho$ for all $\overline{o} \leq o$, and hence no FA occurs for trace s. ■

V. Conclusion

In this article, we studied the prognosis of fault, i.e., its prediction prior to its occurrence, for stochastic DESs. In the prior work [10], the notion of S_m-prognosability for stochastic DESs was formulated, generalizing the corresponding notion from the logical setting. This article complements Chen and Kumar [10] by providing an online recursive prognosis algorithm that relies on converting the computation of least prognostic probability to computation of an MINS, as showed in this article. We showed that the condition for issuing an affirmative decision of an online failure prognoser can be reduced to a verification problem of a controlled Markov chain under a safety constraint. The proposed algorithm then computes the set of initial state distributions from which the probability of future fault is less than a user-specified threshold ρ, denoted as MINS. The online failure prognoser then checks whether the conditional distribution following an observation belongs to MINS or not, and issues prognostic decision accordingly. The termination of the proposed algorithm for computing MINS for any given ρ is guaranteed. Future work includes development of algorithms for computing the decision threshold ρ and the largest possible reaction bound m to satisfy given performance requirements $\phi, \tau > 0$ for FA and MD rates. Extension to the decentralized setting [27]–[29] or distributed setting [8] would be another potential direction for future work.

References

[1] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Hoboken, NJ, USA: Wiley, 2006.

[2] S. Genc and S. Lafortune, "Predictability of event occurrences in partially-observed discrete-event systems," *Automatica*, vol. 45, no. 2, pp. 301–311, 2009.

[3] J. Chen and R. Kumar, "Pattern mining for predicting critical events from sequential event data log," in *Proc. Int. Workshop Discrete Event Syst.*, Paris-Cachan, France, 2014, pp. 1–6.

[4] R. Kumar and S. Takai, "Decentralized prognosis of failures in discrete event systems," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 48–59, Jan. 2010.

[5] S. Takai and R. Kumar, "Inference-based decentralized prognosis in discrete event systems," *IEEE Trans. Autom. Control*, vol. 56, no. 1, pp. 165–171, Jan. 2011.

[6] S. Takai, "Robust prognosability for a set of partially observed discrete event systems," *Automatica*, vol. 51, pp. 123–130, 2015.

[7] M. Yokotani, T. Kondo, and S. Takai, "Abstraction-based verification and synthesis for prognosis of discrete event systems," *Asian J. Control*, vol. 18, pp. 1279–1288, 2016.

[8] S. Takai and R. Kumar, "Distributed failure prognosis of discrete event systems with bounded-delay communications," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1259–1265, May 2012.

[9] F. Cassez and A. Grastien, "Predictability of event occurrences in timed systems," in *Proc. 11th Int. Conf. Formal Model. Anal. Timed Syst.*, 2013, pp. 62–76.

[10] J. Chen and R. Kumar, "Stochastic failure prognosability of discrete event systems," *IEEE Trans. Autom. Control*, vol. 60, no. 6, pp. 1570–1581, Jun. 2015.

[11] X. Yin and Z. Li, "Decentralized fault prognosis of discrete-event systems using state-estimate-based protocols," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1302–1313, Apr. 2019.

[12] X. Yin, "Verification of prognosability for labeled Petri nets," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1828–1834, Jun. 2018.

[13] D. You, S. Wang, and C. Seatzu, "Verification of fault-predictability in labeled Petri nets using predictor graphs," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 4353–4360, Oct. 2019.

[14] R. Ammour, E. Leclercq, E. Sanlaville, and D. Lefebvre, "Faults prognosis using partially observed stochastic Petri-nets: An incremental approach," *Discrete Event Dyn. Syst.*, vol. 28, no. 2, pp. 247–267, 2018.

[15] Z. Ma, X. Yin, and Z. Li, "Marking predictability and prediction in labeled Petri nets," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3608–3623, Aug. 2021.

[16] A. T. Watanabe, R. Sebem, A. B. Leal, and M. D. S. Hounsell, "Fault prognosis of discrete event systems: An overview," *Annu. Rev. Control*, vol. 51, pp. 100–110, 2021.

[17] M. E. Orchard and G. J. Vachtsevanos, "A particle-filtering approach for on-line fault diagnosis and failure prognosis," *Trans. Inst. Meas. Control*, vol. 31, no. 3/4, pp. 221–246, 2009.

[18] R. Ammour, E. Leclercq, E. Sanlaville, and D. Lefebvre, "Fault prognosis of timed stochastic discrete event systems with bounded estimation error," *Automatica*, vol. 82, pp. 35–41, 2017.

[19] A. Arapostathis, R. Kumar, and S. Tangirala, "Controlled Markov chains with safety upper bound," *IEEE Trans. Autom. Control*, vol. 48, no. 7, pp. 1230–1234, Jul. 2003.

[20] A. Arapostathis, R. Kumar, and S.-P. Hsu, "Control of Markov chains with safety bounds," *IEEE Trans. Autom. Sci. Eng.*, vol. 2, no. 4, pp. 333–343, Oct. 2005.

[21] S.-P. Hsu, A. Arapostathis, and R. Kumar, "On controlled Markov chains with optimality requirement and safety constraint," *Int. J. Innov. Comput., Inf., Control*, vol. 6, no. 6, pp. 2497–2511, 2010.

[22] V. K. Garg, R. Kumar, and S. I. Marcus, "A probabilistic language formalism for stochastic discrete-event systems," *IEEE Trans. Autom. Control*, vol. 44, no. 2, pp. 280–293, Feb. 1999.

[23] X. Wang and A. Ray, "A language measure for performance evaluation of discrete-event supervisory control systems," *Appl. Math. Model.*, vol. 28, no. 9, pp. 817–833, 2004.

[24] D. Thorsley and D. Teneketzis, "Diagnosability of stochastic discrete-event systems," *IEEE Trans. Autom. Control*, vol. 50, no. 4, pp. 476–492, Apr. 2005.

[25] J. Chen and R. Kumar, "Failure detection framework for stochastic discrete event systems with guaranteed error bounds," *IEEE Trans. Autom. Control*, vol. 60, no. 6, pp. 1542–1553, Jun. 2015.

[26] S. Takai and R. Kumar, "A generalized inference-based prognosis framework for discrete event systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6819–6824, 2017.

[27] W. Qiu, Q. Wen, and R. Kumar, "Decentralized diagnosis of event-driven systems for safely reacting to failures," *IEEE Trans. Autom. Sci. Eng.*, vol. 6, no. 2, pp. 362–366, Apr. 2009.

[28] J. Chen and R. Kumar, "Decentralized failure diagnosis of stochastic discrete event systems," in *Proc. 9th IEEE Int. Conf. Autom. Sci. Eng.*, Madison, WI, USA, 2013, pp. 1083–1088.

[29] S. Takai and R. Kumar, "Decentralized diagnosis for nonfailures of discrete event systems using inference-based ambiguity management," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 2, pp. 406–412, Mar. 2010.