2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Failure Prognosability of Stochastic Discrete Event Systems

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

***Abstract*— We study the prognosis of fault, i.e., its prediction prior to its occurrence, in stochastic discrete event systems. We introduce the notion of m-steps Stochastic-Prognosability, called S_m-Prognosability, which allows the prediction of a fault at least m-steps in advance. We formalize the notion of a prognoser and also show that S_m-Prognosability is necessary and sufficient for the existence of a prognoser that can predict a fault at least m-steps prior to occurrence, while achieving any arbitrary false alarm and missed detection rates. We also provide a polynomial algorithm for the verification of S_m-Prognosability.**

I. Introduction

The problem of predicting a fault prior to its occurrence is a well researched area [1]-[5]. In [1] the notion of uniformly bounded prognosability of fault was formulated for logical discrete event systems (DESs), where each fault-trace must possess a nonfault-prefix such that for all indistinguishable traces, a future fault is inevitable within a bounded delay that is uniform across all fault-traces. Such a nonfault-prefix from which a future fault is inevitable is termed an *indicator*. The notion was later extended to the decentralized setting in [2] and the requirement of the existence of a uniform bound was also removed. Reference [2] also 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 [4], 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 [5], where the notion of joint-prognosability and its verification was proposed.

In order to generalize the notion of prognosability to stochastic DESs, in this paper, we introduce m-steps Stochastic-Prognosability, or simply S_m-Prognosability, 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. We formalize the notion of a prognoser that maps observations to decisions by comparing a suitable statistic with a threshold, and show 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 and MD rate requirement. In this sense S_m-Prognosability can be viewed as a generalization of the logical prognosability, since it provides a basis for the existence and synthesis of a prognoser that can achieve a user-specified level of FA and MD. In contrast, the logical version is rather rigid, offering no further options for systems that fail to be logically prognosable, even when there may exist a prognoser that can achieve a satisfying performance as measured in terms of FA and MD rates. Also, in the logical setting, an indicator cannot visit a cycle of nonfault-states, which can be restrictive; in contrast in the stochastic setting, an indicator can visit a cycle of nonfault-states as long as the cycle is not absorbing. A polynomial complexity algorithm for verifying S_m-Prognosability is also provided.

The rest of this paper is organized as following: The notations and some preliminaries are presented in Section II, followed by the definition of S_m-Prognosability and stochastic prognoser in Section III and IV, respectively. Section IV also shows necessity and sufficiency of S_m-prognosability for the existence of an m-prognoser that can fulfill any desired level of error bounds over FA and MD. Section V gives an algorithm for verifying S_m-Prognosability and the paper is concluded in Section VI.

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 $\sim \in \{<, \leq, >, \geq, =\}$ and $n \in \mathbb{N}$, where $\mathbb{N}$ denotes the set of all nonnegative integers, define $\Sigma^{\sim n} := \{s \in \Sigma^* : |s| \sim n\}$ and denote $\Sigma^{=n}$ as Σ^n for simplicity. For $L \subseteq \Sigma^*$, its prefix-closure is defined 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 $L_1 L_2 := \{st : s \in L_1, t \in L_2\}$, 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

The research was supported in part by the National Science Foundation under the grants, NSF-ECCS-0801763, NSF-ECCS-0926029, and NSF-CCF-1331390.

The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: junchen@iastate.edu; rkumar@iastate.edu).

$\alpha : X \times \Sigma \times X \to [0,1]$ is the transition probability function [6] 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.) The transition probability function α can be generalized to $\alpha : X \times \Sigma^* \times X$ recursively as follows: $\forall x_i, x_j \in X, s \in \Sigma^*, \sigma \in \Sigma, \alpha(x_i, s\sigma, x_j) = \sum_{x_k\in X}\alpha(x_i,s,x_k)\alpha(x_k,\sigma,x_j)$, and $\alpha(x_i,\epsilon,x_j) = 1$ if $x_i = x_j$ and 0 otherwise. Define a *transition* in G as a triple $(x_i,\sigma,x_j) \in X \times \Sigma \times X$ where $\alpha(x_i,\sigma,x_j) > 0$ and 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 $x \in X_C$, $\sum_{\sigma\in\Sigma}\sum_{x'\in X_C}\alpha_C(x,\sigma,x') = 1$.

A DES G is non-stochastic if $\alpha : X \times \Sigma \times X \to \{0,1\}$, and a non-stochastic 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. Note for non-stochastic DES, we allow the summation $\sum_{\sigma\in\Sigma}\sum_{x'\in X}\alpha(x,\sigma,x')$ to be larger than one, as for non-stochastic DES, α is simply a transition indicator, and not a probability.

To represent the limited sensing capabilities of a prognoser, we introduce an event observation mask, $M : \overline{\Sigma} \to \overline{\Delta}$, 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 $M : 2^\Sigma \to 2^\Delta$ 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 the execution of any fault-trace in $L - K$ prior to its execution, and at least m steps in advance, and with sufficient confidence. Let $K \subseteq L$ be generated by a *deterministic* automaton $R = (Q, \Sigma, \beta, q_0)$ such that $L(R) = K$ (from now on we interchangably 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., $\sum_{x\in X}\alpha(x_0,s,x) = \sum_{(x,\overline{q})\in X\times\overline{Q}}\gamma((x_0,q_0),s,(x,\overline{q}))$.

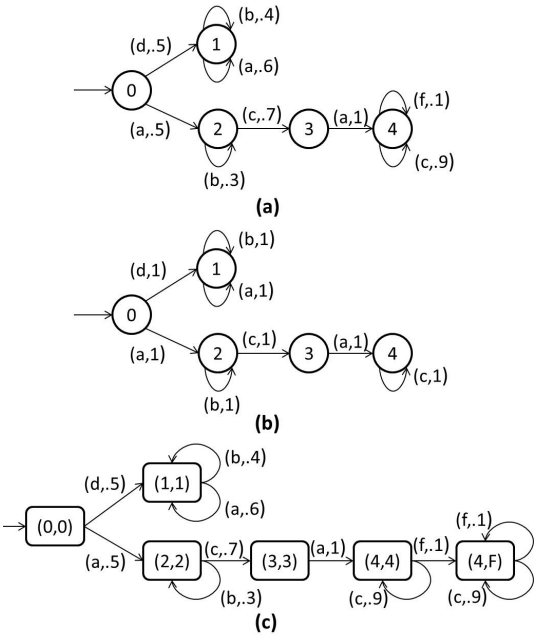

Fig. 1. (a) Stochastic automaton G; (b) Nonfault specification R; (c) Refinement G^R.

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 $L-K$ transition 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 selfloop 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 $\mu_{i,\sigma,j} = \sum_m \lambda_{im}\mu_{m,\sigma,j} + \alpha(x_i, \sigma, x_j)$, where the first term on the right hand side (RHS) involves transitioning in at least two steps via some intermediate states, 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 [7], [8] for a similar matrix equation):

$$\boldsymbol{\mu}(\sigma) \quad = \quad \boldsymbol{\lambda}\boldsymbol{\mu}(\sigma) + \boldsymbol{\alpha}(\sigma), \tag{1}$$

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 of Stochastic DESs

In this section, we formalize the notion of prognosability, called *m-steps Stochastic-Prognosability*, or simply *S_m-Prognosability*, for stochastic DESs, and provide necessary and sufficient conditions for the verification of S_m-Prognosability. In the next section we show that for finite-state systems, S_m-Prognosability is necessary and sufficient for the existence of a prognoser that can predict a fault at least m-steps prior to occurrence, while achieving any arbitrary false alarm and missed detection rates.

Let L be a nonempty closed language and $K \subseteq L$ be a nonempty closed language representing a nonfault-specification. In order to be able to make a prognostic decision, we define the *n-step prognostic probability of no-fault* following an observation $o \in M(L)$ as:

$$\begin{aligned} P_N^n(o) \quad &:= \quad \frac{Pr(\{M^{-1}(o)\}\Sigma^n \cap K)}{Pr(\{M^{-1}(o)\}\Sigma^n \cap L)} \\ &= \quad \frac{Pr(\{M^{-1}(o) \cap K\}\Sigma^n \cap K)}{Pr(M^{-1}(o) \cap L)}, \end{aligned} \tag{2}$$

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

$$\begin{aligned} P_N^*(o) \quad &:= \quad \min_{n \in \mathbb{N}} P_N^n(o) \\ &= \quad \frac{\min_{n \in \mathbb{N}} Pr(\{M^{-1}(o)\}\Sigma^n \cap K)}{Pr(\{M^{-1}(o)\} \cap L)}. \end{aligned} \tag{3}$$

Note $P_N^n(o)$ 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 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 minimization operation in (3) only applies to the numerator.

To help formalize the prognosability for stochastic DESs, we introduce the notions of *boundary* fault-traces whose all strict prefixes are nonfault, *m-steps interior* nonfault-traces for which a fault can occur in the next $(m+1)$th step while no fault can occur within the next m steps, *persistent* nonfault-traces whose all extensions are nonfault, *indicator* nonfault-traces for which a future fault is guaranteed with arbitrary confidence and *nonindicator* nonfault-traces that are not the indicator traces.

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

- *boundary* fault-traces as, $\partial := \{s \in L - K : pr(s) - \{s\} \subseteq K\}$;
- *m-steps interior* nonfault-traces of K with respect to L (where $m \geq 0$) as, $\partial_m^- := \{s \in K : \{s\}\Sigma^{\leq m} \cap (L - K) = \emptyset, \{s\}\Sigma^{m+1} \cap \partial \neq \emptyset\}$;
- *persistent* nonfault-traces of K with respect to L as, $\aleph := \{s \in K : \forall n \in \mathbb{N}, \{s\}\Sigma^n \cap (L - K) = \emptyset\}$;
- *indicator* nonfault-traces of K with respect to L as, $\mathfrak{J} := \{s \in K : \forall \rho > 0, \exists n \in \mathbb{N}, Pr(\{s\}\Sigma^n \cap K) \leq \rho\}$;
- *nonindicator* nonfault-traces of K with respect L as, $\Upsilon := K - \mathfrak{J}$.

Note that $\Upsilon = \{s \in K : \exists \rho > 0, \forall n \in \mathbb{N}, Pr(\{s\}\Sigma^n \cap K) > \rho\}$. Also note that $\aleph$ is "extension-closed" in the sense that if it possesses $s \in K$, then it also possesses all extensions $t \in L$ with $s \leq t$.

Next we introduce the definition of S_m-Prognosability which requires that, for any threshold value $\rho > 0$ and error bound $\tau > 0$, there exists a reaction bound $k \geq m$, such that the set of boundary fault-traces, that are either shorter than k in length or for which a prognostic decision can not be made k steps in advance with confidence level ρ, occurs with probability smaller than τ.

Definition 2: 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

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

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

The next lemma states that we can always choose the reaction bound k in Definition 2 to equal m, thereby simplifying the definition a bit.

Lemma 1: A pair (L, K) of closed languages with $K \subseteq L$ is S_m-Prognosable if and only if $\forall \tau, \rho > 0$,

$$\begin{aligned} &Pr(s \in \partial : [|s| \leq m] \\ &\qquad \vee [\forall u \in s/\Sigma^{>m}, P_N^*(M(u)) > \rho]) < \tau. \end{aligned} \tag{5}$$

Denote $\ell(\partial) = \min\{|s|, s \in \partial\}$ as the length of the shortest fault-trace in $L-K$. Then the following theorem provides a necessary and sufficient condition for S_m-prognosability requiring the reaction bound m to be smaller than the length of the shortest fault-trace, $\ell(\partial)$, and every boundary fault-trace in ∂ to possess a nonfault-prefix which is more than m-steps shorter and is unambiguously an indicator.

Theorem 1: A pair (L, K) of closed languages with $K \subseteq L$ is S_m-Prognosable if and only if $m < \ell(\partial)$ and

$$(\forall s \in \partial)(\exists u \in s/\Sigma^{>m})(M^{-1}M(u) \cap K \subseteq \mathfrak{J}). \tag{6}$$

Example 2: For the system in Fig. 1, $\ell(\partial) = 4$, so by Theorem 1, the system can not be S_m-Prognosable with $m \geq 4$. The set of indicator traces is $\mathfrak{J} = \{a\}\Sigma^* \cap K$, and the set of nonindicator traces is $\Upsilon = \{\epsilon\} \cup \{d\}\Sigma^* \cap L$, while the set of boundary fault-traces is $\partial = ab^*cac^*f$. One can check that for any $s \in \partial$, there exists $u \in s/\Sigma^{>1} \subseteq \{ab^*c\}\Sigma^* \cap K$ such that $M^{-1}M(u) \cap K \subseteq \mathfrak{J}$. Therefore by Theorem 1, (L, K) is S_1-Prognosable. On the other hand, for $s = acaf \in \partial$, $u = a \in s/\Sigma^{>2}$ is such that $M^{-1}M(u) \cap K \cap \Upsilon = \{da\} \neq \emptyset$. Therefore by Theorem 1, (L, K) is not S_2-Prognosable. ■

The following corollary is directly obtained from Theorem 1, and captures the expected property that prognosability continues to hold even with smaller reaction bound.

Corollary 1: Given a pair (L, K) of closed languages with $K \subseteq L$, if (L, K) is S_m-Prognosable, then (L, K) is $S_{m'}$-Prognosable for all nonnegative $m' \leq m$, whereas if (L, K) is not S_m-Prognosable, then (L, K) is not $S_{m'}$-Prognosable for all $m' \geq m$.

For an S_m-Prognosable system, Theorem 1 requires that each boundary fault trace possess a more than m-steps shorter prefix that is unambiguously an indicator. We can strengthen this theorem by requiring that *exactly* the $(m+1)$-shorter prefix possess the said property. This requires the result of the next lemma stating that indicators are "extension-closed" (nonfault-extensions of indicators are also indicators), while nonindicators are prefix-closed (prefixes of nonindicators are also nonindicators).

Lemma 2: For a pair (L, K) of closed languages with $K \subseteq L$, it holds that $\mathfrak{J}\Sigma^* \cap K \subseteq \mathfrak{J}$, and $pr(\Upsilon) \subseteq \Upsilon$.

Using Lemma 2, we can strengthen Theorem 1 to obtain a new result which we employ in Section V for verifying S_m-Prognosability. The new theorem states that S_m-Prognosability holds if and only if the reaction bound $m < \ell(\partial)$, and all m-steps interior traces are distinguishable from any nonindicator trace.

Theorem 2: A pair (L, K) of closed languages with $K \subseteq L$ is S_m-Prognosable if and only if $m < \ell(\partial)$ and

$$M^{-1}M(\partial_m^-) \cap \Upsilon = \emptyset. \tag{7}$$

Example 3: For the system shown in Fig. 1, $\mathfrak{J} = \{a\}\Sigma^* \cap K$, $\Upsilon = \{\epsilon\} \cup \{d\}\Sigma^* \cap L$, $\partial_2^- = ab^*$ and $\partial_1^- = ab^*c$. One can easily check that $M^{-1}M(\partial_2^-) \cap \Upsilon = dab^* \neq \emptyset$ and $M^{-1}M(\partial_1^-) = ab^*c \subseteq \mathfrak{J}$. Therefore (L, K) is S_1-Prognosable but not S_2-Prognosable, as discussed in Example 2. ■

IV. Prognoser and its Existence Condition

In this section we formally define a prognoser with reaction bound at least m, called an m-prognoser, along with its FA and MD rates, and show that the notion of S_m-Prognosability introduced in the previous section acts as a necessary and sufficient condition for the existence of an m-prognoser capable of achieving any FA and MD rates.

In order to predict a fault in advance, the prognoser 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:

$$\forall o \in M(L), [D(o) = F] \Leftrightarrow [\exists \overline{o} \leq o : P_N^*(\overline{o}) \leq \rho], \tag{8}$$

where P_N^* is as defined by (2) and (3). Note that according to (8), once a warning is issued, it remains unchanged for the subsequent extensions.

For a prognoser that aims to predict a fault at least m steps before its occurrence, a *missed detection* (MD) occurs when a fault happens while the prognoser fails to issue a warning m steps in advance. On the other hand a *false alarm* (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 an m-prognoser can be defined as:

$$P^{md} = Pr(s \in \partial : [|s| \leq m] \vee [D(M(s/\Sigma^{m+1})) = \epsilon] \tag{9}$$

$$P^{fa} = Pr(s \in \aleph : D(M(s)) = F). \tag{10}$$

Considering the fact the once the prognoser issues F, it issues F for any subsequent observations, the above equations can also be equivalently presented as:

$$\begin{aligned} P^{md} &= Pr(s \in \partial : [|s| \leq m] \\ &\qquad \vee [\forall u \in s/\Sigma^{>m}, P_N^*(M(u)) > \rho]) \\ P^{fa} &= Pr(s \in \aleph : \exists u \in pr(s), P_N^*(M(u)) \leq \rho). \end{aligned}$$

Example 4: 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 $P_N^*(o)$ approaches 0. Therefore for a 1-prognoser with any threshold ρ, all fault-traces can be prognosed, and hence no missed detection. However, for a 2-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 miss-detected. ■

In order to establish a condition for the existence of an m-prognoser in terms of the property of S_m-prognosability, we first establish the following corollary of Theorem 1 and Lemma 2.

Corollary 2: If a pair (L, K) of closed languages with $K \subseteq L$ is S_m-Prognosable, then $M^{-1}M(\Upsilon) \cap (L - K) = \emptyset$.

The next lemma states that under the assumption of regularity of languages L and K, equivalently the finiteness of the state-space of G^R, no extension of an indicator can be persistently nonfault, whereas some extension of a nonindicator must be persistently nonfault. The lemma requires the finiteness of the state-space that guarantees the probability of staying in a transient state approaches 0 while the system evolves.

Lemma 3: For a pair (L, K) of closed regular languages with $K \subseteq L$, we have $\mathfrak{J}\Sigma^* \cap \aleph = \emptyset$ and $\Upsilon\Sigma^* \cap \aleph \neq \emptyset$.

Now we are ready to present the main result of the section, which shows that for regular languages L and K, S_m-Prognosability is necessary and sufficient for the existence of an m-prognoser to satisfy any level of FA and MD rates.

Theorem 3: 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 an m-prognoser (and its associated prognostic decision threshold) defined by (8) such that the MD and FA rates defined by (9)-(10) satisfy $P^{md} \leq \tau$ and $P^{fa} \leq \phi$ if and only if (L, K) is S_m-Prognosable.

V. Verification of S_m-Prognosability

Having established S_m-Prognosability as a central property, needed for the existence of an m-prognoser, we next provide a polynomial algorithm for the verification of S_m-Prognosability utilizing Theorem 2. We need the following definitions that identify *m-steps interior* nonfault-states from where no fault can occur within m steps but will occur at the $(m+1)$th step, *indicator* nonfault-states from where a future fault is inevitable with arbitrary confidence, and *nonindicator* nonfault-states which are not indicator states.

Definition 3: Given a stochastic DES $G = (X, \Sigma, \alpha, x_0)$, deterministic nonfault-specification $R = (Q, \Sigma, \beta, q_0)$, with their refinement $G^R = (X \times \overline{Q}, \Sigma, \gamma, (x_0, q_0))$, the set of

- *m-steps interior* nonfault-states $\partial_m^-(X \times \overline{Q}) \subseteq X \times \overline{Q}$ (where $m \geq 0$) are states $(x, \overline{q})$ such that $\overline{q} \neq F$, and there exists $(x', \overline{q}')$ with $\overline{q}' = F$ and $s \in \Sigma^{m+1}$ s.t. $\gamma((x, \overline{q}), s, (x', \overline{q}')) > 0$ and for all $(x', \overline{q}')$, $s \in \Sigma^{\leq m}$, $[\gamma((x, \overline{q}), s, (x', \overline{q}')) > 0] \Rightarrow [\overline{q}' \neq F]$;
- *indicator* nonfault-states $\mathfrak{J}(X \times \overline{Q})$ are states $(x, \overline{q})$ such that $\overline{q} \neq F$ and from which the system can not reach a closed SCC in G^R that contains a nonfault-state;
- *nonindicator* nonfault-states $\Upsilon(X \times \overline{Q})$ are states from which the system can reach a closed SCC in G^R that contains a nonfault-state.

The following lemma is immediate from Definition 1, Definition 3 and Lemma 3.

Lemma 4: Given a pair $(L = L(G), K = L(R))$ of closed regular languages with $K \subseteq L$, then for any $s \in K$,

- $[s \in \partial_m^-] \Leftrightarrow [\exists (x, \overline{q}) \in \partial_m^-(X \times \overline{Q}), \gamma((x_0, q_0), s, (x, \overline{q})) > 0]$;
- $[s \in \mathfrak{J}] \Leftrightarrow [\exists (x, \overline{q}) \in \mathfrak{J}(X \times \overline{Q}), \gamma((x_0, q_0), s, (x, \overline{q})) > 0]$;
- $[s \in \Upsilon] \Leftrightarrow [\exists (x, \overline{q}) \in \Upsilon(X \times \overline{Q}), \gamma((x_0, q_0), s, (x, \overline{q})) > 0]$.

The following algorithm verifies the condition of Theorem 2.

Algorithm 1: For a given stochastic automaton $G = (X, \Sigma, \alpha, x_0)$ and a deterministic nonfault-specification $R = (Q, \Sigma, \beta, x_0)$, perform the following steps:

1) Check if the length of the shortest trace to a state $X \times \{F\}$ in G^R is smaller than m, if the answer is yes, proceed to step 2), otherwise (L, K) is not S_m-Prognosable;
2) Construct a testing automaton $T = G^R \times G^R$ such that at each step the first copy of G^R takes lead in executing transitions, whereas the second copy responds by executing an indistinguishable nonfault-trace. This automaton is denoted as $T = (Z, \Sigma \times \overline{\Sigma}, \delta, z_0)$, where
   - $Z = X \times \overline{Q} \times X \times \overline{Q}$;
   - $z_0 = ((x_0, q_0), (x_0, q_0))$ is the initial state;
   - $\delta : Z \times \Sigma \times \overline{\Sigma} \times Z \rightarrow [0, 1]$ is defined as: $\forall ((x_1, \overline{q}_1), (x_2, \overline{q}_2)), ((x_1', \overline{q}_1'), (x_2', \overline{q}_2')) \in Z, (\sigma, \sigma') \in \Sigma \times \overline{\Sigma}$,

$$\delta(((x_1, \overline{q}_1), (x_2, \overline{q}_2)), (\sigma, \sigma'), ((x_1', \overline{q}_1'), (x_2', \overline{q}_2'))) = \begin{cases} \gamma((x_1, \overline{q}_1), \sigma, (x_1', \overline{q}_1')), \\ \quad \text{if } (\sigma \in \Sigma_{uo}) \wedge (\sigma' = \epsilon) \\ \quad \wedge ((x_2, \overline{q}_2) = (x_2', \overline{q}_2')) \wedge (\overline{q}_2' \neq F); \\ \frac{\gamma((x_1, \overline{q}_1), \sigma, (x_1', \overline{q}_1')) \alpha(L_{G^R}((x_2, \overline{q}_2), \sigma', (x_2', \overline{q}_2')))}{\alpha(L_{G^R}((x_2, \overline{q}_2), M(\sigma)))}, \\ \quad \text{if } (\sigma \in \Sigma - \Sigma_{uo}) \wedge (M(\sigma) = M(\sigma')) \\ \quad \wedge (L_{G^R}((x_2, \overline{q}_2), \sigma', (x_2', \overline{q}_2'))) \neq \emptyset) \\ \quad \wedge (\overline{q}_2' \neq F); \\ 0 \quad \text{otherwise.} \end{cases}.$$

   According to the definition of δ, when the first copy of G^R executes an unobservable event, the second copy responds by ϵ (since it observes nothing); if the first copy executes an observable event σ, then the second copy responds by executing a nonfault-trace consisting of sequence of unobservable events followed by an observable event that has the same mask value as $M(\sigma)$. Note a conditioning is applied to limit the executions of the second copy to indistinguishable nonfault-traces.
3) Check if every state $((x_1, \overline{q}_1), (x_2, \overline{q}_2))$ with $(x_1, \overline{q}_1) \in \partial_m^-(X \times \overline{Q})$ satisfies $(x_2, \overline{q}_2) \notin \Upsilon(X \times \overline{Q})$, (L, K) is S_m-Prognosable if and ony if the answer is yes.

The following theorem guarantees the correctness of Algorithm 1.

Theorem 4: A pair $(L = L(G), K = L(R))$ of closed regular languages with $K \subseteq L$ is S_m-Prognosable if and only if any fault-state can only be reached in more than m-steps in G^R and every reachable state $((x_1, \overline{q}_1), (x_2, \overline{q}_2))$ of T with $(x_1, \overline{q}_1) \in \partial_m^-(X \times \overline{Q})$ satisfies $(x_2, \overline{q}_2) \notin \Upsilon(X \times \overline{Q})$.

Example 5: Let us revisit the system shown in Fig. 1. According to Definition 3, $\mathfrak{J}(X \times \overline{Q}) = \{(2, 2), (3, 3), (4, 4)\}$, $\Upsilon(X \times \overline{Q}) = \{(0, 0), (1, 1)\}$, $\partial_1^-(X \times \overline{Q}) = \{(3, 3)\}$ and

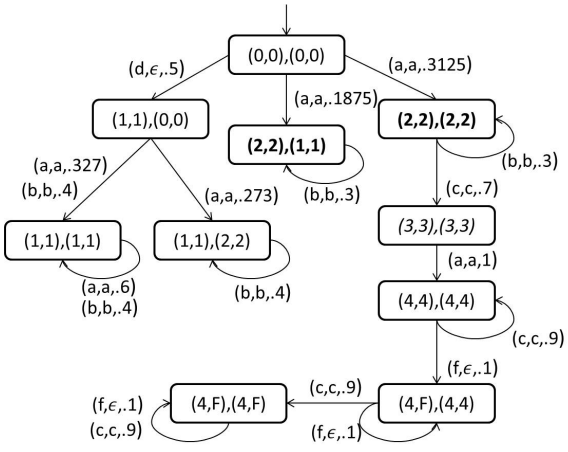

Fig. 2. Testing automaton for the system G^R shown in Fig. 1.

$\partial_2^-(X \times \overline{Q}) = \{(2,2)\}$. It is easy to check that $1 < 2 < \ell(\partial) = 4$. The testing automaton is shown in Fig. 2. The only state $((x_1,\overline{q}_1),(x_2,\overline{q}_2))$ such that $(x_1,\overline{q}_1) \in \partial_1^-(X \times \overline{Q})$ is labeled in *italics*, i.e., state $((3,3),(3,3))$, which satisfies $(x_2,\overline{q}_2) \notin \Upsilon(X \times \overline{Q})$. Therefore (L,K) is S_1-Prognosable. All the states $((x_1,\overline{q}_1),(x_2,\overline{q}_2))$ such that $(x_1,\overline{q}_1) \in \partial_2^-(X \times \overline{Q})$ are labeled in bold, and there exists $((2,2),(1,1))$ such that $(2,2) \in \partial_2^-(X \times \overline{Q})$ and $(1,1) \in \Upsilon(X \times \overline{Q})$. Therefore (L,K) is not S_2-Prognosable. These are as expected from the discussion in Examples 2 and 3. ■

Remark 1: In Algorithm 1. G^R has $O(|X| \times |Q|)$ states and $O(|X|^2 \times |Q| \times |\Sigma|)$ transitions, and the testing automaton $T = G^R \times G^R$ has $O(|X|^2 \times |Q|^2)$ states and $O(|X|^4 \times |Q|^2 \times |\Sigma|^2)$ transitions. The computation of transition probabilities in T requires solving the matrix equation (1) for each $\sigma \in \Sigma - \Sigma_{uo}$ with complexity that is cubic in the number of states in G^R and linear in the number of events in G^R, namely, $O(|X|^3 \times |Q|^3 \times |\Sigma|)$. Thus the complexity of constructing T is $O(|X|^4 \times |Q|^2 \times |\Sigma|^2 + |X|^3 \times |Q|^3 \times |\Sigma|)$. The shortest path to a fault state in G^R can be computed in $O(\sqrt{|X| \times |Q|} \times |X|^2 \times |Q| \times |\Sigma|)$ [9]. Identifying the set of m-steps interior nonfault-states in G^R can be done linearly in the size of G^R, i.e., $O(|X|^2 \times |Q| \times |\Sigma|)$, and identifying the set of indicator nonfault-states can be achieved by determining all the nonfault closed SCC in G^R using the algorithm in [10], which can be done in $O(|X|^3 \times |Q|^3)$. Therefore the overall complexity of Algorithm 1 is $O(|X|^4 \times |Q|^2 \times |\Sigma|^2 + |X|^3 \times |Q|^3 \times |\Sigma|)$, which is polynomial in the number of states and events. Further if G is also deterministic (besides R) so that G^R has a smaller number of transitions, namely, $O(|X| \times |Q| \times |\Sigma|)$, then the verification complexity reduces to $O(|X|^2 \times |Q|^2 \times |\Sigma|^2 + |X|^3 \times |Q|^3 \times |\Sigma|)$. Furthermore, if the mask is "projection-type", the complexity further reduces due to a reduction in the number of transitions in G^R, where each state can now only have at most $|\Sigma|$ outgoing transitions, and thus the $|\Sigma|^2$ term will get replaced by $|\Sigma|$ in the complexity expression.

VI. Conclusion

In this paper, we studied the progosis of fault, i.e., its prediction prior to its occurrence, for stochastic discrete event systems. We formulated the notion of S_m-Prognosability for stochastic DESs, generalizing the corresponding notion from the logical setting [1], [2], and showed that it is a necessary and sufficient condition for the existence of a prognoser that can predict a fault at least m-steps prior to its occurrence, while achieving any arbitrary false alarm and missed detection rates. A polynomial complexity algorithm for the verification of S_m-Prognosability was also provided. There are several directions for future research: 1) An online recursive prognosis algorithm to compute the state distribution resulted by an observation o, $\pi(o)$, so as to be able to check whether $P_N^*(o) \leq \rho$ by checking if $\pi(o)$ falls within a suitable range, and 2) algorithms for computing the decision threshold ρ and the largest possible reaction bound m for a given performance requirement $\phi, \tau > 0$ of FA and MD rates. Also, the applications of probabilistic model checking (see [11] and the references therein) for stochastic prognosis computations, and the prediction of violation of requirement expressed as temporal logic ([12], [13], [14]) would be other directions for future research.

References

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

[2] 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.

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

[4] 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.

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

[6] 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.

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

[8] J. Chen and R. Kumar, "Online failure diagnosis of stochastic discrete event systems," in *Proc. 2013 IEEE Multi-Conf. Syst. and Control*, Hyderabad, India, Aug. 2013, pp. 194–199.

[9] A. V. Goldberg, "Scaling algorithms for the shortest paths problem," *SIAM J. Comput.*, vol. 24, no. 3, pp. 494–504, Jun. 1995.

[10] A. Xie and P. A. Beerel, "Efficient state classification of finite-state Markov chains," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 12, pp. 1334–1339, Dec. 1998.

[11] C. Baier and M. Kwiatkowska, "Preface to the special issue on probabilistic model checking," *Formal methods Syst. Des.*, vol. 43, no. 2, pp. 121–123, Oct. 2013.

[12] J. Chen and R. Kumar, "Failure diagnosis of discrete-time stochastic systems subject to temporal logic correctness requirements," in *Proc. 2014 IEEE Int. Conf. Netw. Sensing, and Control*, Miami, FL, Apr. 2014.

[13] S. Jiang and R. Kumar, "Failure diagnosis of discrete-event systems with linear-time temporal logic specifications," *IEEE Trans. Autom. Control*, vol. 49, no. 6, pp. 934–945, Jun. 2004.

[14] ——, "Diagnosis of repeated failures for discrete event systems with linear-time temporal-logic specifications," *IEEE Trans. Auto. Sci. Eng.*, vol. 3, no. 1, pp. 47–59, Jan. 2006.