D5.3 : Supervision and probability

Collaboration LSV - INRIA Rennes Atlantique

November 6, 2014

En vue d'augmenter la robustesse du fonctionnement d'un sytème partiellement observable et susceptible d'occurrences de fautes, il importe de diagnostiquer ces fautes. D'autre part, afin de prendre en compte les incertitudes de comportement généralement liées à l'environnement, les systèmes sont modélisés par des modèles probabilistes tels que les chaînes de Markov ou les processus de décision markoviens.

Nos premiers travaux (présentés à la conférence FOSSACS 2014) ont donc porté sur le *diagnostic actif* de systèmes probabilistes partiellement observables. Le diagnostic actif consiste à contrôler le système de telle sorte qu'il reste vivant et qu'il soit diagnosticable. Le diagnostic actif avait été étudié dans le cadre de systèmes à événements discrets. Nous avons caractérisé la complexité du problème de décision ainsi que la taille optimale d'un contrôleur dans le cas positif. Nous avons ensuite introduit le *diagnostic actif sain* qui requiert que le système contrôlé conserve une probabilité non nulle de rester correct. De manière surprenante, nous avons démontré que le problème devenait alors indécidable. Ce résultat a aussi des conséquences sur l'étude théorique des POMDP. Enfin en se restreignant à des contrôleurs à mémoire finie, le problème redevient décidable et nous avons de nouveau caractérisé sa complexité exacte.

En examinant la notion de diagnostic utilisée dans les travaux précédents, nous avons remarqué qu'elle présentait des différences subtiles avec la définition originale (cette différence n'apparaît qu'avec les probabilités). Nos travaux suivants (présentés à la conférence FSTTCS 2014) ont porté sur l'étude sémantique et algorithmique du diagnostic (passif) dans les systèmes probabilistes. Nous avons établi une hiérarchie complexe des différentes notions en y incluant la *prédictabilité* et en introduisant la *prédiagnosabilité* qui combine les avantages de la diagnosabilité et de la prédictabilité. Nous avons aussi caractérisé la complexité exacte de chacun des problèmes de décision montrant qu'ils étaient soient NLOGSPACE-complets, soit PSPACE-complets. Enfin nous avons proposé des constructions de diagnostiqueurs de taille optimale.

Active diagnosis for probabilistic systems (long version)*

Nathalie Bertrand[1], Éric Fabre[1],
Stefan Haar[1,2], Serge Haddad[2], and Loïc Hélouët[1]

[1] Inria, France
[2] LSV, ENS Cachan & CNRS & Inria, France

Abstract. The diagnosis problem amounts to deciding whether some specific "fault" event occurred or not in a system, given the observations collected on a run of this system. This system is then diagnosable if the fault can always be detected, and the active diagnosis problem consists in controlling the system in order to ensure its diagnosability. We consider here a stochastic framework for this problem: once a control is selected, the system becomes a stochastic process. In this setting, the active diagnosis problem consists in deciding whether there exists some observation-based strategy that makes the system diagnosable with probability one. We prove that this problem is EXPTIME-complete, and that the active diagnosis strategies are belief-based. The *safe* active diagnosis problem is similar, but aims at enforcing diagnosability while preserving a positive probability to non faulty runs, i.e. without enforcing the occurrence of a fault. We prove that this problem requires non belief-based strategies, and that it is undecidable. However, it belongs to NEXPTIME when restricted to belief-based strategies. Our work also refines the decidability/undecidability frontier for verification problems on partially observed Markov decision processes.

1 Introduction

Diagnosis for discrete event systems was introduced in [SSL+95], and can be described as follows: a labeled transition system performs a run, which may contain some specific events called *faults*. Some of the transition labels are observable, so one gets information about the performed run through its trace, i.e. its sequence of observed labels. The diagnosis problem then amounts to determining whether a fault event occurred or not given the observed trace. The trace is called faulty (resp. correct) if all runs that can have produced it contain (resp. do not contain) a

* This work was supported by project ImpRo ANR-2010-BLAN-0317 and the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement 257462 HYCON2 NOE.

fault. In the remaining cases the trace is called ambiguous. Along with the diagnosis problem comes the diagnosability question: does there exist an infinite ambiguous trace (thus forbidding diagnosis)? For finite transition systems, checking diagnosability was proved to have a polynomial complexity [YL02].

Diagnosis and diagnosability checking have been extended to numerous models (Petri nets [CGLS09], pushdown systems [MP09], etc.) and settings (centralized, decentralized, distributed), and have had an impact on important application areas, e.g. for telecommunication network failure diagnosis. Several contributions have considered enforcing the diagnosability of a system. Under the generic name of active diagnosis, the problems take quite different shapes. They range from the selection of minimal sets of observable labels that make the system diagnosable [CT08], to the design of controllers that select a diagnosable sublanguage of a system [SLT98], and to online aspects that either turn on and off sensors [TT07, CT08] or modify an action plan [CP09] in order to reduce the amount of ambiguity. Probabilistic systems have also received some attention [TT05, FJ10], with two essential motivations: determining the likelihood of a fault given an observed trace and defining diagnosability for probabilistic systems. Two definitions have been proposed: The A-diagnosability, which requires that the ambiguous traces have a null probability, and the weaker AA-diagnosability, which requires that fault likelihood will converge to one with probability one. Interestingly, the A-diagnosability does not depend on the specific values of transition probabilities, but only on their support: it is thus a structural property of a system, which can be checked in polynomial time on finite state systems.

Here we address the question of active diagnosis for stochastic systems. We elaborate on two recent contributions. The first one [HHMS13] improves the work in [SLT98] and designs an observation-based controller that enables a subset of actions in the system in order to make it diagnosable while preserving its liveness. Optimal constructions are then proposed the most relevant for our work being the characterization of unambiguous traces by a deterministic Büchi automaton with minimal size. The second one [BBG12] considers probabilistic Büchi automata, a subclass of partially observed Markov decision processes (POMDP), and proves that checking the existence of strategies that almost surely achieve a Büchi condition on POMDP is EXPTIME-complete. The result was later extended in [BGG09]. This motivates the use of POMDP as semantics for the models we consider.

The first contribution of this paper is a framework for the active diagnosis problem of probabilistic systems. The models we consider are weighted and labeled transition systems, where some transitions represent a fault. Some of the transition labels are observable, and similarly some are controllable. From a given state of the system, and given a set of enabled labels, one derives a transition probability by normalization of transition weights. The active diagnosis problem amounts to designing a label activation strategy that enforces the stochastic diagnosability of the system while preserving its liveness. As a second contribution, this problem is proved to be decidable, and EXPTIME complete. The resulting strategies are belief-based, i.e. they only depend on the set of possible states of the system given past observations, regardless of the exact values of transition weights. As a third contribution, we introduce and analyze the *safe* active diagnosis problem. It extends the active diagnosis by enforcing a positive probability of correct runs. In other words, this rules out strategies that would reach diagnosability only by enforcing the occurrence of a fault. We prove that safe active diagnosis may require non belief-based strategies, and that the existence of such strategies is an undecidable problem. This result refines the decidability/undecidability frontier for POMDP: the existence of a strategy simultaneously ensuring a Büchi condition almost-surely and a safety condition with positive probability is undecidable. This may seem surprising since the existence of strategies for each objective taken separately is decidable. As a last contribution, we prove that, restricted to belief-based strategies, the safe active diagnosis problem becomes decidable and belongs to NEXPTIME.

The paper is organized as follows: section 2 introduces the active diagnosis problem for probabilistic systems, and compares it with the state of the art. Section 3 proposes resolution techniques for active diagnosis. Section 4 analyzes the safe active diagnosis problem. Section 5 concludes this work. Due to lack of space, several proofs are provided in appendix.

2 The active diagnosis problem

This section recalls diagnosis problems from the literature, and formalizes the new problems we are interested in.

2.1 Passive (probabilistic) diagnosis

When dealing with stochastic discrete event systems diagnosis, systems are often modeled using labeled transition systems.

Definition 1. *A* probabilistic labeled transition system *(pLTS) is a tuple* $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P} \rangle$ *where:*

- Q *is a set of states with* $q_0 \in Q$ *the initial state;*
- Σ *is a finite set of events;*
- $T \subseteq Q \times \Sigma \times Q$ *is a set of transitions;*
- $\mathbf{P}$ *is the transition matrix from* T *to* $\mathbb{Q}_{\geq 0}$ *fulfilling for all* $q \in Q$*:* $\sum_{(q,a,q') \in T} \mathbf{P}[q, a, q'] = 1$.

Observe that a pLTS is a labeled transition system (LTS) equipped with transition probabilities. The transition relation of the underlying LTS is defined by: $q \xrightarrow{a} q'$ for $(q, a, q') \in T$; this transition is then said to be *enabled* in q. A *run* over the word $\sigma = a_1 a_2 \ldots \in \Sigma^\omega$ is a sequence of states $(q_i)_{i \geq 0}$ such that $q_i \xrightarrow{a_{i+1}} q_{i+1}$ for all $i \geq 0$, and we write $q_0 \stackrel{\sigma}{\Rightarrow}$ if such a run exists. A finite run over $w \in \Sigma^*$ is defined analogously, and we write $q \stackrel{w}{\Rightarrow} q'$ if such a run ends at state q'. A state q is *reachable* if there exists a run $q_0 \stackrel{w}{\Rightarrow} q$ for some $w \in \Sigma^*$. On the other hand, forgetting the labels and merging the transitions with same source and target, one obtains a discrete time Markov chain (DTMC).

Definition 2 (Languages of a pLTS). *Let* $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P} \rangle$ *be a pLTS. The finite language* $\mathcal{L}^*(\mathcal{A}) \subseteq \Sigma^*$ *of* $\mathcal{A}$ *and the infinite language* $\mathcal{L}^\omega(\mathcal{A}) \subseteq \Sigma^\omega$ *of* $\mathcal{A}$ *are defined by:*

$$\mathcal{L}^*(\mathcal{A}) = \{\, w \in \Sigma^* \mid \exists q : q_0 \stackrel{w}{\Rightarrow} q \,\} \qquad \mathcal{L}^\omega(\mathcal{A}) = \{\, \sigma \in \Sigma^\omega \mid q_0 \stackrel{\sigma}{\Rightarrow} \,\}$$

Observations. In order to formalize problems related to diagnosis, we partition Σ into two disjoint sets Σ_o and Σ_u, the sets of *observable* and of *unobservable events*, respectively. Moreover, we distinguish a special *fault* event $f \in \Sigma_u$. Let σ be a finite word; its length is denoted $|\sigma|$. For $\Sigma' \subseteq \Sigma$, define $\mathcal{P}_{\Sigma'}(\sigma)$, the projection of σ on Σ', inductively by: $\mathcal{P}_{\Sigma'}(\varepsilon) = \varepsilon$; for $a \in \Sigma'$, $\mathcal{P}_{\Sigma'}(\sigma a) = \mathcal{P}_{\Sigma'}(\sigma)a$; and $\mathcal{P}_{\Sigma'}(\sigma a) = \mathcal{P}_{\Sigma'}(\sigma)$ for $a \notin \Sigma'$. Write $|\sigma|_{\Sigma'}$ for $|\mathcal{P}_{\Sigma'}(\sigma)|$, and for $a \in \Sigma$, write $|\sigma|_a$ for $|\sigma|_{\{a\}}$. When σ is an infinite word, its projection is the limit of the projections of its finite prefixes. This projection can be either finite or infinite. As usual the projection is extended to languages. In the rest of the paper, we will only use $\mathcal{P}_{\Sigma_o}$, the projection onto observable events, and hence we will drop the subscript and simply write $\mathcal{P}$ instead of $\mathcal{P}_{\Sigma_o}$.

With respect to the partition of $\Sigma = \Sigma_o \uplus \Sigma_u$, a pLTS $\mathcal{A}$ is *convergent* if $\mathcal{L}^\omega(\mathcal{A}) \cap \Sigma^* \Sigma_u^\omega = \emptyset$ (i.e. there is no infinite sequence of unobservable events from any reachable state). When $\mathcal{A}$ is convergent, then for all $\sigma \in \mathcal{L}^\omega(\mathcal{A})$, one has $\mathcal{P}(\sigma) \in \Sigma_o^\omega$. In the rest of the paper we assume that pLTS are

convergent and we will call a *sequence* a finite or infinite word over Σ, and an *observed sequence* a finite or infinite sequence over Σ_o. Clearly, the projection of a sequence on Σ_o yields an observed sequence. Intuitively, a sequence describes the behavior of a system during an execution, and an observed sequence represents how such a run is perceived. Now, the role of diagnosis is to decide, for any observed sequence, whether a fault has occurred or not.

Ambiguity. A finite (resp. infinite) sequence σ is *correct* if it belongs to $(\Sigma \setminus \{f\})^*$ (resp. $(\Sigma \setminus \{f\})^\omega$). Otherwise σ is called *faulty*. A correct sequence and a faulty sequence may have the same observed projection, yielding ambiguity.

Definition 3 (Classification of observed sequences). *Let $\mathcal{A}$ be a pLTS. An observed sequence $\sigma \in \Sigma_o^\omega$ is called* ambiguous *if there exist two sequences $\sigma_1, \sigma_2 \in \mathcal{L}^\omega(\mathcal{A})$ such that $\mathcal{P}(\sigma_1) = \mathcal{P}(\sigma_2) = \sigma$, σ_1 is correct and σ_2 is faulty. An observed sequence $\sigma' \in \mathcal{P}(\mathcal{L}^\omega(\mathcal{A}))$ is* surely faulty *if $\mathcal{P}^{-1}(\sigma') \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq \Sigma^* f \Sigma^\omega$. An observed sequence $\sigma' \in \mathcal{P}(\mathcal{L}^\omega(\mathcal{A}))$ is* surely correct *if $\mathcal{P}^{-1}(\sigma') \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq (\Sigma \setminus \{f\})^\omega$. These notions are defined analogously for finite observed sequences.*

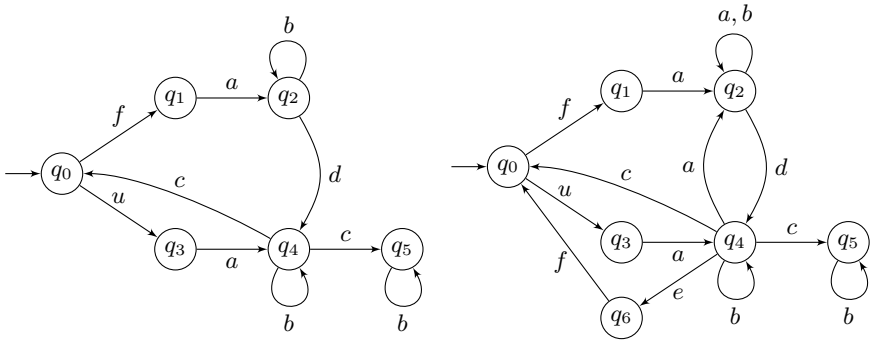

Fig. 1. Two examples of pLTS (cLTS), with $\Sigma_u = \{f, u\}$ and $\Sigma_o = \{a, b, c, d, e\}$.

Example. Consider the (convergent) pLTS to the left in Fig. 1, where $\Sigma_u = \{f, u\}$. We assume uniform distributions so we do not represent the probability matrix $\mathbf{P}$. This pLTS contains infinite ambiguous sequences: immediately after a is observed, an ambiguity appears, and this ambiguity remains in all infinite observed sequences without occurrence of d and finishing with ab^ω. Removing the loop at q_2 and/or q_4 makes all infinite ambiguous sequences disappear.

In the sequel, we will use the characterization of unambiguous sequences using deterministic Büchi automata [HHMS13].

Definition 4 (Büchi automaton). *A* Büchi automaton *over Σ is a tuple $\mathcal{B} = \langle Q, q_0, \Sigma, T, F\rangle$ with $\langle Q, q_0, \Sigma, T\rangle$ its underlying LTS and $F \subseteq Q$ an* acceptance condition*. A* run $(q_i)_{i\geq 0}$ *is* accepting *if $q_i \in F$ for infinitely many values of i. The language $\mathcal{L}(\mathcal{B})$ consists of all words in Σ^ω for which there exists an accepting run. A Büchi automaton is* deterministic *if for all q, a, $\{q' \mid q \xrightarrow{a} q'\}$ is either a singleton or the empty set.*

Theorem 1 ([HHMS13]). *Given a pLTS $\mathcal{A}$ with n states, one can build in exponential time a deterministic Büchi automaton $\mathcal{B}$ with $2^{O(n)}$ states whose language is the set of unambiguous sequences of $\mathcal{A}$.*

We briefly sketch the structure of $\mathcal{B}$. Its states are triples $\langle U, V, W\rangle$, where $U, V, W \subseteq Q$, $U \cup V \cup W \neq \emptyset$ and $V \cap W = \emptyset$, and its transitions are labeled by events from Σ_o, that is $\mathcal{B}$ recognizes observed sequences. The initial state of $\mathcal{B}$ is $\langle\{q_0\}, \emptyset, \emptyset\rangle$. Given an observed sequence σ reaching state $\langle U, V, W\rangle$, U is the set of states of $\mathcal{A}$ reached by a correct sequence with projection σ, and $V \cup W$ is the set of states of $\mathcal{A}$ reached by a faulty sequence with projection σ. When $U = \emptyset$, σ is the projection of faulty sequences of $\mathcal{A}$. The decomposition between V and W reflects the fact that $\mathcal{B}$ tries to "solve the ambiguity" between U and W (when both are non empty), while V corresponds to a waiting room of states reached by faulty sequences that will be examined when the current ambiguity is resolved. Given some new observation a, a transition from $\langle U, V, W\rangle$ to the new state $\langle U', V', W'\rangle$ is defined as follows. U' is the set of states reached from U by a correct sequence with projection a. Let Y be the set of states reached from U by a faulty sequence with projection a, or reached from V by a sequence with projection a. When W is non empty then W' is the set of states reached from W by a sequence with projection a and $V' = Y$. Otherwise, the faulty sequences ending in states memorized by W cannot be extended by a sequences with projection a, and we set $V' = \emptyset$ and $W' = Y$. The ambiguity between U and W has been resolved, but new ambiguity may arise between U' and W'. Accepting states in F are triples $\langle U, V, W\rangle$ with $U = \emptyset$ or $W = \emptyset$. Hence, all infinite observed sequence of $\mathcal{A}$ passing infinitely often through F are not ambiguous (they resolve ambiguities one after another) and are accepted by $\mathcal{B}$.

We are now in position to define diagnosability. It is well-known that given a pLTS $\mathcal{A}$ and a Büchi automaton $\mathcal{B}$, the set of sequences of $\mathcal{A}$ accepted by $\mathcal{B}$ is measurable [Var85]. So the following definition is sound.

Definition 5 (Diagnosability). *A pLTS $\mathcal{A}$ is* diagnosable *if the set of sequences yielding ambiguous observed sequences has null measure.*
It is safely diagnosable *if it is diagnosable and the set of correct sequences has positive measure.*

The notion of a safely diagnosable pLTS is introduced to ensure that fault occurrence is not almost sure. This property is important: a diagnosable system which is not safely diagnosable contains only faulty infinite runs. In the rest of the paper, we will consider active diagnosis, that is, ways to force a system to become diagnosable using a controller. If a controlled system is not safely diagnosable, then the diagnosis solution enforced by the controller is not acceptable.

Example. Consider again the pLTS to the left in Fig. 1. The only ambiguous observed (infinite) sequences necessarily terminate with ab^ω. But the probability to produce such a sequence is null, as the system will reach q_5 with probability one. In other words, ambiguity vanishes at the first occurrence of d or cb. Since cb occurs with probability one, this pLTS is diagnosable. This pLTS is also safely diagnosable, as it can produce correct sequences with a positive probability: there is a positive probability to reach q_5 by sequence uac. If one removes state q_5 and its connected transitions, the system remains diagnosable, but is not safely diagnosable anymore: as the graph of the pLTS is strongly connected, every transition will be visited (infinitely often) with probability 1 implying that f occurs.

2.2 Active probabilistic diagnosis

In order to allow control over the actions of a system while preserving the possibility of a probabilistic semantic, we introduce controllable weighted labelled transition system where probabilities are replaced by weights.

Definition 6. *A* controllable weighted labelled transition system *(cLTS) is a tuple $\mathcal{C} = \langle Q, q_0, \Sigma, T\rangle$ where:*

- *Q is a finite set of states with $q_0 \in Q$ the initial state;*
- *the event alphabet Σ is partitionned into observable Σ_o and unobservable Σ_u events, and also partitionned into controllable Σ_c and uncontrollable Σ_e (e for environment) events;*
- *$\Sigma_u = \{f, u\}$ contains a faulty event, and a non-faulty one;*
- *$T : S \times \Sigma \times S \to \mathbb{N}$ is the transition function, labelling transitions with integer weights.*

A cLTS has an underlying LTS where the transition relation is defined by $q \xrightarrow{a} q'$ if $T(q,a,q') > 0$. All previous definitions that do not depend on probabilities equally apply to cLTS. We denote by $\mathsf{Ena}(q)$ the set of events that are enabled in q: $\mathsf{Ena}(q) = \{a \in \Sigma \mid \exists q',\ T(q,a,q') > 0\}$. We assume that the cLTS is convergent and *live*: for all q, $\mathsf{Ena}(q) \neq \emptyset$.

Let $\mathcal{C} = \langle Q, q_0, \Sigma, T\rangle$ be a cLTS. For $q \in Q$ and $\Sigma^\bullet \subseteq \Sigma$, we define

$$G^{\Sigma^\bullet}(q) = \sum_{a \in \Sigma^\bullet,\ q' \in Q} T(q,a,q')$$

as the (possibly null) global outgoing weight from q restricted to $\Sigma^\bullet$-events. Similarly, we define a normalization of the transition relation restricted to $\Sigma^\bullet$ by

$$T^{\Sigma^\bullet}(q,a,q') = \begin{cases} \frac{T(q,a,q')}{G^{\Sigma^\bullet}(q)} \text{ if } a \in \Sigma^\bullet \text{ and } T(q,a,q') > 0 \\ 0 \text{ otherwise} \end{cases}$$

For a given finite set X, we define by $\mathsf{Dist(X)}$ the set of probabilistic distributions over X. Let $x \in X$, we denote by $\mathbf{1}_x$ the Dirac distribution on x. For a distribution $\delta \in \mathsf{Dist(X)}$, the support of δ is the set $\mathsf{Supp}(\delta) = \{x \in X \mid \delta(x) > 0\}$.

A *strategy* for a cLTS $\mathcal{C}$ is a mapping $\pi : \Sigma_o^* \to \mathsf{Dist}(2^\Sigma)$ such that for every $\sigma \in \Sigma_o^*$, for every $\Sigma' \in \mathsf{Supp}(\pi(\sigma))$, $\Sigma' \supseteq \Sigma_e$. A strategy consists in, given some observation, randomly choosing a subset of allowed events that includes the uncontrollable events. Given a cLTS $\mathcal{C}$ and a strategy π, we consider configurations of the form $(\sigma, q, \Sigma^\bullet) \in \Sigma_o^* \times Q \times 2^\Sigma$ where σ is the observed sequence, q is the current state and $\Sigma^\bullet$ is a set of events allowed by π after observing σ. We define inductively the set $\mathsf{Reach}_\pi(\mathcal{C})$ of reachable configurations under π:

- for all $\Sigma^\bullet \in \mathsf{Supp}(\pi(\varepsilon))$, $(\varepsilon, q_0, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$;
- for all $(\sigma, q, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, for all $a \in \Sigma_u \cap \Sigma^\bullet$, such that $q \xrightarrow{a} q'$ $(\sigma, q', \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, denoted $(\sigma, q, \Sigma^\bullet) \xrightarrow{a}_\pi (\sigma, q', \Sigma^\bullet)$;
- for all $(\sigma, q, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, for all $a \in \Sigma_o \cap \Sigma^\bullet$ such that $q \xrightarrow{a} q'$ and $\Sigma^{\bullet\prime} \in \mathsf{Supp}(\pi(\sigma a))$, $(\sigma a, q', \Sigma^{\bullet\prime}) \in \mathsf{Reach}_\pi(\mathcal{C})$, denoted $(\sigma, q, \Sigma^\bullet) \xrightarrow{a}_\pi (\sigma a, q', \Sigma^{\bullet\prime})$.

A strategy π is said to be *live* if for every configuration $(\sigma, q, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, $G^{\Sigma^\bullet}(q) \neq 0$. Live strategies are the only relevant strategies as the other strategies introduce deadlocks. We are now in position to introduce the semantics of a cLTS. It is defined w.r.t. to some live strategy π as a pLTS. Its set of states is $\mathsf{Reach}_\pi(\mathcal{C})$ with an initial state whose goal

is to randomly select w.r.t. π the initial control. The transition probabilities are defined by $T^{\Sigma^\bullet}$ accordingly to the current control $\Sigma^\bullet$ except that when an observable action occurs it must be combined with the random choice (w.r.t. π) of the next control.

Definition 7. *Let $\mathcal{C}$ be a CLTS and π be a live strategy, the pLTS $\mathcal{C}_\pi$ induced by strategy π on $\mathcal{C}$ is defined as $\mathcal{C}_\pi = \langle Q_\pi, \Sigma, q_{0\pi}, T_\pi, \mathbf{P}_\pi \rangle$ where:*

- $Q_\pi = \{q_{0\pi}\} \cup \mathsf{Reach}_\pi(\mathcal{C})$;
- *for all* $(\varepsilon, q_0, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, $(q_{0\pi}, u, (\varepsilon, q_0, \Sigma^\bullet)) \in T_\pi$;
- *for all* $(\sigma, q, \Sigma^\bullet), (\sigma', q', \Sigma^{\bullet\prime}) \in \mathsf{Reach}_\pi(\mathcal{C})$, $\big((\sigma, q, \Sigma^\bullet), a, (\sigma', q', \Sigma^{\bullet\prime})\big) \in T_\pi$ *iff* $(\sigma, q, \Sigma^\bullet) \xrightarrow{a}_\pi (\sigma', q', \Sigma^{\bullet\prime})$;
- *for all* $(\varepsilon, q_0, \Sigma^\bullet) \in \mathsf{Reach}_\pi(\mathcal{C})$, $\mathbf{P}_\pi(q_{0\pi}, u, (\varepsilon, q_0, \Sigma^\bullet)) = \pi(\varepsilon)(\Sigma^\bullet)$;
- *for all* $((\sigma, q, \Sigma^\bullet), a, (\sigma, q', \Sigma^\bullet)) \in T_\pi$, *for all* $a \in \Sigma_u \cap \Sigma^\bullet$, $\mathbf{P}_\pi\left((\sigma, q, \Sigma^\bullet), a, (\sigma, q', \Sigma^\bullet)\right) = T^{\Sigma^\bullet}(q, a, q')$;
- *for all* $\Big((\sigma, q, \Sigma^\bullet), a, (\sigma a, q', \Sigma^{\bullet\prime})\Big) \in T_\pi$, *for all* $a \in \Sigma_o \cap \Sigma^\bullet$, $\mathbf{P}_\pi\Big((\sigma, q, \Sigma^\bullet), a, (\sigma a, q', \Sigma^{\bullet\prime})\Big) = T^{\Sigma^\bullet}(q, a, q') \cdot \pi(\sigma.a)(\Sigma^{\bullet\prime})$.

We can now formalize the decision problems we are interested in.

Definition 8 ((Safe) Active probabilistic diagnosis). *Given a cLTS $\mathcal{C} = \langle Q, q_0, \Sigma, T \rangle$, the* active probabilistic diagnosis problem *asks, whether there exists a live strategy π in $\mathcal{C}$ such that the pLTS $\mathcal{C}_\pi$ is diagnosable. The* safe active probabilistic diagnosis problem *asks whether there exists a live strategy π in $\mathcal{C}$ such that the pLTS $\mathcal{C}_\pi$ is safely diagnosable. The* synthesis problems *consists in building a live strategy π in $\mathcal{C}$ such that the pLTS $\mathcal{C}_\pi$ is (safely) diagnosable.*

Example. Consider the cLTS to the right in Fig. 1 with all weights equal to 1 and $\Sigma_o = \Sigma_c$. Without control, the system is not diagnosable as the observed sequence $aadcb^\omega$ is ambiguous, and it has a positive probability. So the strategy should disable action a for each correct observed sequence ending by ab^*. In addition, if this strategy always forbids c, the system becomes diagnosable, but the occurrence of a fault is enforced: so it is not safely diagnosable. Alternatively, if the strategy always forbids e, the system becomes safely diagnosable, as we obtain a pLTS "weakly probabilistically bisimilar" to the one on the left in Fig. 1.

3 Analysis of the active probabilistic diagnosis problem

To solve the active probabilistic diagnosis problem, we reduce it to a decidable problem on POMDP: namely, the existence of a strategy ensuring a Büchi objective with probability one [BBG12, BGG09].

Definition 9 (POMDP). *A* partially observable *Markov decision process (POMDP) is a tuple* $\mathsf{M}_{\mathcal{C}} = \langle Q, q_0, \mathsf{Obs}, \mathsf{Act}, T\rangle$ *where*

- Q *is a finite set of states with* q_0 *the initial state;*
- $\mathsf{Obs} : Q \to \mathcal{O}$ *assigns an observation* $\mathsf{O} \in \mathcal{O}$ *to each state.*
- Act *is a finite set of actions;*
- $T : Q \times \mathsf{Act} \to \mathsf{Dist}(\mathsf{Q})$ *is a partial transition function. Letting* $\mathsf{Ena}(q) = \{a \in \mathsf{Act} \mid T(q,a)$ *is defined*$\}$*, we assume that:*
  - *for all* $q \in Q$, $\mathsf{Ena}(q) \neq \emptyset$*, and*
  - *whenever* $\mathsf{Obs}(q) = \mathsf{Obs}(q') = \mathsf{O}$*, then* $\mathsf{Ena}(q) = \mathsf{Ena}(q')$ *and slightly abusing our notation, we will denote by* $\mathsf{Ena}(\mathsf{O})$ *the set of events enabled in every state with observation* O*.*

A *decision rule* is an item of $\mathsf{Dist}(\mathsf{Act})$ that resolves non-determinism by randomization. A *strategy* maps histories of observations to decision rules. Formally, a strategy is a function $\pi : \mathcal{O}^+ \to \mathsf{Dist}(\mathsf{Act})$ such that for all $\mathsf{O}_1 \cdots \mathsf{O}_i$, $\mathsf{Supp}(\pi(\mathsf{O}_1 \cdots \mathsf{O}_i)) \subseteq \mathsf{Ena}(O_i)$. Given a strategy π and an initial distribution δ over states, a POMDP M becomes a stochastic process that can be represented by a possibly infinite pLTS denoted $\mathsf{M}(\pi)$. One denotes $\mathbb{P}^{\delta}_{\pi}(\mathsf{Ev})$ the probability that event Ev is realized in this process.

A *belief* is a subset of $\mathsf{Obs}^{-1}(\mathsf{O})$ for some observation O that corresponds to the possible reachable states w.r.t. some sequence of observations. The initial belief is $\{q_0\}$ and given a current belief B, a decision rule δ and a observation O, the belief $\Delta(B, (\delta, \mathsf{O}))$ obtained after δ has been applied and O has been observed is defined by: $\bigcup_{q \in B, a \in \mathsf{Supp}(\delta)} \mathsf{Supp}(T(q,a)) \cap \mathsf{Obs}^{-1}(\mathsf{O})$. A strategy which only depends on the current belief is called a *belief-based strategy.*

In order to provide a POMDP $\mathsf{M}_{\mathcal{C}}$ for the diagnosis problems of a cLTS $\mathcal{C}$, we face several difficulties. First, in a cLTS the observations are related to actions while in a POMDP they are related to states. Fortunately all the information related to ambiguity is included in the deterministic Büchi automaton described in section 2. Thus (with one exception) the states are pairs of a state of the Büchi automaton and a state of the cLTS. In $\mathcal{C}$, the control is performed by allowing a subset of events. Thus actions of $\mathsf{M}_{\mathcal{C}}$ are subset of events that includes the uncontrollable events. Given some control Σ', for defining the transition probability of $\mathsf{M}_{\mathcal{C}}$ from (l, q) to (l', q'), one must consider all paths in $\mathcal{C}$ labelled by events of Σ' from q to q' such that the last event (say b) is the single observable one. The probability of any such path is obtained by the product of the individual step probabilities. The latter are then defined by the normalization of weights w.r.t. Σ'. They cannot be infinite paths of unobservable events

due to the convergence of $\mathcal{C}$. However some path can reach a state where no event of Σ' is possible. In other words, the control Σ' applied in (l, q) has a non null probability to reach a deadlock (i.e. the chosen decision rule leads to a non live strategy for the cLTS). In order to capture this behaviour and to obtain a non defective probability distribution, we add an additional state `lost`, that corresponds to such deadlocks. The next definition formalizes our approach.

Definition 10. *The POMDP* $\mathsf{M}_{\mathcal{C}} = \langle Q^{\mathsf{M}}, q_0^{\mathsf{M}}, \mathsf{Obs}, \mathsf{Act}, T^{\mathsf{M}} \rangle$ *derived from a cLTS* $\mathcal{C} = \langle Q, q_0, \Sigma, T \rangle$ *and its associated deterministic Büchi automaton* $\mathcal{B} = \langle L, l_0, \Sigma_o, T^{\mathcal{B}}, F \rangle$ *is defined by:*

- $Q^{\mathsf{M}} = L \times Q \uplus \{\mathbf{lost}\}$ *with* $q_0^{\mathsf{M}} = ((l_0, q_0)$*;*
- *the set of observations is* $\mathcal{O} = L \cup \{\mathbf{lost}\}$*, with* $\mathsf{Obs}((l, q)) = l$ *and* $\mathsf{Obs}(\mathbf{lost}) = \mathbf{lost}$*;*
- $\mathsf{Act} = \{\Sigma' \mid \Sigma' \supseteq \Sigma_e\}$*;*
- *for all* $(l, q) \in Q^{\mathsf{M}}$ *and* $\Sigma' \in \mathsf{Act}$*,* $T^{\mathsf{M}}((l, q), \Sigma') = \mu$ *where:*
  - $\mu((l', q'))$ *is defined by:*

$$\sum_{\substack{l \xrightarrow{b} l' \\ b \in \Sigma' \cap \Sigma_o}} \sum_{\substack{q \xrightarrow{a_1} q_1 \cdots \xrightarrow{a_n} q_n \xrightarrow{b} q' \\ a_1 \cdots a_n \in \Sigma' \cap \Sigma_u}} T^{\Sigma'}(q, a_1, q_1) \cdot \Big(\prod_{i=1}^{n-1} T^{\Sigma'}(q_i, a_{i+1}, q_{i+1}) \Big) \cdot T^{\Sigma'}(q_n, b, q')$$

  - $\mu(\mathtt{lost})$ *is defined by:*

$$\sum_{\substack{q \xrightarrow{a_1} q_1 \cdots \xrightarrow{a_n} q_n \\ a_1 \cdots a_n \in \Sigma' \cap \Sigma_u \\ G^{\Sigma'}(q_n) = 0}} T^{\Sigma'}(q, a_1, q_1) \cdot \prod_{i=1}^{n-1} T^{\Sigma'}(q_i, a_{i+1}, q_{i+1})$$

- $T^{\mathsf{M}}(\mathtt{lost}, \Sigma') = \mathbf{1}_{\mathtt{lost}}$ *for all* $\Sigma' \in \mathsf{Act}$*.*

Given $\mathcal{C}$, the construction of the Büchi automaton $\mathcal{B}$ is performed in exponential time. The construction of $\mathsf{M}_{\mathcal{C}}$ is also done in exponential time. Indeed, there is an exponential blowup for Act but again w.r.t. $\mathcal{C}$. Finally, while the distributions μ of action effects are presented in the definition as sums over paths of $\mathcal{C}$, each one can be computed by a matrix inversion in polynomial time (as done in discrete time Markov chains).

The next lemma is a straightforward consequence of the properties of $\mathcal{B}$ and the above definition of $\mathsf{M}_{\mathcal{C}}$. Here we use LTL notations to denote sets of paths in a POMDP, such as $\Diamond$, $\Box$ and $\Box\Diamond$ for eventually, always and infinitely often respectively.

Lemma 1. *$\mathcal{C}$ is actively diagnosable if and only if there exists a strategy π in $\mathsf{M}_{\mathcal{C}}$ such that $\mathbb{P}^{q_0}_{\pi}(\mathsf{M}_{\mathcal{C}} \models \Box\Diamond(W = \emptyset \vee U = \emptyset)) = 1$.*

Moreover, $\mathcal{C}$ is safely actively diagnosable if and only if there exists a strategy π in $\mathsf{M}_{\mathcal{C}}$ such that $\mathbb{P}^{q_0}_{\pi}(\mathsf{M}_{\mathcal{C}} \models \Box\Diamond(W = \emptyset \vee U = \emptyset)) = 1$ and $\mathbb{P}^{q_0}_{\pi}(\mathsf{M}_{\mathcal{C}} \models \Box(U \neq \emptyset)) > 0$.

In the statement of Lemma 1, $W = \emptyset \vee U = \emptyset$ is a shorthand to denote the set of states $(\langle U, V, W\rangle, q)$ in $\mathsf{M}_{\mathcal{C}}$ such that either $W = \emptyset$ or $U = \emptyset$; similarly, $U \neq \emptyset$ represents the set of states $(\langle U, V, W\rangle, q)$ such that $U \neq \emptyset$. As a consequence of Lemma 1, the active diagnosis problem for controllable LTS reduces to the existence of an almost-sure winning strategy for a Büchi objective on some exponential size POMDP.

Theorem 2. *The active probabilistic diagnosis decision and synthesis problems are EXPTIME-complete. There exists a family $(\mathcal{C}_n)_{n\in\mathbb{N}}$ of actively diagnosable cLTS with the size of $\mathcal{C}_n$ in $O(n)$, and such that any winning strategy for $\mathsf{M}_{\mathcal{C}_n}$ diagnosable requires at least $2^{\Omega(n)}$ memory-states.*

The EXPTIME upper bound may seem surprising, since $\mathsf{M}_{\mathcal{C}}$ is exponential in the size of $\mathcal{C}$, and the procedure to decide whether there exists a strategy in a POMDP to ensure a Büchi objective with probability 1 is in EXPTIME, due to the use of beliefs. However, in the POMDP $\mathsf{M}_{\mathcal{C}}$ we consider, the information on the belief is already contained in the state $(\langle U, V, W\rangle, q)$, as $U \cup V \cup W$. Therfore, a second exponential blowup, due to the beliefs, is avoided and the active probabilistic diagnosis problem remains in EXPTIME.

4 Analysis of the safe active probabilistic diagnosis problem

As will be shown below, the status of the active diagnosis problem changes when the safety requirement is added. The next proposition highlights this difference and it is the basis for the undecidability result of Theorem 3.

Proposition 1. *There exists a cLTS which is safely actively diagnosable and such that all belief-based strategies are losing.*

Proof. Let us consider the cLTS of Figure 2 with $\Sigma_u = \{u, f\}$ and $\Sigma_e = \{u, f, c\}$, and where all weights are equal to 1.
Pick any sequence of positive integers $\{\alpha_i\}_{i\geq 1}$ such that $\prod_{i\geq 1} 1-2^{-\alpha_i} > 0$. Define $A = \{a\} \cup \Sigma_e$ and $\overline{A} = \{\overline{a}\} \cup \Sigma_e$. We claim that the strategy π that consists in selecting, after n observations, the n^{th} subset in the following sequence $A^{\alpha_1}\overline{A}A^{\alpha_2}\overline{A}\ldots$, is winning. Observe that after an observable sequence of length $i \leq \alpha_1$, the system is either after a faulty sequence in

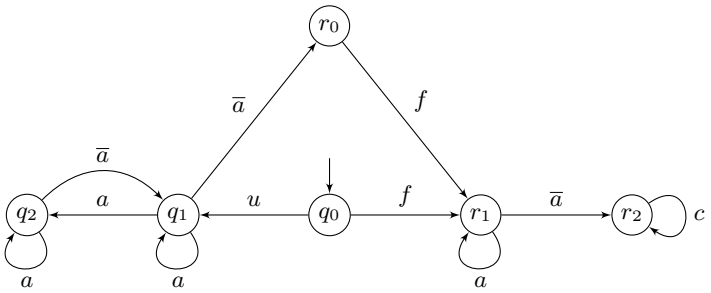

Fig. 2. A cLTS with only non belief-based strategies for safe diagnosis.

r_1 with probability $\frac{1}{2}$, or after a correct sequence in q_1 with probability 2^{-i-1}, or after a correct sequence in q_2 with probability $\frac{1}{2}(1-2^{-i})$. So, after an observable sequence of length α_1+1, the system is either after a faulty sequence in r_2 with probability $\frac{1}{2}$, or after a faulty sequence in r_1 (*via* r_0) with probability $2^{-\alpha_1-1}$, or after a correct sequence in q_1 with probability $\frac{1}{2}(1-2^{-\alpha_1})$. At the next step, the faulty sequence in r_2 is then detected by the occurrence of c.
Iterating this process we conclude that:

- any fault that may occur after π is applied up to $A^{\alpha_1}\overline{A}A^{\alpha_2}\overline{A}\dots A^{\alpha_i}\overline{A}$, is detected after π is applied up to $A^{\alpha_1}\overline{A}A^{\alpha_2}\overline{A}\dots A^{\alpha_{i+1}}\overline{A}A$. So the (full) strategy $\pi = A^{\alpha_1}\overline{A}A^{\alpha_2}\overline{A}\dots$ *surely* detects faults.
- the probability that there is an infinite correct sequence is equal to $\frac{1}{2}\prod_{i\geq 1} 1-2^{-\alpha_i} > 0$, due to our choice of the α_i's. Therefore, correct sequences have positive probability under π.

Consider a belief-based strategy π. There are three possible subsets of allowed events: A, $\overline{A}$ and Σ. The decision rule associated with belief $\{q_0\}$ must allow a in order to get the possibility of a correct sequence which, in case a occurs, leads to belief $\{q_1, q_2, r_1\}$. We should clarify here that beliefs do not correspond to the possible current states. They represent the possible states after the last observed event. For instance, when the belief is $\{q_0\}$, the current state may either be q_0, or q_1 after action u, or r_1 after fault f. Consider the (randomized) decision rule of π associated with belief $\{q_1, q_2, r_1\}$: $p_A \cdot A + p_{\overline{A}} \cdot \overline{A} + p_\Sigma \cdot \Sigma$ (denoted $\boldsymbol{p}$). If $p_A = 1$, then the possible first fault remains undetected, and π is losing. So $\overline{a}$ may occur leading to belief $\{q_1, r_0, r_2\}$.
Consider the decision rule of π associated with belief $\{q_1, r_0, r_2\}$: $p'_A \cdot A + p'_{\overline{A}} \cdot \overline{A} + p'_\Sigma \cdot \Sigma$ (denoted $\boldsymbol{p'}$). If $p'_{\overline{A}} = 1$, then at the next instant, there is no possible correct sequence, and π is losing.

So $p'_{\overline{A}} < 1$ and $p_A < 1$. Assume now that the current distribution of states is $\alpha q_1 + \beta r_0 + (1-\alpha-\beta) r_2$ (with belief $\{q_1, r_0, r_2\}$). The distribution after the next occurrence of $\overline{a}$ is defined by $\alpha_{\boldsymbol{p},\boldsymbol{p}'} \alpha q_1 + (1-\alpha_{\boldsymbol{p},\boldsymbol{p}'})\alpha r_0 + (1-\alpha) r_2$, where $\alpha_{\boldsymbol{p},\boldsymbol{p}'} < 1$ only depends on $\boldsymbol{p}$ and $\boldsymbol{p}'$. A correct sequence implies an infinite number of $\overline{a}$; after n occurrences of $\overline{a}$ the probability of a correct sequence is bounded by $\alpha^n_{\boldsymbol{p},\boldsymbol{p}'}$. So the probability of an infinite correct sequence is null, and π is losing. □

Theorem 3. *The safe active diagnosis problem for cLTS is undecidable.*

Proof (sketch). We perform a reduction from the following undecidable problem: given a blind POMDP and a set F of states, does there exist a strategy that ensures the Büchi objective $\Box\Diamond F$ with positive probability. The structure of the cLTS we construct is similar to the one of the example from Fig. 2, except that the states q_1 and q_2 are replaced with two copies of the POMDP. Consistently a and $\overline{a}$ are replaced by two copies of the alphabet of the POMDP with one of them bared. From F states in the first copy, with a non bared action one moves to the second one, and from any state, with bared actions, one moves back from the second copy to the first one, or moves from the first copy to r_0.

The following immediate corollary is interesting since both the existence of a strategy achieving a Büchi objective almost surely, and the existence of strategy achieving a safety objective with positive probability are decidable for POMDP [BGG09, CDGH10].

Corollary 1. *The problem whether, given a POMDP* M *with subsets of states* F *and* I*, there exists a strategy* π *with* $\mathbb{P}_\pi(\mathsf{M} \models \Box\Diamond F) = 1$ *and* $\mathbb{P}_\pi(\mathsf{M} \models \Box I) > 0$*, is undecidable.*

Given that the general safe active diagnosis problem is undecidable, and that belief-based strategies are not sufficient to achieve safe diagnosability, we consider now the restriction of the safe active diagnosis problem to belief-based strategies. Similarly to the case of active diagnosis, we reduce the safe active probabilistic diagnosis for belief-based-strategies to some verification question on POMDP.

Theorem 4. *The safe active probabilistic diagnosis problem restricted to belief-based strategies is in NEXPTIME and EXPTIME-hard.*

5 Conclusion

We studied the active diagnosis and safe active diagnosis problems for probabilistic discrete event systems, within a unifying POMDP framework. While the active diagnosis problem is EXPTIME-complete, the safe

active diagnosis problem is undecidable in general, and belongs to NEXPTIME when restricted to belief-based strategies. Since the lower and upper bounds do not coincide for the latter problem, we strive to close the gap between these bounds in future work. Another problem, closely related to diagnosability, is the predictability problem: given any observation, can we detect that the occurrence of a fault *before* it happens? Last, given the tight relation probabilistic diagnosis has with verification problems for POMDP, we plan to investigate further POMDP problems with multiple objectives.

References

[BBG08] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic büchi automata. In *Proceedings of FoSSaCS'08*, volume 4962 of *Lecture Notes in Computer Science*, pages 287–301. Springer, 2008.

[BBG12] C. Baier, N. Bertrand, and M. Grösser. Probabilistic ω-automata. *Journal of the ACM*, 59(1):1–52, 2012.

[BD08] D. Berwanger and L. Doyen. On the power of imperfect information. In *Proceedgins of FSTTCS'08*, volume 2 of *LIPIcs*, Bangalore, India, 2008. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

[BGG09] N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proceedings of LICS'09*, pages 319–328. IEEE Computer Society, 2009.

[CDGH10] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *Proceedings of MFCS'10*, volume 6281 of *Lecture Notes in Computer Science*, pages 246–257. Springer, 2010.

[CGLS09] M.P. Cabasino, A. Giua, S. Lafortune, and C. Seatzu. Diagnosability analysis of unbounded Petri nets. In *Proceedings of CDC'09*, pages 1267–1272. IEEE, 2009.

[CP09] E. Chanthery and Y. Pencolé. Monitoring and active diagnosis for discrete-event systems. In *Proceedings of SP'09*, pages 1545–1550. Elsevier, 2009.

[CT08] F. Cassez and S. Tripakis. Fault diagnosis with static and dynamic observers. *Fundamenta Informaticae*, 88:497–540, 2008.

[FJ10] E. Fabre and L. Jezequel. On the construction of probabilistic diagnosers. In *Proceeeding of WODES'10*, pages 229–234. Elsevier, 2010.

[HHMS13] S. Haar, S. Haddad, T. Melliti, and S. Schwoon. Optimal constructions for active diagnosis. In *Proceedings of FSTTCS'13*, volume 24 of *LIPIcs*, pages 527–539. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.

[MP09] C. Morvan and S. Pinchinat. Diagnosability of pushdown systems. In *Proceedings of HVC'09*, LNCS 6405, pages 21–33. Springer, 2009.

[SLT98] M. Sampath, S. Lafortune, and D. Teneketzis. Active diagnosis of discrete-event systems. *IEEE Transactions on Automatic Control*, 43(7):908–929, 1998.

[SSL+95] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *IEEE Trans. Aut. Cont.*, 40(9):1555–1575, 1995.

[TT05] D. Thorsley and D. Teneketzis. Diagnosability of stochastic discrete-event systems. *IEEE Transactions on Automatic Control*, 50(4):476–492, 2005.

[TT07] D. Thorsley and D. Teneketzis. Active acquisition of information for diagnosis and supervisory control of discrete-event systems. *Journal of Discrete Event Dynamic Systems*, 17:531–583, 2007.

[Var85] M. Y. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *Proceedings of FOCS'85*, pages 327–338. IEEE Computer Society Press, 1985.

[YL02] T-S. Yoo and S. Lafortune. Polynomial-time verification of diagnosability of partially observed discrete-event systems. *IEEE Trans. Automat. Contr.*, 47(9):1491–1495, 2002.

A Additional proofs

A.1 Proof of Theorem 2

EXPTIME upper-bound

Proposition 2. *The active probabilistic diagnosis problem is in EXPTIME.*

Proof. For the sake of completeness, and to justify the EXPTIME upper-bound, we recall the decision algorithm for POMDP with Büchi condition F. The correctness proof can be found in [BGG09], in the more general framework of 2-player stochastic games with signals. Given a POMDP $\mathsf{M} = \langle Q, q_0, \mathsf{Obs}, \mathsf{Act}, T\rangle$, we define its *belief automaton*:
Recall that a *belief* B for M is a non empty subset of states included in some observation $\mathsf{O} \in \mathsf{Obs}(Q)$. We write $\mathcal{B}l$ for the set of all beliefs, and we define the deterministic belief automaton $\mathcal{A}_{\mathcal{B}l}(\mathsf{M}) = \langle \mathcal{B}l, \{q_0\}, \mathsf{Act} \times \mathsf{Obs}, \Delta\rangle$ such that: for $B \in \mathcal{B}l$, $\alpha \in \mathsf{Act}$ and $\mathsf{O} \in \mathsf{Obs}(Q)$, $\Delta(B, (\alpha, \mathsf{O})) = \bigcup_{q\in B} \mathsf{Supp}(T(q, \alpha)) \cap \mathsf{O}$. In words, $\Delta(B, (\alpha, \mathsf{O}))$ updates the possible set of states the system is in, given the action that has been triggered and the observation that was made.
For a sequence of actions and observations $(\alpha_1, \mathsf{O}_1)\cdots(\alpha_n, \mathsf{O}_n)$, we write $\Delta(B, (\alpha_1, \mathsf{O}_1)\cdots(\alpha_n, \mathsf{O}_1))$ for $\Delta(\cdots\Delta(B, (\alpha_1, \mathsf{O}_1)), \cdots), (\alpha_n, \mathsf{O}_n))$.
For almost-sure Büchi objectives in POMDP, it was proven that belief-based strategies are sufficient, that is, there exists a strategy to achieve a given Büchi objective iff there exists a belief-based strategy for it. Building on the belief automaton, the set Win of beliefs from which there exists a winning strategy (to ensure the Büchi condition almost-surely), can be computed as a greatest fixpoint. Let $\mathcal{B}l_F = \{B \in \mathcal{B}l \mid B \subseteq F\}$. Then, Win is the limit of the non-increasing sequence that starts with $\mathsf{Win}_0 = \mathcal{B}l$ and is defined inductively by:

$$\mathsf{Win}_{n+1} = \{B \in \mathsf{Win}_n \mid \exists(\alpha_1, \mathsf{O}_{i_1})\cdots(\alpha_n, \mathsf{O}_{i_n}),\ \Delta\big(B, (\alpha_1, \mathsf{O}_{i_1})\cdots(\alpha_n, \mathsf{O}_{i_n})\big) \in \mathcal{B}l_F$$
$$\wedge\, \forall k,\ \forall \mathsf{O}_{j_k},\ \Delta\big(B, (\alpha_1, \mathsf{O}_1)\cdots(\alpha_k, \mathsf{O}_{j_k})\big) \neq \emptyset \Rightarrow \Delta\big(B, (\alpha_1, \mathsf{O}_1)\cdots(\alpha_k, \mathsf{O}_{x_k})\big) \in \mathsf{Win}_n\}.$$

Then, there exists a strategy to ensure the Büchi objective $\Box\Diamond F$ with probability 1, if and only if $\{q_0\} \in \mathsf{Win}$.

Note that this procedure is exponential w.r.t. the size of the input POMDP, due to the construction of the beliefs. However in the case of the POMDP $\mathsf{M}_{\mathcal{C}}$ we consider, the belief is already contained in the state (U, V, W, q), as $U \cup V \cup W$. Therefore, there is no exponential blowup due to

the resolution on the POMDP; the only exponential blowup comes from the Büchi automaton component, hence the active probabilistic diagnosis problem is in EXPTIME. □

EXPTIME-hardness The proof relies on a reduction from safety games with imperfect information [BD08] and it is adapted from an original proof in [HHMS13] in a non probabilistic context.

Proposition 3 (hardness). *The following problems are EXPTIME-hard.*

- *The existence of a winning strategy for the active diagnosis of a cLTS.*
- *The existence of a winning belief-based strategy for the safe active diagnosis of a cLTS.*

Proof. A safety game $\mathcal{G} = (L, l_0, \Sigma, \Delta, O, F, obs)$ with imperfect information is defined by:

- L a finite set of locations with $l_0 \in L$ the initial location;
- Σ a finite alphabet;
- $\Delta \subseteq L \times \Sigma \times L$ the transition relation such that for all $l \in L$ and $a \in \Sigma$ there exists at least one l' with $(l, a, l') \in \Delta$;
- O a finite set of observations with $F \subseteq O$ the final observations;
- $obs : L \mapsto O$ the observation mapping.

$\mathcal{G}$ is a turn-based game played by two players A and B. It starts in location l_0 with A to play. In the first round, A chooses a letter a_0 in Σ, and then B chooses a location l_1 such that $(l_0, a_0, l_1) \in \Delta$. A only observes $o_1 = obs(l_1)$. The next rounds are played similarly. Player A wins if for all i, $o_i \notin F$.

The problem of existence of a winning strategy for player A is EXPTIME-complete [BD08]. We now describe the reduction of this problem to diagnosis problems for a cLTS $\mathcal{C}$ defined as follows.

- Q, the set of states, is defined by $Q = L \uplus ((L \setminus obs^{-1}(F)) \times \Sigma) \uplus \{\bot\}$ and $q_0 = l_0$.
- The alphabet $\Sigma' = \Sigma \uplus O \uplus \{u, f, z\}$. The unobservable events are u and f and the uncontrollable events are $O \uplus \{u, f, z\}$.
- T the transition relation is defined as follows.
  1. For all $l \in L \setminus obs^{-1}(F)$ and $a \in \Sigma$, $T(l, a, (l, a)) = 1$.
  2. For all $l \in L \setminus obs^{-1}(F)$, $a \in \Sigma$ and $l' \in L$, $T((l, a), obs(l'), l') = 1$ if $(l, a, l') \in \Delta$.
  3. For all $l \in obs^{-1}(F)$, $T(l, u, \bot) = 1$ and $T(l, f, \bot) = 1$.

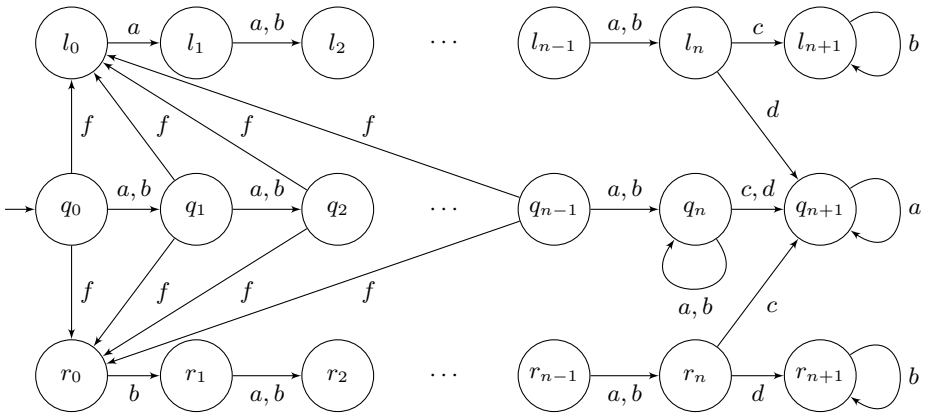

Fig. 3. A cLTS $\mathcal{C}_n$ with $\Sigma_o = \{a, b, c, d\}$, $\Sigma_c = \{c, d\}$ and weights are all one.

4. $T(\bot, z, \bot) = 1$.
5. All other weights are null.

From the very definition of $\mathcal{C}$, a sequence is ambiguous if and only if it contains an occurrence of z. Thus a strategy of the $\mathcal{C}$ is winning for the active diagnosis problem if and only if it avoids states $obs^{-1}(F)$. In addition, such a strategy only "controls" the subset of states $L \setminus obs^{-1}(F)$ and due to the assumptions on $\mathcal{G}$, it can safely restrict the allowed events to a single one. Furthermore since the information available to the strategy is exactly that of player A, a winning strategy for player A in $\mathcal{G}$ provides a winning strategy for the active diagnosis problem of $\mathcal{C}$ and vice versa.

In addition, a winning strategy for the active diagnosis problem of $\mathcal{C}$ only allows correct sequences. So it also solves the safe active diagnosis problem. Finally, it is known that in safety games with imperfect information if there is a winning strategy then there is a belief-based winning strategy. So the second problem of the proposition is also EXPTIME-hard. □

Optimality of belief-based strategies Again, the proof of the next proposition is adapted from an original proof in [HHMS13] in a non probabilistic context.

Proposition 4 (memory lower bound for strategy). *There exists a family $(\mathcal{C}_n)_{n\geq 1}$ of actively diagnosable cLTS with the size of $\mathcal{C}_n$ in $\mathcal{O}(n)$ such that any winning strategy has at least 2^n different memory states.*

Proof. The family of LTS $(\mathcal{C}_n)_{n\geq 1}$ is depicted in Figure 3, where $\Sigma_o = \{a, b, c, d\}$, $\Sigma_c = \{c, d\}$, and the initial state is q_0. Intuitively, during the n first steps a fault can occur leading to the upper (resp. lower) "branch" of the LTS when followed by a (resp. b).

Formally, let $\sigma = w_1 w_2 y a^\omega \in \Sigma_o^*$ be an observed sequence, where $w_1 w_2 \in \{a, b\}^*$, $1 \leq |w_1| \leq n$, $|w_2| = n - 1$, $y \in \{c, d\}$. Such a sequence has a positive probability to occur. Let $x_1 \cdots x_{|w_1|}$ be the letters of w_1. There are two possible execution sequences that have triggered $\sigma' = w_1 w_2 y$: the correct sequence σ' itself and the faulty sequence $x_1 \cdots x_{|w_1|-1} f x_{|w_1|} w_2 y$. If $x_{|w_1|} = a$, before the occurrence of y, the current state is q_n in the correct sequence and ℓ_n in the faulty sequence. So if $y = d$ the two sequences will lead to the same state q_{n+1} while if $y = c$ one sequence will lead to ℓ_{n+1} and the other one to q_{n+1} and they will be discriminated by the next observation. The case $x_{|w_1|} = b$ is symmetrical. So σ is ambiguous iff $x_{|w_1|} = a$ and $y = d$ or $x_{|w_1|} = b$ and $y = c$.

The LTS $\mathcal{C}_n$, is actively diagnosable. However assume that one observes a word $\sigma = a_1 \dots a_m \in \{a, b\}^*$ such that $n \leq m \leq 2n - 1$. Then when $a_{m-n+1} = a$, $\mathcal{C}$ may be in either q_n or l_n, and when $a_{m-n+1} = b$, $\mathcal{C}$ may be in either q_n or r_n. In the former case the controller must forbid d while in the latter it must forbid c. This implies that a winning strategy π must be in two different states after seeing two different words from $\{a, b\}^n$, therefore it must have at least 2^n states. □

A.2 Proof of Theorem 3

Proof. We proceed by establishing a reduction from the problem of achieving a positive probability in a blind POMDP with a Büchi objective. This problem was shown undecidable in [CDGH10] by establishing that pure strategies are enough for POMDP. Indeed, in the case of a blind POMDP, a pure strategy is an infinite word, and so the problem boils down to the undecidable problem of the existence of an infinite word achieving a positive probability in a Büchi probabilistic automaton [BBG08].

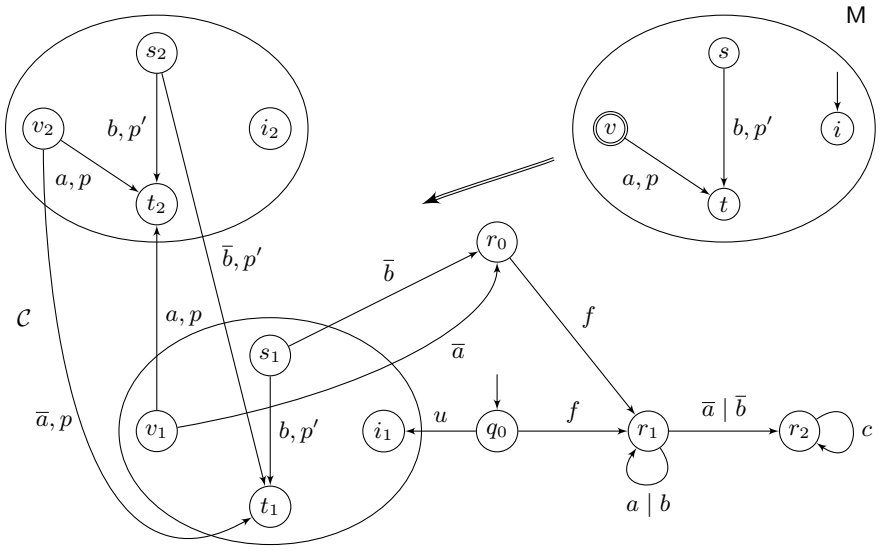

Fig. 4. A POMDP M and the derived cLTS $\mathcal{C}$ for the undecidability proof.

Here the cLTS has rational weights that could be transformed in integers by an appropriate scaling without changing the problem.
Let $\mathsf{M} = \langle Q^{\mathsf{M}}, i, \mathsf{Obs}, \mathsf{Act}, T^{\mathsf{M}} \rangle$ be a blind POMDP (that is $\mathsf{Obs}(Q^{\mathsf{M}})$ is a constant function) over the action alphabet $\mathsf{Act} = \{a, b\}$, and $F \subseteq Q^{\mathsf{M}}$ a set of states. From M one builds a cLTS $\mathcal{C} = \langle Q, q_0, \Sigma, T \rangle$ over $\Sigma = \{a, b, \overline{a}, \overline{b}, u, f, c\}$, and with $\Sigma_u = \{u, f\}$ and $\Sigma_e = \{u, f, c\}$. The construction is illustrated in Figure 4, where the POMDP is represented above right, with $v \in F$, and the cLTS below. Let us detail the reduction:

– The set of states is $Q = Q_1^{\mathsf{M}} \cup Q_2^{\mathsf{M}} \cup \{q_0, r_0, r_1, r_2\}$, where Q_1^{M} and Q_2^{M} are two copies of the states Q^{M} of the POMDP. The structure of

the CLTS in Figure 2 mimics that of the example of Figure 1. While in the example of Figure 1, the objective was to design a strategy increasing the probability to be in state q_2 when allowing action $\bar{a}$ to be safely diagnosable, the main objective for the CLTS of Figure 2 is to increase the probability for the system to be in a state of Q_2^{M} before allowing $\bar{a}$ or $\bar{b}$.

- The transitions from states in $\{q_0, r_0, r_1, r_2\}$ are fully depicted in the figure, with i_1 being the first copy of the initial state of the POMDP. The weights labelling these transitions are all equal to 1.
- For $s, t \in Q^{\mathsf{M}}$ and $x \in \{a, b\}$, we let $T(s_1, \overline{x}, r_0) = 1$, $T(s_2, x, t_2) = T^{\mathsf{M}}(s, x)(t)$ and $T(s_2, \overline{x}, t_1) = T^{\mathsf{M}}(s, x)(t)$.
- For $s \in S \setminus F$, $t \in S$ and $x \in \{a, b\}$, we let $T(s_1, x, t_1) = T^{\mathsf{M}}(s, x)(t)$.
- For $s \in F$, $t \in S$ and $x \in \{a, b\}$, we let $T(s_1, x, t_2) = T^{\mathsf{M}}(s, x)(t)$.
- All other weights are null.

This reduction ensures that there exists a strategy π in M such that $\mathbb{P}^i_\pi(\mathsf{M} \models \Box\Diamond F) > 0$ if and only if the cLTS $\mathcal{C}$ is safely active diagnosable. To prove it, we start with some preliminary remarks. A reasoning similar to the one in Proposition 1 implies that the faulty unambiguous sequences are those ending by $\overline{x}c^\omega$ with $x \in \{a, b\}$, while the correct unambiguous sequences are those belonging $\{a, b, \overline{a}, \overline{b}\}^\omega$ with an infinite number of occurrences of actions in $\{\overline{a}, \overline{b}\}$.

Assume first that there is a winning strategy in the blind POMDP M for the Büchi objective with positive probability. W.l.o.g., this strategy may be assumed to be pure [CDGH10], and thus can be given as an infinite word $\sigma = \sigma_1\sigma_2 \ldots \in \{a, b\}^\omega$. We write p for the probability of infinitely meeting F under σ: $p = \mathbb{P}^i_\sigma(\mathsf{M} \models \Box\Diamond F)$.
Pick some infinite sequence $(\beta_j)_{j\in\mathbb{N}}$ with $0 < \beta_j < 1$ and such that $\prod_{j\geq 0} \beta_j > 0$. We iteratively build an infinite increasing sequence $(n_j)_{j\in\mathbb{N}}$ of integers as follows. First $n_0 = 0$, and if $n_0, \ldots, n_j$ are defined, $n_{j+1} > n_j$ is set as the smallest integer that satisfies

$$\mathbb{P}^i_\sigma\Big(\mathsf{M} \models \Diamond^{[n_j+1, n_{j+1}]} F \mid \mathsf{M} \models \bigwedge_{k=0}^{i} \Diamond^{[n_k+1, n_{k+1}]} F \wedge \Box\Diamond F\Big) \geq \beta_j \ ,$$

where the notation $\Diamond^{[m,M]}F$ stands for the event "F is visited between the m-th and M-th time instants". Because σ is a winning strategy, and due to the induction hypothesis, the above conditional probability is well defined, and it tends to 1 as n_{j+1} goes to infinity. So n_{j+1} is well defined. By construction:

$$\mathbb{P}^i_\sigma\Big(\mathsf{M} \models \bigwedge_{j\geq 0} \Diamond^{[n_j+1, n_{j+1}]} F\Big) \geq p \prod_{j\geq 0} \beta_j > 0 \ .$$

We are now in a position to define a winning strategy π in the cLTS $\mathcal{C}$. Writing $X = \{x\} \cup \Sigma_e$ and $\overline{X} = \{\overline{x}\} \cup \Sigma_e$ for $x \in \{a, b\}$, at time instant k different from any n_j, π selects X with $x = \sigma_k$, and at time instant n_j, π selects $\overline{X}$ with $x = \sigma_{n_j}$. Due to the choice of time instants n_j, any sequence triggered by π is unambiguous. Furthermore, the probability that, for all j, at time instant n_j the current state is in Q_2^{M}, is at least $\frac{p}{2} \prod_{j \geq 0} \beta_j$. Thus with at least this probability, the random sequence in $\mathcal{C}$ generated by π will never leave $Q_1^{\mathsf{M}} \cup Q_2^{\mathsf{M}}$, and will be a correct sequence. Putting all together, π ensures the safe diagnosis of cLTS $\mathcal{C}$.

Assume now that there is a strategy π in the cLTS, that renders it safe diagnosable. Observe that every decision rule for $\mathcal{C}$ corresponds to a possibly randomized decision rule in M. For instance, if π chooses $\Sigma' = \{a, \overline{b}, u, o, f\}$ as set of enabled actions, then for a state in $Q_1^{\mathsf{M}} \cup Q_2^{\mathsf{M}}$ it corresponds to choosing with probability $\frac{1}{2}$ either a or $\overline{b}$ (due to the definition of weights).

Using the preliminary observations, the set Ex of executions that contain infinitely often actions in $\{\overline{a}, \overline{b}\}$ or an occurence of c, is exactly the set of unambiguous sequences and therefore has probability one under π. Let $\mathsf{Ex}' \subseteq \mathsf{Ex}$ be the subset of executions that contain infinitely often actions in $\{\overline{a}, \overline{b}\}$. By assumption on π, Ex' has positive probability and corresponds to executions that only visit states of $Q_1^{\mathsf{M}} \cup Q_2^{\mathsf{M}}$. Observe that such sequences visit infinitely often F_1.

One builds a winning strategy π' for M as follows. Recall that the POMDP is blind, and therefore a strategy in M can only base its decisions on the events that have happened thus far, not on the current state. Thus, π' will be defined on $\{a, b\}^*$. Define $\mathsf{proj}(x) = \mathsf{proj}(\overline{x}) = x$ for $x \in \{a, b\}$. Let $\sigma = \sigma_1 \dots \sigma_n \in \{a, b\}^*$ be a finite sequence such that $\mathbb{P}_\pi^{q_0}(\mathcal{C} \models \mathsf{proj}^{-1}(\sigma)) > 0$, and $\sigma' \in \mathsf{proj}^{-1}(\sigma)$. Intuitively, with positive probability, σ is the projection of an execution following π. We further define $p_{\sigma'} = \mathbb{P}_\pi^{q_0}(\mathcal{C} \models \rho = \sigma' \mid \mathsf{proj}(\rho) = \sigma)$ where ρ is a random sequence of length n. Then $\pi'(\sigma)$ is defined by $\pi'(\sigma) = \sum_{\sigma' \in \mathsf{proj}^{-1}(\sigma)} p_{\sigma'} \pi(\sigma')$, *i.e.* the appropriate weighted combination on decision rules applied in the cLTS. For σ such that $\mathbb{P}_\pi^{q_0}(\mathcal{C} \models \mathsf{proj}(\rho) = \sigma) = 0$, $\pi'(\sigma)$ is arbitrarily defined. By construction (and induction) one gets: $\mathbb{P}_{\pi'}^{i}(\mathsf{M} \models \rho = \sigma) = \sum_{\sigma' \in \mathsf{proj}^{-1}(\sigma)} \mathbb{P}_\pi^{q_0}(\mathcal{C} \models \rho = \sigma')$. Thus $\mathbb{P}_{\pi'}^{i}(\mathsf{M} \models \mathsf{proj}(\mathsf{Ex}')) = \mathbb{P}_\pi^{q_0}(\mathcal{C} \models \mathsf{Ex}')$ and $\mathbb{P}_{\pi'}^{i}(\mathsf{M} \models \mathsf{proj}(\mathsf{Ex}')) > 0$. On the other hand, any sequence of $\mathsf{proj}(\mathsf{Ex}')$ visits infinitely often F. As a consequence, π' is a winning strategy. □

A.3 Proof of Theorem 4

Proof. Here we solve the more general problem of the existence of a belief-based strategy in a POMDP that simultaneously achieves a Büchi condition F and a safety condition I. Let π be a belief-based strategy. We define a finite discrete-time Markov chain (DTMC) as follows.

- The set of states is $\{(q, B) \mid q \in Q \wedge q \in B \subseteq \mathsf{Obs}(q)\}$.
- The initial state is $(q_0, \{q_0\})$.
- The transition matrix is defined by:
  $\mathbf{P}[(q, B), (q', B')] = \sum_{a \in \mathsf{Act}} \pi(B)(a) T(q, a)(q')$
  if $B' = \{q'' \in \mathsf{Obs}(q') \mid \exists q^* \in B \ \sum_{a \in \mathsf{Act}} \pi(B)(a) T(q^*, a)(q'') > 0\}$
  and $\mathbf{P}[(q, B), (q', B')] = 0$ otherwise.

Assume that π is winning from the initial state q_0 for the almost-sure Büchi objective F and positive safe objective I. This property is equivalent to the fact that in the the underlying graph of the DTMC the following holds:

- every bottom strongly connected component (BSCC) reachable from $(q_0, \{q_0\})$ contains a state (q, B) with $B \subseteq F$,
- and there is a BSCC reachable from $(q_0, \{q_0\})$ by some path such that all the beliefs along this path and in the BSCC belong to I.

These properties only depend on the underlying graph of the DTMC which in turns only depends on the supports of the decisions of strategy π. Therefore a NEXPTIME procedure consists in guessing the supports of some belief-based strategy, building the underlying graph of the corresponding DTMC and checking the two previous properties.
The EXPTIME hardness has been proved in proposition 3. □

Foundation of Diagnosis and Predictability in Probabilistic Systems

Nathalie Bertrand, Serge Haddad,
Engel Lefaucheux

July 2014

Research report LSV-14-09 (Version 1)

Laboratoire Spécification & Vérification

École Normale Supérieure de Cachan
61, avenue du Président Wilson
94235 Cachan Cedex France

Foundation of Diagnosis and Predictability in Probabilistic Systems*

Nathalie Bertrand[1], Serge Haddad[2], and Engel Lefaucheux[1,2]

1 **Inria, France** nathalie.bertrand@inria.fr
2 **LSV, ENS Cachan & CNRS & Inria, France**
{serge.haddad,engel.lefaucheux}@ens-cachan.fr

Abstract

In discrete event systems prone to unobservable faults, a diagnoser must eventually detect fault occurrences. The diagnosability problem consists in deciding whether such a diagnoser exists. Here we investigate diagnosis issues in a probabilistic framework modelled by partially observed Markov chains (denoted pLTS). First we study different specifications of diagnosability and establish their relations both in finite and infinite pLTS. Then we analyze the complexity of the diagnosability problem for finite pLTS: we show that the polynomial time procedure earlier proposed is erroneous and that in fact for all considered specifications, the problem is PSPACE-complete. We also establish tight bounds for the size of diagnosers. Afterwards we consider the dual notion of predictability which consists in predicting that in a safe run, fault will eventually occur. Predictability is easier than diagnosability: it is NLOGSPACE-complete. Yet the predictor synthesis is as hard as the diagnoser synthesis. Finally we introduce and study the more flexible notion of *prediagnoser* that generalizes predictor and diagnoser.

1998 ACM Subject Classification D.2.5 Testing and Debugging

Keywords and phrases Partially observed systems – Diagnosis – Markov chains

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

Diagnosis. In computer science, diagnosis may refer to different kinds of activities. For instance, in artificial intelligence it can describe the process of identifying a desease from its symptoms, as performed by the expert system MYCIN [2]). In this work, we concentrate on diagnosis as studied in control theory, where it is applied to partially observable systems prone to faults. A sequence of observations of such a system is said to be surely correct (respectively surely faulty) if all possible runs corresponding to this sequence are correct (respectively faulty); otherwise the observed sequence is ambiguous. While monitoring the system, the *diagnoser* should rule out ambiguities, and in particular detect that a fault occurred; and the problem of existence of such a diagnoser is refered to as *diagnosability* [12]. In order to anticipate problems triggered by fault occurrences, one can also be interested into *predictors* that detect that fault will eventually occur, and the *predictability* problem [5] is concerned with the existence of a predictor.

Diagnosis of discrete event systems. Diagnosability and predictability were first defined and studied in the framework of finite discrete event systems modelled by labelled transition

* This work has been supported by project ImpRo ANR-2010-BLAN-0317 and the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement 257462 HYCON2 NOE.

Conference title on which this volume is based on.
Editors: Billy Editor and Bill Editors; pp. 1–43

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

systems (LTS), and the problems were shown to be solvable in PTIME (see [7] and [5], respectively). Despite the polynomial time complexity of the decision problems, for diagnosable (respectively predictable) LTS, the size of the diagnoser (respectively predictor) contructed by the algorithms may be exponential. Diagnosers as well as predictors must ensure two requirements: *correctness*, meaning that the information provided by the diagnoser/predictor is accurate, and *reactivity*, ensuring that a fault will eventually be detected.

Diagnosis of probabilistic sytems. Building on the work for LTS, the notion of diagnosability was later extended to Markov chains with labelled transitions, also called probabilitic labelled transition systems (pLTS) [13]. In a probabilistic context, the reactivity requirement now requires that a fault will be almost surely eventually detected. Regarding correctness, two specifications have been proposed: either one sticks to the original definition and requires that the provided information is accurate, defining *A-diagnosability*; or one weakens the correctness by admitting errors in the provided information that should, however, have an arbitrary small probability when the delay before the diagnostic is long enough, defining *AA-diagnosability*. From a computational viewpoint, PTIME algorithms have been proposed to solve these two specifications of probabilistic diagnosability [3].

In case a system is not diagnosable, one may be able to control it, by forbidding some controllable actions, so that is becomes diagnosable. This property of *active diagnosability* has been studied for probabilistic systems in [1] pursuing the work of [11, 6] for discrete-event systems. Decidability and complexity issues are considered and optimal size diagnosers are synthesized. Interestingly the notion of diagnosability in [1] slightly differs from the original one.

Remaining issues. Some issues remained untouched in the above line of work. First, diagnosability was only considered with respect to finite faulty runs. It seems as important to also consider diagnosability of correct runs, and ambiguity can also be defined for infinite computations. Second, in most work, the complexity of the varied diagnosability problems and of the diagnosers synthesis were left open. Moreover, optimizing the delay between the fault occurrence and its detection is an important issue. Yet the search for diagnosers (or predictors) with optimal reactivity was not even considered. Last predictability and diagnosability were independently studied while combining them is obviously a fruitful direction.

Contributions. In this paper, we address the above mentioned gaps, and revisit diagnosability and predictability for probabilistic systems, from a semantical as well as a computational perspectives.

- In order to give a firm semantical classification of diagnosability notions, we define criteria for diagnosability in probabilistic systems, depending on (1) whether the information provided by a diagnoser is related to faulty runs only or to all runs and, (2) whether ambiguity is defined at the level of infinite runs, or for longer and longer finite subruns. These two dimensions yield three main specifications: FF-diagnosability defined in [13] and named A-diagnosability, IA-diagnosability used in [1] and FA-diagnosability, and we establish the connections between them.
- For finite state probabilistic systems, we show that these three notions of diagnosability can be characterized based on deterministic (finite or Büchi) automata acting as *monitors*, and synchronized with the pLTS. We further prove that the complexity of the diagnosability problem (for all three specifications) is PSPACE-complete, contradicting the polynomial time result for FF-diagnosability [3], and we indeed establish that their algorithm is wrong.
- Afterwards, we design algorithms for the synthesis of finite-memory diagnosers and prove

that their size $2^{\Theta(n)}$ (where n is the number of states of the pLTS model) is optimal.

- Since predictability is an interesting alternative to diagnosability, we introduce two possible specifications for predictability in probabilistic systems, and show that in both cases the predictability problem is NLOGSPACE-complete. Yet, as for diagnosers, the optimal size of predictors is also in $2^{\Theta(n)}$.
- Last, we introduce and study *prediagnosability* that combines the benefits of predictability and diagnosability: depending on the current observation, a prediagnoser behaves either as a diagnoser or as a predictor. Prediagnosability is of interest since generally predictability is more difficult to achieve than diagnosability, also prediagnosers can be seen as as soon as possible diagnosers. For the varied notions of prediagnosability we define, we establish that the prediagnosability problem is PSPACE-complete and design prediagnosers with optimal size.
- Summarizing we provide a full picture of the hierarchy for the different notions and the border between NLOGSPACE and PSPACE-complete problems.

Organization. In Section 2, we introduce probabilistic LTS, define the possible diagnosability specifications, establish their connection, and provide characterizations. In Section 3, we determine the exact complexity of the diagnosability problems. In Section 4, we design algorithmes for synthesis of diagnosers with optimal size. In Section 5, we study predictability and prediagnosis, and focus on optimal diagnosers. Finally in Appendix, we show that the algorithms in [3] are erroneous.

2 Diagnosability specification

2.1 Probabilistic labelled transition systems

In the context of stochastic discrete event systems diagnosis, systems are often modeled using a labeled transition systems.

▶ **Definition 1.** A *probabilistic labeled transition system* (pLTS) is a tuple $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$ where:

- Q is a set of states with $q_0 \in Q$ the initial state;
- Σ is a finite set of events;
- $T \subseteq Q \times \Sigma \times Q$ is a set of transitions;
- $\mathbf{P}$ is the transition matrix from T to $\mathbb{Q}_{>0}$ fulfilling for all $q \in Q$: $\sum_{(q,a,q')\in T} \mathbf{P}[q,a,q'] = 1$.

Observe that a pLTS is a labeled transition system (LTS) equipped with transition probabilities. The transition relation of the underlying LTS is defined by: $q \xrightarrow{a} q'$ for $(q,a,q') \in T$; this transition is then said to be *enabled* in q. A pLTS is said to be *live* if in every state q of the pLTS, a transition is enabled. We assume the pLTS we consider are a countably branching, *i.e.*, in every state q, only countably many transitions are enabled, so that the summation $\sum_{(q,a,q')\in T} \mathbf{P}[q,a,q']$ is well-defined.

Let us now introduce some important notions and notations that will be used throughout the paper. A *run* ρ of a pLTS $\mathcal{A}$ is a (finite or infinite) sequence $\rho = q_0a_0q_1\dots$ such that for all i, $q_i \in Q$, $a_i \in \Sigma$ and when q_{i+1} is defined, $q_i \xrightarrow{a_i} q_{i+1}$. The notion of run can be generalized, starting from an arbitrary state q. We write Ω for the set of all infinite runs of $\mathcal{A}$ starting from q_0, assuming the pLTS is clear from context. When it is finite, ρ ends in a state q and its *length*, denoted $|\rho|$, is the number of actions occurring in it. Given a finite run $\rho = q_0a_0q_1\dots q_n$ and a (finite or infinite) run $\rho' = q_na_nq_{n+1}\dots$, we call concatenation of

ρ and ρ' and we write $\rho\rho'$ the run $q_0a_0q_1\dots q_na_nq_{n+1}\dots$; the run ρ is then a *prefix* of $\rho\rho'$, which we denote $\rho \preceq \rho\rho'$. The *cylinder* defined by a finite run ρ is the set of all infinite runs that extend ρ: $C(\rho) = \{\rho' \in \Omega \mid \rho \preceq \rho'\}$. The sequence associated with $\rho = qa_0q_1\dots$ is the word $\sigma_\rho = a_0a_1\dots$, and we write indifferently $q \overset{\rho}{\Rightarrow}$ or $q \overset{\sigma_\rho}{\Longrightarrow}$ (resp. $q \overset{\rho}{\Rightarrow} q'$ or $q \overset{\sigma_\rho}{\Longrightarrow} q'$) for an infinite (resp. finite) run ρ. A state q is *reachable* (from q_0) if there exists a run such that $q_0 \overset{\rho}{\Rightarrow} q$, which we alternatively write $q_0 \Rightarrow q$. The language of pLTS $\mathcal{A}$ consists of all infinite words that label runs of $\mathcal{A}$ and is formally defined as $\mathcal{L}^\omega(\mathcal{A}) = \{\, \sigma \in \Sigma^\omega \mid q_0 \overset{\sigma}{\Rightarrow} \,\}$.

Forgetting the labels and merging (and summing the probabilities of) the transitions with same source and target, a pLTS yields a discrete time Markov chain (DTMC). As usual for DTMC, the set of infinite runs of $\mathcal{A}$ is the support of a probability measure defined by Caratheodory's extension theorem from the probabilities of the cylinders:

$$\mathbb{P}(C(q_0a_0q_1\dots q_n)) = \mathbf{P}[q_0, a_1, q_1]\cdots\mathbf{P}[q_{n-1}, a_{n-1}, q_n] \ .$$

2.2 Partial observation and ambiguity

In order to formalize problems related to fault diagnosis, we partition Σ into two disjoint sets Σ_o and Σ_u, the sets of *observable* and of *unobservable events*, respectively. Moreover, we distinguish a special *fault* event $\mathbf{f} \in \Sigma_u$. Let σ be a finite word; its length is denoted $|\sigma|$. The projection of σ onto Σ_o is defined inductively by: $\mathcal{P}(\varepsilon) = \varepsilon$; for $a \in \Sigma_o$, $\mathcal{P}(\sigma a) = \mathcal{P}(\sigma)a$; and $\mathcal{P}(\sigma a) = \mathcal{P}(\sigma)$ for $a \notin \Sigma_o$. Write $|\sigma|_o$ for $|\mathcal{P}(\sigma)|$. When σ is an infinite word, its projection is the limit of the projections of its finite prefixes. This projection is applicable to runs via their associated sequence; it can be either finite or infinite. As usual the projection is extended to languages. With respect to the partition of $\Sigma = \Sigma_o \uplus \Sigma_u$, a pLTS $\mathcal{A}$ is *convergent* if there is no infinite sequence of unobservable events from any reachable state: $\mathcal{L}^\omega(\mathcal{A}) \cap \Sigma^*\Sigma_u^\omega = \emptyset$. When $\mathcal{A}$ is convergent, for every $\sigma \in \mathcal{L}^\omega(\mathcal{A})$, $\mathcal{P}(\sigma) \in \Sigma_o^\omega$. In the rest of the paper we assume that pLTS are convergent. We will refer to a *sequence* for a finite or infinite word over Σ, and an *observed sequence* for a finite or infinite sequence over Σ_o. Clearly, the projection onto Σ_o of a sequence yields an observed sequence.

The *observable length* of a run ρ denoted $|\rho|_o \in \mathbb{N} \cup \{\infty\}$, is the number of observable actions that occur in it: $|\rho_o| = |\sigma_\rho|_o$. A *signalling* run is a finite run whose last action is observable. Signalling runs are precisely the relevant runs w.r.t. partial observation issues since each observable event provides an additional information about the execution to an external observer. In the sequel, SR denotes the set of signalling runs, and SR_n the set of signalling runs of observable length n. Since we assume that the pLTS are convergent, for all $n > 0$, SR_n is equipped with a probability distribution defined by assigning measure $\mathbb{P}(\rho)$ to each $\rho \in \mathsf{SR}_n$. Given ρ a finite or infinite run, and $n \leq |\rho|_o$, $\rho_{\downarrow n}$ denotes the signalling subrun of ρ of observable length n. For convenience, we consider the empty run q_0 to be the single signalling run, of null length.

Let $\mathcal{A}$ be a pLTS. A run ρ is *faulty* if σ_ρ contains $\mathbf{f}$, otherwise it is *correct*. W.l.o.g., by considering two copies of each state, we assume that the states are partitioned into correct states and faulty states: $Q = Q_f \uplus Q_c$ where Q_f are faulty states, and Q_c correct states. Faulty (resp. correct) states are only reachable by faulty (resp. correct) runs. An observed sequence $\sigma \in \Sigma_o^\omega$ is *surely correct* if $\mathcal{P}^{-1}(\sigma) \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq (\Sigma \setminus \mathbf{f})^\omega$; it is *surely faulty* if $\mathcal{P}^{-1}(\sigma) \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq \Sigma^*\mathbf{f}\Sigma^\omega$; otherwise, it is *ambiguous*. For finite sequences, we need to rely on signalling runs: a finite observed sequence $\sigma \in \Sigma_o^*$ is *surely faulty* (resp. *surely correct*) if for every signalling run ρ with $\mathcal{P}(\sigma_\rho) = \sigma$, ρ is faulty (resp. correct); otherwise it is ambiguous. A (finite signalling or infinite) run ρ is *surely faulty* (resp. *surely correct*, *ambiguous*) if $\mathcal{P}(\rho)$ is surely faulty (resp. surely correct, ambiguous).

We introduce different kinds of diagnosability and study their connection. In order to do so we introduce different subsets of infinite runs.

▶ **Definition 2** (Ambiguous runs). Let $\mathcal{A}$ be a pLTS and $n \in \mathbb{N}$ with $n \geq 1$. Then:

- FAmb_∞ is the set of infinite faulty ambiguous runs of $\mathcal{A}$;
- CAmb_∞ is the set of infinite correct ambiguous runs of $\mathcal{A}$;
- FAmb_n is the set of infinite runs of $\mathcal{A}$ whose signalling subrun of observable length n is faulty and ambiguous;
- CAmb_n is the set of infinite runs of $\mathcal{A}$ whose signalling subrun of observable length n is correct and ambiguous.

2.3 Which diagnosis for pLTS?

We propose four possible specifications of diagnosability for probabilistic systems. There are two discriminating criteria: whether the non ambiguity requirement holds for faulty runs only or for all runs, and whether ambiguity is defined at the infinite run level or for longer and longer finite signalling subruns.

▶ **Definition 3** (Diagnosability specifications). Let $\mathcal{A}$ be a pLTS. Then:

- A pLTS $\mathcal{A}$ is IF-diagnosable if $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$.
- A pLTS $\mathcal{A}$ is IA-diagnosable if $\mathbb{P}(\mathsf{FAmb}_\infty \uplus \mathsf{CAmb}_\infty) = 0$.
- A pLTS $\mathcal{A}$ is FF-diagnosable if $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) = 0$.
- A pLTS $\mathcal{A}$ is FA-diagnosable if $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \uplus \mathsf{CAmb}_n) = 0$.

The next theorem summarizes the connections between these definitions.

▶ **Theorem 4.** *The different kinds of diagnosability for pLTS are related according to the following table. Moreover, the implications hold for infinite-state pLTS while the non implications hold already for finite-state pLTS.*

Diagnosability	*All runs*		*Faulty runs*
Signalling runs	FA	$\Rightarrow$ $\not\Leftarrow$	FF
	$\Downarrow\not\Uparrow$		$\Downarrow\Uparrow^*$
Infinite runs	IA	$\Rightarrow$ $\not\Leftarrow$	IF

The implication marked with * requires pLTS to be finitely branching.

The rest of this section is devoted to the proof of Theorem 4. It relies on the forthcoming observations, lemmas and counter-examples. First of all, the implications from left to right are immediate by definition.

In all following examples, when the probabilities are omitted, we assume an equidistribution between outgoing edges of a state. Consider the pLTS $\mathcal{A}$ of Figure 1 where $\{u, \mathbf{f}\}$ is the set of unobservable events. A faulty run will almost surely produce a b-event that cannot be mimicked by the single correct run. Thus this pLTS is A-diagnosable. The unique correct run $\rho = q_0 u q_1 a q_1 \dots$ has probability $\frac{1}{2}$ and its corresponding observed sequence a^ω is ambiguous. Thus the pLTS is not AS-diagnosable. This simple example shows that, already for finite-state pLTS, A-diagnosability does not imply AS-diagnosability.

We now focus on "vertical" implications. To study them, let us introduce new subsets of infinite runs. C_∞ is the set of correct runs (ambiguous or not) and Sf_∞ the set of surely faulty runs. In addition C_n (resp. Sf_n) is the set of infinite runs whose signalling subrun of observable length n is correct (resp. surely faulty).

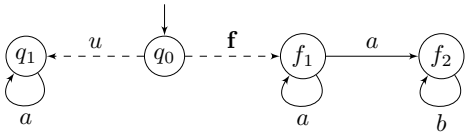

Figure 1 A pLTS which is IF-diagnosable but not IA-diagnosable.

▶ **Lemma 5.** *Let $\mathcal{A}$ be a pLTS. Then:*

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0$$

If $\mathcal{A}$ is finitely branching then:

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty) = 0$$

Proof. Observe that Ω admits the following partitions $\Omega = \mathsf{FAmb}_\infty \uplus \mathsf{C}_\infty \uplus \mathsf{Sf}_\infty$ and for every $n \in \mathbb{N}$, $\Omega = \mathsf{FAmb}_n \uplus \mathsf{C}_n \uplus \mathsf{Sf}_n$. Thus, for every $n \in \mathbb{N}$,

$$\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n = (\mathsf{C}_n \uplus \mathsf{Sf}_n) \cap \mathsf{FAmb}_\infty = (\mathsf{C}_n \uplus \mathsf{Sf}_n) \setminus (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \subseteq (\mathsf{C}_n \setminus \mathsf{C}_\infty) \uplus (\mathsf{Sf}_n \setminus \mathsf{Sf}_\infty) \ .$$

Since for all n, $\mathsf{Sf}_n \subseteq \mathsf{Sf}_\infty$, one gets:

$$\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n \subseteq \mathsf{C}_n \setminus \mathsf{C}_\infty \ .$$

$\{\mathsf{C}_n\}_{n\in\mathbb{N}}$ is a non increasing family of sets and we claim that $\mathsf{C}_\infty = \bigcap_{n\in\mathbb{N}} \mathsf{C}_n$. Indeed an infinite run ρ is correct if and only if f does not occur in it if and only if all its signalling subruns are correct. Thus,

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{C}_n \setminus \mathsf{C}_\infty) = 0 \text{ implying } \lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0 \ .$$

Using again the two partitions we obtain:

$$\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty = (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \cap \mathsf{FAmb}_n = (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \setminus (\mathsf{C}_n \uplus \mathsf{Sf}_n) \subseteq (\mathsf{C}_\infty \setminus \mathsf{C}_n) \uplus (\mathsf{Sf}_\infty \setminus \mathsf{Sf}_n)$$

Since for all n, $\mathsf{C}_\infty \subseteq \mathsf{C}_n$, one gets:

$$\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty \subseteq \mathsf{Sf}_\infty \setminus \mathsf{Sf}_n$$

Let us show that, under the assumption that $\mathcal{A}$ is finitely branching, then $\mathsf{Sf}_\infty \subseteq \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. Let $\rho \notin \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. We build a tree as follows:

- Nodes at level n correspond to the correct signalling runs whose observed sequence is $\mathcal{P}(\rho_{\downarrow n})$;
- The node associated with ρ' is a child of ρ'' if $\rho'' \preceq \rho'$.

Because $\rho \notin \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$, for every $n \in \mathbb{N}$, there exist a correct run with observed sequence $\mathcal{P}(\rho_{\downarrow n})$, so that the above-defined tree is infinite. Since the pLTS is finitely branching and convergent, the tree is also finitely branching. By König's lemma, it must contain an infinite branch, thus there exists an infinite correct run whose observed sequence is $\mathcal{P}(\rho)$. As a consequence ρ is not surely faulty: $\rho \notin \mathsf{Sf}_\infty$. This proves that $\mathsf{Sf}_\infty \subseteq \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. Thus:

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{Sf}_\infty \setminus \mathsf{Sf}_n) = 0 \text{ implying } \lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty) = 0$$

which concludes the proof. ◀

So we immediately get the following relation between IF-diagnosability and FF-diagnosability.

▶ **Corollary 6.** *A pLTS $\mathcal{A}$ is* IF*-diagnosable if it is* FF*-diagnosable. The converse implication holds when $\mathcal{A}$ is finitely branching.*

Note that the restriction to finitely branching pLTS is necessary. It is also necessary for IA-diagnosability to imply FF-diagnosability. Indeed, consider the infinitely branching pLTS of Figure 2, in which the probabilities on outgoing transitions from q_0 are $1/2$ to f_1 and $1/2^{i+1}$ to q_{i1}. It contains no infinite ambiguous sequence, so that it is IA-diagnosable, and thus IF-diagnosable. Yet, for every $n \in \mathbb{N}$, FAmb_n has probability $1/2$, so that it is not FF-diagnosable.

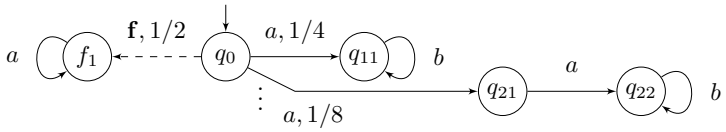

Figure 2 An infinitely branching pLTS that is IA- and IF-diagnosable but not FF-diagnosable.

▶ **Lemma 7.** *A pLTS $\mathcal{A}$ is* IA*-diagnosable if it is* FA*-diagnosable.*

Proof. In this proof, we use another subset of infinite runs: for $n \in \mathbb{N}$, $\mathsf{CAmb}_{n,\infty}$ is the set of correct ambiguous runs that admit an observationally equivalent run which is faulty before its n^{th} observable event. Observe that $\{\mathsf{CAmb}_{n,\infty}\}_{n\in\mathbb{N}}$ is a non decreasing sequence of sets and we claim that $\mathsf{CAmb}_\infty = \bigcup_{n\in\mathbb{N}} \mathsf{CAmb}_{n,\infty}$. Indeed, an infinite run ρ is correct and ambiguous if and only if there is a faulty run whose observed sequence is $\mathcal{P}(\rho)$ if and only if there exists some $n \in \mathbb{N}$ and some run ρ' such that $\rho'_{\downarrow n}$ is faulty and such that $\mathcal{P}(\rho') = \mathcal{P}(\rho)$. Morevover, by definition, $\mathsf{CAmb}_{n,\infty} \subseteq \mathsf{CAmb}_n$. Assume that $\limsup_{n\to\infty} \mathbb{P}(\mathsf{CAmb}_n) = 0$. For every $\varepsilon > 0$, there exists $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$, $\mathbb{P}(\mathsf{CAmb}_n) < \varepsilon$ and thus $\mathbb{P}(\mathsf{CAmb}_{n,\infty}) < \varepsilon$. On the other hand, there exists $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$, $\mathbb{P}(\mathsf{CAmb}_\infty) - \mathbb{P}(\mathsf{CAmb}_{n,\infty}) < \varepsilon$. Combining these two inequalities for $n = \max(n_1, n_2)$, one obtains $\mathbb{P}(\mathsf{CAmb}_\infty) < 2\varepsilon$, which concludes the proof. ◀

Let us look at the pLTS of Figure 3 where $\{u, \mathbf{f}\}$ is the set of unobservable events. Any infinite faulty run will contain a b-event, and cannot be mimicked by a correct run, therefore $\mathsf{FAmb}_\infty = \emptyset$. The two infinite correct runs have a^ω as observed sequence, and cannot be mimicked by a faulty run, thus $\mathsf{CAmb}_\infty = \emptyset$. As a consequenc, this pLTS is IA-diagnosable. Consider now the infinite correct run $\rho = q_0 u q_1 a q_1 \ldots$. It has probability $\frac{1}{2}$, and all its finite signalling subruns are ambiguous since their observed sequence is a^n, for some $n \in \mathbb{N}$. Thus for all $n \geq 1$, $\mathbb{P}(\mathsf{CAmb}_n) \geq \frac{1}{2}$, so that this pLTS is not FA-diagnosable.

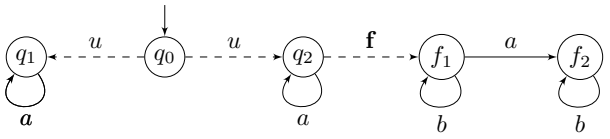

Figure 3 A pLTS that is IA-diagnosable but not FA-diagnosable.

2.4 Characterizations of diagnosability

In this section, we provide characterizations of the different diagnosability notions we introduced. Having algorithmic developments in mind, we focus here on finite-state pLTS. As a consequence, FF-diagnosability and IF-diagnosability coincide, and we stick to IF-diagnosability in the sequel. For each notion of diagnosability, we proceed similarly. First, given a pLTS $\mathcal{A}$ we design a deterministic automaton that accepts some (finite or infinite) observed sequences of $\mathcal{A}$. Then we build the synchronized product of this automaton with $\mathcal{A}$, to obtain another pLTS with the same stochastic behaviour as $\mathcal{A}$ but augmented with additional information about the current run, that will be useful for diagnosability. Finally, we characterize diagnosability by graph properties on the synchronized product.

2.4.1 IF-diagnosability

The first automaton $\mathsf{IF}(\mathcal{A})$, which is a complete automaton, tracks the subset of states reached by a correct signalling run associated with a given observed sequence. The set of states and transitions is inductively defined by:

- $s_0 = \{q_0\}$ is the initial state of $\mathsf{IF}(\mathcal{A})$;
- Given U a state of $\mathsf{IF}(\mathcal{A})$ and any $a \in \Sigma_o$, there is a transition $U \xrightarrow{a} U'$ where:
  $U' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k}$ run of $\mathcal{A}, q_{\alpha_0} \in U, \forall i < k\ a_i \in \Sigma_u \setminus \{\mathbf{f}\}, a_k = a, q_{\alpha_k} = q\}$.

Figure 4 illustrates this construction on the pLTS of Figure 3. As long as b is not observed the current signalling run may be correct leading to either q_1 or q_2, and once b happens, the current signalling run is surely faulty.

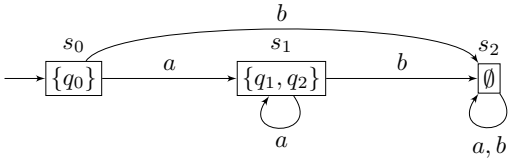

Figure 4 The IF-automaton of pLTS of Figure 3.

We now define the pLTS $\mathcal{A}_{\mathsf{IF}} = \mathcal{A} \times \mathsf{IF}(\mathcal{A})$ as the product of $\mathcal{A}$ and $A(\mathcal{A})$ synchronized over observed events. Since $\mathsf{IF}(\mathcal{A})$ is deterministic and complete, $\mathcal{A}_{\mathsf{IF}}$ is still a pLTS, with same stochastic behaviour as $\mathcal{A}$. In addition, the U-component of a state (q, U) of $\mathcal{A}_{\mathsf{IF}}$ stores the relevant information w.r.t A-diagnosability of the observed sequence so far. Figure 5 illustrates this construction on the pLTS of Figure 3. Observe that this pLTS has two bottom strongly connected components (BSCC): the absorbing states (q_1, s_1) and (f_2, s_2).

In finite DTMC every run almost surely ends up in a BSCC, and A-diagnosability is concerned with (faulty) ambiguous infinite runs, so unsurprisingly, our characterization of diagnosability is based on the BSCC of $\mathcal{A}_{\mathsf{IF}}$.

▶ **Proposition 8.** *Let $\mathcal{A}$ be a finite pLTS. Then $\mathcal{A}$ is IF-diagnosable if and only if $\mathcal{A}_{\mathsf{IF}}$ has no BSCC containing a state (q, U) with $q \in Q_f$ and $U \neq \emptyset$.*

Proof. Suppose first that there exists a reachable BSCC C of $\mathcal{A}_{\mathsf{IF}}$ and a state $s = (q, U)$ in C such that $q \in Q_f$ and $U \neq \emptyset$. Let ρ be a signalling run leading from the initial state s_0 of $\mathcal{A}_{\mathsf{IF}}$ to s. Now, for every state $s' = (q', U') \in C$, necessarily $q' \in Q_f$ and $U' \neq \emptyset$, because C is strongly connected. So for every signalling run ρ' that extends ρ, writing $s' = (q', U')$ for

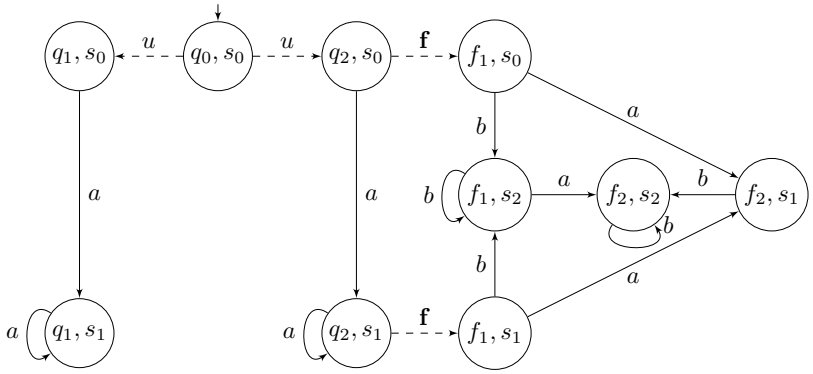

Figure 5 The synchronized product of pLTS of Figure 3 and its IF-automaton.

the state ρ' leads to, there exists a correct signalling run ρ'' such that $\mathcal{P}(\rho'') = \mathcal{P}(\rho')$ and $q_0 \xrightarrow{\rho''} q''$ with $q'' \in U'$. As a consequence the observed sequence $\mathcal{P}(\rho'')$ is ambiguous, and for every $n \geq |\rho|_o$, $\mathbb{P}(\mathsf{FAmb}_n) \geq \mathbb{P}(\rho)$, so that $\mathcal{A}$ is not IF-diagnosable.

Suppose now that for every state $s = (q, U)$ of a BSCC C, either $q \in Q_c$, or $U = \emptyset$. This property is in fact uniform by BSCC: for every BSCC C, either for every state $(q, U) \in C$, $q \in Q_c$, or, for every state $(q, U) \in C$, $U = \emptyset$. This is a straightforward consequence of C being strongly connected. Moreover, $q \in Q_c$ and $U = \emptyset$ cannot hold at the same time, since there is at least one state q' reached by an observed event corresponding to some state (q', U') in $\mathcal{A}_{\mathsf{IF}}$ with $q' \in U'$ implying that $U \neq \emptyset$. Therefore in $\mathcal{A}_{\mathsf{IF}}$ the BSCC are partitioned in non-faulty ones, in case all q- components of states in C are non-faulty, and faulty ones, in case all U- components of states in C are empty ensuring non ambiguity of faulty runs ending in a BSCC. Thus an infinite faulty ambiguous run must only visit transient states. Since almost surely runs leave the transient states in a finite DTMC, this implies that $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$. ◀

Based on the previous characterization, we establish that our definition of IF-diagnosability is equivalent to the original so-called A-diagnosability presented in [13] for finite pLTS.

▶ **Theorem 9.** *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is IF-diagnosable if and only if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$, for every faulty signalling run ρ and every $n \geq N$*

$$\mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) < \varepsilon \mathbb{P}(\rho) \tag{1}$$

Proof. Assume first that $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$, for every faulty signalling run ρ and every $n \geq N$, $\mathbb{P}(\{\rho' \mid \rho' \in \mathsf{FAmb}_{n+|\rho_o|}, \rho \preceq \rho'\}) < \varepsilon \mathbb{P}(\rho)$. Let us fix $\varepsilon > 0$, and $N \in \mathbb{N}$ corresponding to ε in the hypothesis. Using both statements of Lemma 5, there exists n_0 such that for all $n \geq n_0$

$$\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_\infty) < \varepsilon$$

where Δ stands for the symmetric difference: for sets A and B, $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

So for $n, n' \geq n_0$ one has $\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_{n'}) < 2\varepsilon$. Now, for every $n \geq N$

$$\begin{aligned}\mathbb{P}(\mathsf{FAmb}_{n+n_0} \cap \mathsf{FAmb}_{n_0}) &= \mathbb{P}(\bigcup_{\rho \in \mathsf{FAmb}_{n_0}} \{\rho' \in \mathsf{FAmb}_{n+n_0} \mid \rho \preceq \rho'\}) \\ &= \sum_{\rho \in \mathsf{FAmb}_{n_0}} \mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+n_0} \mid \rho \preceq \rho'\}) \\ &< \varepsilon \sum_{\rho \in \mathsf{FAmb}_{n_0}} \mathbb{P}(\rho) \\ &= \varepsilon \mathbb{P}(\mathsf{FAmb}_{n_0}) \leq \varepsilon \ .\end{aligned}$$

Thus $\mathbb{P}(\mathsf{FAmb}_n) < 3\varepsilon$ for all $n \geq N + n_0$. This proves that $\mathcal{A}$ is IF-diagnosable.

Assume now that $\mathcal{A}$ is IF-diagnosable, and let $\varepsilon > 0$. Select N such that for all (q, U) with $q \in Q_f$, the probability that a signalling run of observable length greater or equal than N, starting from (q, U) stays in a transient state is less than ε.
Consider any faulty signalling run ρ of $\mathcal{A}$. In $\mathcal{A}_{\mathsf{IF}}$, this run reaches some state (q, U) with $q \in Q_f$. Since $\mathcal{A}$ is IF-diagnosable and using Proposition 8, one gets for every $n \geq N$

$$\begin{aligned}&\mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) \\ &\qquad \leq \mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge s_0 \stackrel{\rho'}{\Rightarrow} s \text{ with } s \text{ transient state in } \mathcal{A}_{\mathsf{IF}}\}) < \varepsilon \mathbb{P}(\rho) \ .\end{aligned}$$

◀

Note that in the latter theorem, finitess of pLTS is crucial for the proof of necessity. Consider the pLTS of Figure 6. Almost surely a faulty run eventually produces a b-event, that cannot be mimicked by the single correct run. Thus this pLTS is IF-diagnosable. For the faulty signalling run $\rho = q_0 \mathbf{f} f_1 a \dots f_k a f'_k$, for every $n \leq k$, $\mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) = \mathbb{P}(\rho)$. This implies that given any $0 < \varepsilon < 1$, there cannot exist $N \in \mathbb{N}$ such that for every faulty signalling run and every $n \geq N$, Equation (1) holds.

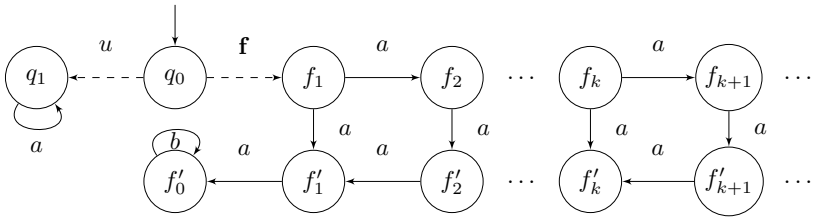

Figure 6 An infinite IF-diagnosable pLTS.

2.4.2 FA-diagnosability

The automaton $\mathsf{FA}(\mathcal{A})$ tracks the subsets of possible states reached by a correct and a faulty signalling run associated with an observed sequence, at the same time. It resembles the on-the-fly determinization of $\mathcal{A}$ viewing unobserved events as silent transitions, yet, in view of the forthcoming characterization, the subsets of correct and faulty states are kept separately. Note also that this construction only considers signalling runs. Formally, the states and transitions of $\mathsf{FA}(\mathcal{A})$ are inductively defined by:

- $s_0 = (\{q_0\}, \emptyset)$ is the initial state of $\mathsf{FA}(\mathcal{A})$;

- Given (U, V) a state of $\mathsf{FA}(\mathcal{A})$ and $a \in \Sigma_o$, there is a transition $(U, V) \xrightarrow{a} (U', V')$ if:
  1. $E = \{\rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \mid \rho \text{ is a run of } \mathcal{A}, q_{\alpha_0} \in U \cup V, \forall i < k\ a_i \in \Sigma_u, a_k = a\} \neq \emptyset$
  2. $U' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ correct run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}$,
  3. $V' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ run of } E, q_{\alpha_0} \in V, q_{\alpha_k} = q\} \cup$
     $\{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ faulty run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}$.

Figure 7 illustrates this construction on the pLTS of Figure 3. As could be expected, the FA-automaton is a refinement of the IF-automaton: the U-component of a state in $\mathsf{FA}(\mathcal{A})$ corresponds to a state in $\mathsf{IF}(\mathcal{A})$. For instance, state s_2 of Figure 4 is split into s_2, s_3 and s_4 .

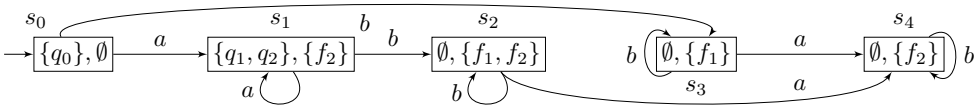

Figure 7 The FA-automaton of pLTS of Figure 3.

We now define the pLTS $\mathcal{A}_{\mathsf{FA}} = \mathcal{A} \times \mathsf{FA}(\mathcal{A})$ as the product of $\mathcal{A}$ and $\mathsf{FA}(\mathcal{A})$ synchronized over observed events. $\mathcal{A}_{\mathsf{FA}}$ is still a pLTS with same stochastic behaviour as $\mathcal{A}$ augmented with the relevant information of the observed sequence w.r.t FA-diagnosability. Figure 8 shows this synchronized product for the pLTS of Figure 3.

As before our characterization of FA-diagnosability is based on the BSCC of $\mathcal{A}_{\mathsf{FA}}$.

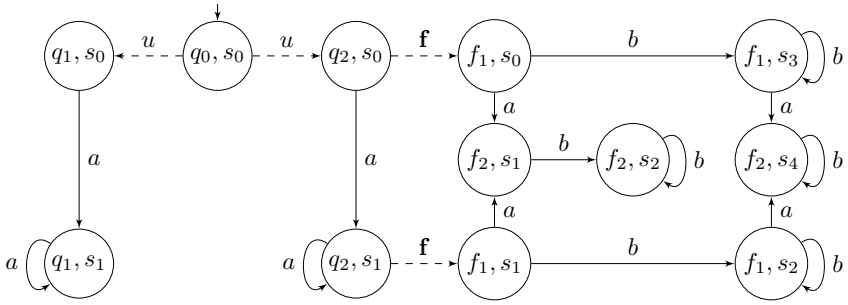

Figure 8 The synchronized product of pLTS of Figure 3 and its FA-automaton.

▶ **Proposition 10.** *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is* FA*-diagnosable if and only if $\mathcal{A}_{\mathsf{FA}}$ has no BSCC that:*

- *either contains a state (q, U, V) with $q \in Q_f$ and $U \neq \emptyset$;*
- *or contains a state (q, U, V) with $q \in Q_c$ and $V \neq \emptyset$.*

Note that the characterization of FA-diagnosability is symmetric for correct states and V set (resp. faulty states and U set). This reflects that the definition of FA-diagnosability itself is symmetric.

Proof. Suppose first that there exists a reachable BSCC C of $\mathcal{A}_{\mathsf{FA}}$ and a state $s = (q, U, V)$ in C such that $q \in Q_f$ and $U \neq \emptyset$. Let ρ be a signalling run leading from the initial state s_0 of $\mathcal{A}_{\mathsf{FA}}$ to s. Now, for every state $s' = (q', U', V') \in C$, necessarily $q' \in Q_f$ and $U' \neq \emptyset$, because C is strongly connected. So for every signalling run ρ' that extends ρ, writing $s' = (q', U', V')$

for the state ρ' leads to, there exists a correct signalling run ρ'' such that $\mathcal{P}(\rho'') = \mathcal{P}(\rho')$ and $q_0 \xrightarrow{\rho''} q''$ with $q'' \in U'$. As a consequence the observed sequence $\mathcal{P}(\rho'')$ is ambiguous, and for every $n \geq |\rho|_o$, $\mathbb{P}(\mathsf{FAmb}_n) \geq \mathbb{P}(\rho)$, so that $\mathcal{A}$ is not FA-diagnosable.
Suppose now that there exists a reachable BSCC C of $\mathcal{A}_{\mathsf{FA}}$ and a state $s = (q, U, V)$ in C such that $q \in Q_c$ and $V \neq \emptyset$. Since the pair (U, V) is unchanged by unobservable transitions, w.l.o.g we assume that s is the successor of some state of C by an observable event and we denote C' the set of such states.
Observe that a signalling run that reaches s is ambiguous. Denote $\pi_i(s')$ the probability that a random path visits a state s' at instant i. In a finite DTMC, for every state s' of a BSCC the Cesaro-limit $\pi_\infty(s') = \lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \pi_i(s')$ exists and is greater than 0. For $s' \in C'$ denote by $p_{s',s}$ the probability of an observable transition from s' to s. Then $0 < \sum_{s' \in C'} \pi_\infty(s') p_{s',s} \leq \liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \alpha_i(s)$ where $\alpha_i(s)$ is the probability that a random path at time i is a signalling run visiting s. From time 0 to time n, a run can be a signalling run at most $n+1$ times. Thus:

$$\frac{1}{n+1} \sum_{i=0}^{n} \alpha_i(s) \leq \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{P}(\mathsf{CAmb}_i)$$

which implies that

$$0 < \liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{P}(\mathsf{CAmb}_i) \leq \limsup_{n\to\infty} \mathbb{P}(\mathsf{CAmb}_n)$$

This shows that $\mathcal{A}$ is not FA-diagnosable.
The proof of proposition 8 has established that a signalling run reaching a BSCC C where for every state $s = (q, U, V)$ q is faulty and $U = \emptyset$ is surely faulty. Similarly a signalling run that reaches a BSCC where for every state $s = (q, U, V)$ q is correct and $V = \emptyset$ is surely correct. Thus an signalling run must only visit transient states. Since a run almost surely leaves the transient states in a finite DTMC, this implies that:

$$\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) + \mathbb{P}(\mathsf{CAmb}_n) = 0 \ .$$

◀

2.4.3 IA-diagnosability

The IA-automaton is the deterministic Büchi automaton introduced in [6]. It refines the FA-automaton by splitting the set V into two disjoints subsets V and W of possible faulty states. The decomposition between V and W reflects the fact that the IA-automaton tries to resolve the ambiguity between U and W (when both are non empty), while V corresponds to a waiting room of states reached by faulty runs that will be examined when the current ambiguity is resolved. Formally, states and transitions of $\mathsf{IA}(\mathcal{A})$ is defined inductively by:

- $s_0 = (\{q_0\}, \emptyset, \emptyset)$ is the initial state of $\mathsf{IA}(\mathcal{A})$;
- Given (U, V, W) a state of $\mathsf{IA}(\mathcal{A})$ and $a \in \Sigma_o$, there is a transition $(U, V, W) \xrightarrow{a} (U', V', W')$ if:
  1. $E = \{\rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \mid \rho \text{ is a run of } \mathcal{A}, q_{\alpha_0} \in U \cup V \cup W, \forall i < k \ a_i \in \Sigma_u, a_k = a\} \neq \emptyset$
  2. $U' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ correct run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}$,
  3. If $W = \emptyset$ then $V' = \emptyset$ and
     $W' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ run of } E, q_{\alpha_0} \in V, q_{\alpha_k} = q\} \cup$
     $\{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ faulty run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}$,

4. If $W \neq \emptyset$ then $W' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ run of } E, q_{\alpha_0} \in W, q_{\alpha_k} = q\}$ and $V' = \Big(\{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ run of } E, q_{\alpha_0} \in V, q_{\alpha_k} = q\} \cup \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ faulty run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}\Big) \setminus W'$.

The set F of accepting states consists of all triples (U, V, W) with $U = \emptyset$ or $W = \emptyset$. When $U = \emptyset$, the current signalling run is surely faulty as for the IF-automaton and the FA-automaton. When $W = \emptyset$ the current signalling run may be ambiguous (if $V \neq \emptyset$) but the "oldest" possible faulty runs have been discarded. Hence, any infinite observed sequence of $\mathcal{A}$ passing infinitely often through F is not ambiguous (ambiguities are resolved one after another).

Figure 9 shows the IA-automaton of the pLTS depicted in Figure 3. Observe that the state s_1 of its FA-automaton (see Figure 7) has been split into two states s_1 and s_1' with $s_1' \in F$. Thus a^ω which is indeed unambiguous is accepted by the IA-automaton.

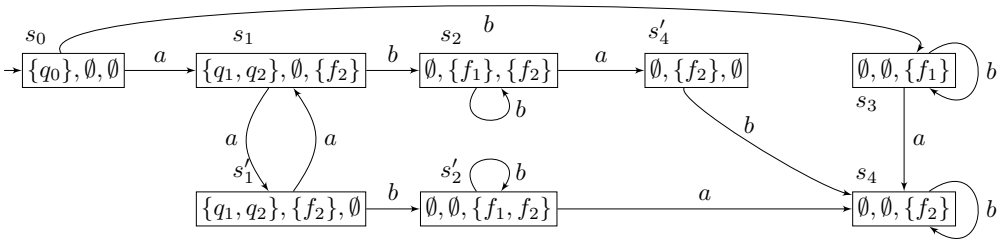

Figure 9 The IA-automaton of pLTS of Figure 3.

The next proposition recalls the property of this automaton.

▶ **Proposition 11** ([6]). *Let $\mathcal{A}$ be a finite pLTS. Then the deterministic Büchi automaton* $\mathsf{IA}(\mathcal{A})$ *accepts the infinite unambiguous observed sequences of $\mathcal{A}$.*

$\mathsf{IA}(\mathcal{A})$ is thus sufficient to characterize IA-diagnosability for $\mathcal{A}$.

As before, to come up with a characterization, one builds $\mathcal{A}_{\mathsf{IA}} = \mathcal{A} \times \mathsf{IA}(\mathcal{A})$, the product of $\mathcal{A}$ and $\mathsf{IA}(\mathcal{A})$ synchronized over observed events. Figure 10 shows the synchronized product corresponding to the pLTS depicted in Figure 3. Among the BSCC, all the faulty ones (*i.e.* the ones reached after a faulty event) have $U = \emptyset$, while $\{(q_1, s_1), (q_1, s_1')\}$, the single one that is reached by a correct run, has a state (q_1, s_1') with $W = \emptyset$.

▶ **Proposition 12.** *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is* IA*-diagnosable if and only if $\mathcal{A}_{\mathsf{IA}}$ has no BSCC such that:*

- *either, all its states (q, U, V, W) fulfill $q \in Q_f$ and $U \neq \emptyset$;*
- *or all its states (q, U, V, W) fulfill $q \in Q_c$ and $W \neq \emptyset$.*

Proof. Assume first that $\mathcal{A}_{\mathsf{IA}}$ has a BSCC with (at least) some state (q, U, V, W) with $q \in Q_f$ and $U \neq \emptyset$. Using Proposition 8, $\mathcal{A}$ is not IF-diagnosable and thus not IA-diagnosable either. If now some BSCC C of $\mathcal{A}_{\mathsf{IA}}$ has all its states (q, U, V, W) with $q \in Q_c$ and $W \neq \emptyset$. In particular none of these states are accepting for the deterministic Büchi automaton $\mathsf{IA}(\mathcal{A})$. Let ρ be a finite signalling run that hits C. By Proposition 11, any infinite run ρ' that extends ρ is ambiguous. From $q \in Q_c$ we deduce that $\mathbb{P}(\mathsf{CAmb}_\infty) \geq \mathbb{P}(\rho) > 0$. Therefore $\mathcal{A}$ is not IA-diagnosable.

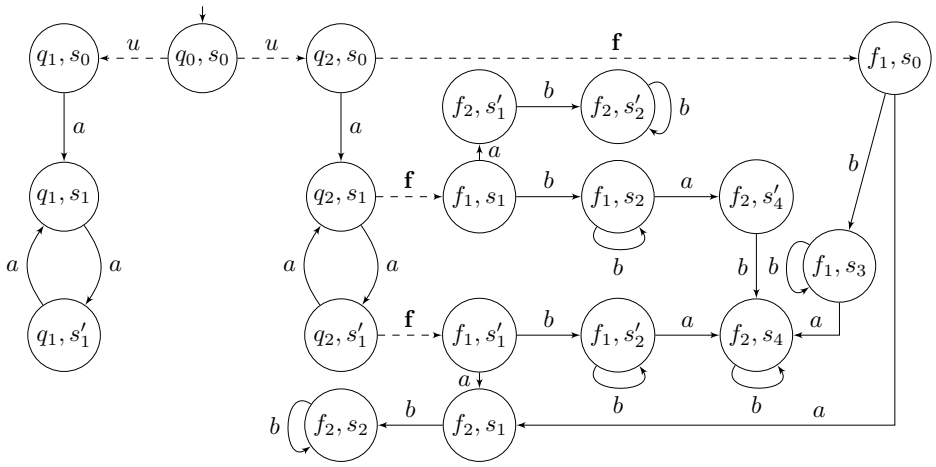

Figure 10 The synchronized product of pLTS of Figure 3 and its IA-automaton.

Assume now $\mathcal{A}_{\mathsf{IA}}$ has no BSCC such that either, all its states (q, U, V, W) fulfill $q \in Q_f$ and $U \neq \emptyset$, or all its states (q, U, V, W) fulfill $q \in Q_c$ and $W \neq \emptyset$. First observe that in case some BSCC of $\mathcal{A}_{\mathsf{IA}}$ contains some state (q, U, V, W) with $q \in Q_f$ and $u \neq \emptyset$, then all its states satisfy the same constraints. Moreover, if some state (q, U, V, W) of a BSCC has $q \in Q_c$, then all states of this BSCC have their q-component in Q_c. Therefore, the condition can be reformulated as follows: all BSCC C of $\mathcal{A}_{\mathsf{IA}}$ satisfy:

- either all states (q, U, V, W) of C fulfill $q \in Q_f$ and $U = \emptyset$;
- or all states (q, U, V, W) of C fulfill $q \in Q_c$ and some state (q, U, V, W) of C fulfills $W = \emptyset$.

Whatever the case, all contain (at least) an accepting state for the Büchi condition of $\mathsf{IA}(\mathcal{A})$. Since every run almost surely ends up in a BSCC and visits each of its states infinitely often, using Proposition 11, almost all runs of $\mathcal{A}_{\mathsf{IA}}$ are unambiguous. This proves that $\mathcal{A}$ is IA-diagnosable. ◀

3 Complexity of diagnosability

In this section we prove the diagnosability problem to be PSPACE-complete, for all variants that we introduced.

3.1 Complexity of IF-diagnosability

IF-diagnosability coincides with A-diagnosability, introduced in [13] and further studied in [3] where a PTIME decision procedure is provided.

▶ **Fact 13.** *The decision procedure of [3] for A-diagnosability is erroneous.*

In Appendix A, we describe their algorithm, give an example of pLTS on which it is not sound, and explain where the error lies in their correctness proof.

In order to establish a lower bound for the complexity of IF-diagnosability, we introduce a variant of language universality. A language $\mathcal{L}$ over an alphabet Σ is said *eventually universal* if there exists a word $v \in \Sigma^*$ such that $v^{-1}\mathcal{L} = \Sigma^*$. Recently, several variants of

the universality problem were shown to be PSPACE-complete [10] but, to the best of our knowledge, eventual universality has not been considered.

Because of our diagnosis framework, we focus on live non deterministic finite automata (NFA). Similarly to pLTS, an NFA is *live* if from every state there is at least one outgoing transition. The language of an NFA $\mathcal{A}$, denoted $\mathcal{L}(\mathcal{A})$, is defined as the set of finite words that are accepted by $\mathcal{A}$.

We reduce the universality problem for NFA, which is known to be PSPACE-complete [9] to the eventual universality problem.

▶ **Proposition 14.** *Let $\mathcal{A}$ be a live NFA where all states are terminal. Then deciding whether $\mathcal{L}(\mathcal{A})$ is eventually universal is* PSPACE*-hard.*

Proof. Let $\mathcal{A} = (Q, \Sigma, T, q_0, F)$ be an NFA. From $\mathcal{A}$ we define the NFA $\mathcal{A}' = (Q', \Sigma', T', q_0, Q')$ where $\Sigma' = \Sigma \uplus \{\sharp\}$, $Q' = Q \uplus \{s\}$, and

$$T' = T \cup \{(q, \sharp, q_0) \mid q \in F\} \cup \{(s, a, s) \mid a \in \Sigma\} \cup \{(q, a, s) \mid a \in \Sigma,\ q \not\to_{\mathcal{A}}\} \ .$$

Assume first that $\mathcal{L}(\mathcal{A}) = \Sigma^*$. Any word w over the alphabet Σ' can be decomposed into $w = w_0\sharp w_1\sharp \dots \sharp w_n$ with $w_i \in \Sigma^*$. For each factor w_i, since $\mathcal{A}$ is universal, there exists a run ρ_i on w ending in some terminal state q_i in $\mathcal{A}$. Therefore w is accepted in $\mathcal{A}'$ by the run $\rho_0\sharp\rho_1\sharp \dots \sharp\rho_n$. Hence $\mathcal{A}'$ is universal, and thus eventually universal: $\varepsilon^{-1}\mathcal{L}(\mathcal{A}') = \Sigma'^*$.
Assume that $\mathcal{A}'$ is eventually universal and let $v \in \Sigma'^*$ be such that $v^{-1}\mathcal{L}(\mathcal{A}') = \Sigma'^*$. Given $w \in \Sigma^*$, we consider the word $w' = v\sharp w\sharp$. Since $\mathcal{A}'$ is eventually universal with witness v, $w' \in \mathcal{L}(\mathcal{A}')$ and an accepting run can be decomposed as: $\rho\sharp\rho'\sharp q_0$ where run ρ' which corresponds to word w has q_0 as initial state, ends in a final state of $\mathcal{A}$, and only uses transitions of $\mathcal{A}$. So ρ' is a run of $\mathcal{A}$ that accepts w. Therefore $w \in \mathcal{L}(\mathcal{A})$, and $\mathcal{A}$ is universal. ◀

▶ **Proposition 15.** *The* IF*-diagnosability problem is* PSPACE*-hard.*

Proof. The proof is by reduction from the eventual universality problem, and relies on the characterization of Proposition 8. Let $\mathcal{A}$ be a live NFA over Σ, in which all states are final. From $\mathcal{A}$, one builds a live pLTS $\mathcal{A}'$ as depicted in Figure 11. In $\mathcal{A}'$, the set of events is

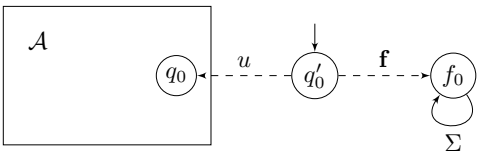

Figure 11 A reduction for PSPACE-hardness of IF-diagnosability.

$\Sigma \uplus \{u, \mathbf{f}\}$, and u and $\mathbf{f}$ are the unobservable events. Under this construction, we will show that $\mathcal{A}$ is eventually universal if and only if $\mathcal{A}'$ is not IF-diagnosable.
Let us have a closer look to the product pLTS $\mathcal{A}'_{\mathsf{IF}}$. For $w \in \Sigma^+$, if $\mathbf{f}w$ leads to some state (f_0, U) in $\mathcal{A}'_{\mathsf{IF}}$, by contruction of $\mathcal{A}'$, U corresponds to the subset of states reachable in $\mathcal{A}$ after reading w. Otherwise stated, either $w \in \mathcal{L}(\mathcal{A})$ and the non empty set U consists of the on-the-fly determinization of $\mathcal{A}$, or $w \notin \mathcal{L}(\mathcal{A})$ and $U = \emptyset$.
Assume first that $\mathcal{A}'$ is not IF-diagnosable. By Proposition 8, $\mathcal{A}'_{\mathsf{IF}}$ contains a reachable BSCC $\mathcal{C}$ with some state $s = (f_0, U) \in \mathcal{C}$ such that $U \neq \emptyset$. In fact, since $\mathcal{C}$ is a BSCC, and because f_0 is a sink state in $\mathcal{A}'$ where all events are enabled, for every state s' of $\mathcal{C}$ there exists $U' \neq \emptyset$

such that $s' = (f_0, U')$. As a consequence, for any $v \in \Sigma^*$ such that $\mathbf{f}v$ leads from the initial state to $s \in \mathcal{C}$ in $\mathcal{A}'_A$, we must have $v^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$. Therefore, $\mathcal{A}$ is eventually universal.
Conversely assume that there exists a word $v \in \Sigma^*$ such that $v^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$. Of course, any word extending v is also a witness that $\mathcal{A}$ is eventually universal. Let $v' \in \Sigma^*$ be such that, in $\mathcal{A}'_A$, the run reading $\mathbf{f}vv'$ ends in a BSCC $\mathcal{C}$. Since $(vv')^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$, all states of $\mathcal{C}$ are of the form (f_0, U) with $U \neq \emptyset$. Therefore, by Proposition 8, $\mathcal{A}'$ is not IF-diagnosable. ◀

▶ **Proposition 16.** *The* IF*-diagnosability problem is in* PSPACE*.*

Proof. We use the characterization of IF-diagnosability given in Proposition 8 without building explicitely the product pLTS $\mathcal{A}_{\mathsf{IF}}$. Given two states s, s' of $\mathcal{A}_{\mathsf{IF}}$, we can in polynomial space check whether one is reachable from the other. Using this procedure, we can check whether a state s is not in a BSCC by guessing another state s' such that s' is reachable from s but s is not reachable from s' (here we use Savitch theorem).
Thus the procedure which decides whether $\mathcal{A}$ is not IF-diagnosable consists in guessing a state $s = (q, U)$ with $q \in Q_f$ and $U \neq \emptyset$ and checking whether s is in a BSCC (here again, we use Savitch theorem). ◀

Propositions 15 and 16 determine the precise complexity of IF-diagnosability, as summarized below:

▶ **Theorem 17.** *The* IF*-diagnosability problem is* PSPACE*-complete.*

3.2 Complexity of FA-diagnosability

▶ **Proposition 18.** *The* FA*-diagnosability problem is* PSPACE*-hard.*

Proof. The proof is again by reduction from the eventual universality problem. Let $\mathcal{A}$ be an NFA and $\mathcal{A}'$ the pLTS built from $\mathcal{A}$ as depicted in Figure 12. For any $n \in \mathbb{N}$, the probability

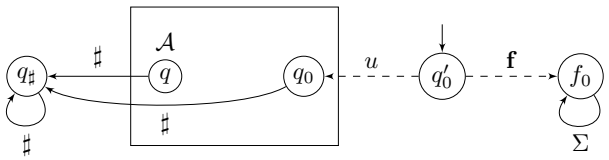

Figure 12 A reduction for PSPACE-hardness of FA- and IA-diagnosability.

of ambiguity can be decomposed as:

$$\limsup_{n\to\infty} \mathbb{P}_{\mathcal{A}'}(\mathsf{CAmb}_n \cup \mathsf{FAmb}_n) =$$
$$\limsup_{n\to\infty} \mathbb{P}_{\mathcal{A}'}(\{q'_0 u q_0 \rho \mid q'_0 u q_0 \rho \in \mathsf{CAmb}_n\}) + \mathbb{P}_{\mathcal{A}'}(\{q'_0 f f_0 \rho \mid q'_0 f f_0 \rho \in \mathsf{FAmb}_n\})$$

Observe that an infinite run starting by $q'_0 u q_0$ almost surely contains a $\sharp$ event. On the other hand, an ambiguous signalling run $\rho \in \mathsf{SR}_n$ starting by $q'_0 u q_0$ must not contain a $\sharp$ event. Thus the probability that such runs are ambiguous is decreasing and tends to 0, when n goes to infinity. We deduce

$$\limsup_{n\to\infty} \mathbb{P}_{\mathcal{A}'}(\mathsf{CAmb}_n \cup \mathsf{FAmb}_n) = \limsup_{n\to\infty} \mathbb{P}_{\mathcal{A}'}(\mathsf{FAmb}_n)$$

and therefore, $\mathcal{A}'$ is FA-diagnosable if and only if it is FF- (or IF-)diagnosable.

Since the faulty runs never contain a $\sharp$, only runs that do not contain $\sharp$ are relevant for ambiguity of faulty runs. We are thus is the same position as in the proof of Proposition 15, and $\mathcal{A}'$ is not IF-diagnosable if and only if $\mathcal{A}$ is eventually universal. ◀

▶ **Proposition 19.** *The* FA*-diagnosability problem is in* PSPACE*.*

Proof. We use the characterization of FA-diagnosability given in Proposition 10 without explicitely building the product pLTS $\mathcal{A}_{\mathsf{FA}}$. Similarly to the proof of Proposition 16, we heavily use Savitch theorem here. First given a state (q, U, V) of $\mathcal{A}_{\mathsf{FA}}$ we can check in polynomial space whether it belongs to a BSCC (as in the proof of Proposition 16). We can also check in polynomial space whether it can be reached from some state (q', U', V') with $U' = \emptyset$ or $V' = \emptyset$ by guessing such a state. Combining the two, this provides a polynomial space algorithm to check whether (q, U, V) belongs to a BSCC in which no state (q', U', V') fulfills $U' \neq \emptyset$ and $V' \neq \emptyset$.
Thus the procedure that decides whether $\mathcal{A}$ is not FA-diagnosable consists in guessing a state $s = (q, U, V)$, checking that it is reachable from s_0 and belongs to a BSCC where all states (q', U', V') of the BSCC fulfill $U' \neq \emptyset$ and $V' \neq \emptyset$. ◀

Propositions 18 and 19 determine the precise complexity of FA-diagnosability, as summarized below:

▶ **Theorem 20.** *The* FA*-diagnosability problem is* PSPACE*-complete.*

3.3 Complexity of IA-diagnosability

▶ **Proposition 21.** *The* IA*-diagnosability problem is* PSPACE*-hard.*

Proof. The proof is again by reduction from the eventual universality problem. Let $\mathcal{A}$ be an NFA and consider again the pLTS $\mathcal{A}'$ depicted in Figure 12. The probability of ambiguity can be decomposed as:

$$\mathbb{P}_{\mathcal{A}'}(\mathsf{CAmb}_\infty \cup \mathsf{FAmb}_\infty) = \\ \mathbb{P}_{\mathcal{A}'}(\{q'_0 u q_0 \cdot \rho \mid q'_0 u q_0 \cdot \rho \in \mathsf{CAmb}_\infty\}) + \mathbb{P}_{\mathcal{A}'}(\{q'_0 f f_0 \cdot \rho \mid q'_0 f f_0 \cdot \rho \in \mathsf{FAmb}_\infty\}) \ .$$

Observe that an infinite run starting with $q'_0 u q_0$ almost surely contains a $\sharp$ event. Thus the probability that such runs are ambiguous is null and we obtain:

$$\mathbb{P}_{\mathcal{A}'}(\mathsf{CAmb}_\infty \cup \mathsf{FAmb}_\infty) = \mathbb{P}_{\mathcal{A}'}(\mathsf{FAmb}_\infty)$$

and therefore, $\mathcal{A}'$ is IA-diagnosable if and only if it is IF-diagnosable. Now, we exploit the end of the proof of Proposition 18 to obtain that $\mathcal{A}'$ is not IF-diagnosable if and only if $\mathcal{A}$ is eventually universal. ◀

▶ **Proposition 22.** *The* IA*-diagnosability problem is in* PSPACE*.*

Proof. We use the characterization of IA-diagnosability given in Proposition 12 without building explicitely $\mathcal{A}_{\mathsf{IA}}$. First, given a state (q, U, V, W) of $\mathcal{A}_{\mathsf{IA}}$, we can check in polynomial space that it belongs to a BSCC (as in the proof of Proposition 16). We can also check in polynomial space whether it is coreachable from a state (q', U', V', W') that fulfills $U' = \emptyset$ or $W' = \emptyset$ by guessing such a state (we use Savitch theorem here). Combining the two procedures, we can check in polynomial space whether (q, U, V, W) belongs to a BSCC where all states (q', U', V', W') of the BSCC fulfill $U' \neq \emptyset$ and $W' \neq \emptyset$.

Thus the procedure that decides whether $\mathcal{A}$ is not IA-diagnosable consists in guessing a state $s = (q, U, V, W)$, checking that it is reachable from s_0 and belongs to a BSCC where all states (q', U', V', W') of the BSCC fulfill $U' \neq \emptyset$ and $W' \neq \emptyset$. ◀

Propositions 21 and 22 determine the precise complexity of IA-diagnosability, as summarized below:

▶ **Theorem 23.** *The* IA*-diagnosability problem is* PSPACE*-complete.*

4 Diagnoser construction

In this section we focus on the construction of diagnosers. A diagnoser is a function $D : \Sigma_o^* \to \{?, \top, \bot\}$ assigning to every finite observation sequence a verdict. Informally when a diagnoser outputs ? it does not provide any information, while $\top$ ensures that a fault is certain and $\bot$ that some information about correctness has been provided. We consider the natural partial order $\prec$ on these values defined by $? \prec \top$ and $? \prec \bot$.

A finite memory diagnoser is given by a tuple $(M, \Sigma, m_0, \mathsf{up}, D_{fm})$ where M is a finite set of memory states, $m_0 \in M$ is the initial memory state, $\mathsf{up} : M \times \Sigma_o \to M$ is a memory update function, and finally $D_{fm} : M \to \{?, \top, \bot\}$ is a diagnoser function. The mapping up is extended into a function $\mathsf{up} : M \times \Sigma_o^* \to M$ defined inductively by $\mathsf{up}(m, \varepsilon) = m$ and $\mathsf{up}(m, wa) = \mathsf{up}(\mathsf{up}(m, w), a)$. A finite memory diagnoser is not a diagnoser as defined above, yet it induces the diagnoser defined by $D(w) = D_{fm}(\mathsf{up}(m_0, w))$.

Diagnosers we define in the sequel will have two important properties: soundness and reactivity. Soundness ensures that the information provided is accurate and reactivity specifies which pieces of information the diagnoser must provide. The precise soundness and reactivity requirements will depend on the considered diagnosability notion. Moreover, we only consider diagnosers that, once they output $\top$, never change their verdict in the future. Note that any sound diagnoser can be turned into one that is sound and satisfies this commitment property.

4.1 IF-diagnoser

We start with IF-diagnoser, that performs diagnosis of IF-diagnosable pLTS. In the sequel we fix $\mathcal{A}$ a finite pLTS.

▶ **Definition 24.** An IF*-diagnoser* for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, ?\}$ such that,
soundness For every $w \in \Sigma_o^*$, if $D(w) = \top$, then w is surely faulty.
reactivity For every finite faulty run ρ, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) = ?\}) = 0$ where for $w \in \Sigma_o^\omega$, $D(w) = \lim_{n\to\infty} D(w_{\leq n})$.
Note that in the above definition the limit is well-defined because we assumed that the diagnoser commits to $\top$.

▶ **Proposition 25.** *A finite pLTS is* IF*-diagnosable if and only if it admits an* IF*-diagnoser.*

Proof. Let $\mathcal{A}$ be a pLTS, and assume there exists an IF-diagnoser D for $\mathcal{A}$.
Let $\varepsilon > 0$. Using Lemma 5, first select n_0 such that for all $n \geq n_0$

$$\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_\infty) < \varepsilon \ ,$$

where Δ stands for the symmetric difference. So for $n \geq n_0$ one has $\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_{n_0}) < 2\varepsilon$.

Since D is sound, for any signalling faulty run $\rho \in \mathsf{SR}_{n_0}$, and for any $n \geq n_0$

$$\{\rho' \in \mathsf{FAmb}_n \mid \rho \preceq \rho'\} \subseteq \{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\} \ .$$

By the reactivity condition, D almost surely detects faults, and because the number of signalling runs of fixed observable length is finite (since $\mathcal{A}$ is convergent by hypothesis), there exists $N \in \mathbb{N}$ such that for every $n \geq N + n_0$ and every signalling faulty run $\rho \in \mathsf{SR}_{n_0}$, $\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) < \varepsilon \cdot \mathbb{P}(\rho)$.
Thus for every $n \geq N + n_0$,

$$\mathbb{P}(\mathsf{FAmb}_n \cap \mathsf{FAmb}_{n_0}) = \mathbb{P}\Big(\biguplus_{\rho \in \mathsf{FAmb}_{n_0}} \{\rho' \mid \rho \in \mathsf{FAmb}_n, \rho \preceq \rho'\} \Big) =$$

$$\sum_{\rho \in \mathsf{FAmb}_{n_0}} \mathbb{P}(\{\rho' \mid \rho \in \mathsf{FAmb}_n, \rho \preceq \rho'\}) < \varepsilon \sum_{\rho \in \mathsf{FAmb}_{n_0}} \mathbb{P}(\rho) = \varepsilon \mathbb{P}(\mathsf{FAmb}_{n_0}) \leq \varepsilon \ .$$

Thus $\mathbb{P}(\mathsf{FAmb}_n) < 3\varepsilon$ for every $n \geq N + n_0$, which proves that $\mathcal{A}$ is IF-diagnosable.

Assume that $\mathcal{A}$ is IF-diagnosable. We define the function $D : \Sigma_o^* \to \{\top, ?\}$ by $D(w) = \top$ if and only if w is a surely faulty observed sequence. Let us check that D is an IF-diagnoser. Since $D(w) = \top$ iff w is a surely faulty sequence, D is sound. Now, let ρ be a faulty run.

$$\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) = \lim_{n \to \infty} \mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) \ .$$

For every $n \in \mathbb{N}$, $\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\} \subseteq \mathsf{FAmb}_{n+|\rho|_o}$ and $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) = 0$.
Therefore $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) = 0$ and D is reactive. ◀

▶ **Proposition 26.** *If $\mathcal{A}$ is an IF-diagnosable pLTS with n correct states, one can build an IF-diagnoser with at most 2^n memory states where $n = |Q_c|$.*

Proof. For an IF-diagnosable pLTS $\mathcal{A}$ with $\mathsf{IF}(\mathcal{A}) = (Q^*, \Sigma_o, T^*, \{q_0\})$ its deterministic and complete IF-automaton, we define the finite memory diagnoser $(M, \Sigma, \mathsf{up}, m_0, D_{fm})$ with $M = Q^*$ and $m_0 = \{q_0\}$, $\mathsf{up}(q, a) = T^*(q, a)$ and $D_{fm}(U) = \top$ iff $U = \emptyset$. Let us show that the induced diagnoser D is indeed an IF-diagnoser, and that it has at most 2^n memory states, where n is the number of correct states of $\mathcal{A}$.

soundness When D outputs the verdict $\top$, $\mathsf{IF}(\mathcal{A})$ is in the state associated with $\emptyset$. Thus the observed sequence is surely faulty.

reactivity If an infinite faulty run ρ is such that $D(\mathcal{P}(\rho)) =?$ then, by construction of $\mathsf{IF}(\mathcal{A})$ and definition of D, for every length $n \in \mathbb{N}$, there exists a finite correct signalling run $\rho_n \in \mathsf{SR}_n$ such that $\mathcal{P}(\rho_n) = \mathcal{P}(\rho_{\downarrow n})$. Using König's lemma, since $\mathcal{A}$ is finitely branching, one can extract an infinite correct run ρ_∞ such that $\mathcal{P}(\rho_\infty) = \mathcal{P}(\rho)$, so that $\rho \in \mathsf{FAmb}_\infty$. Moreover $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$ as $\mathcal{A}$ is IF-diagnosable. Putting everything together, for every faulty run ρ, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) = 0$.

size The memory states are states of $\mathsf{IF}(\mathcal{A})$, which are themselves subsets of correct states of $\mathcal{A}$. Therefore, D uses at most 2^n memory states, with $n = |Q_c|$.

◀

▶ **Proposition 27.** *There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of IF-diagnosable pLTS such that $\mathcal{A}_n$ has $n+1$ correct states and it admits no IF-diagnoser with less than 2^n states.*

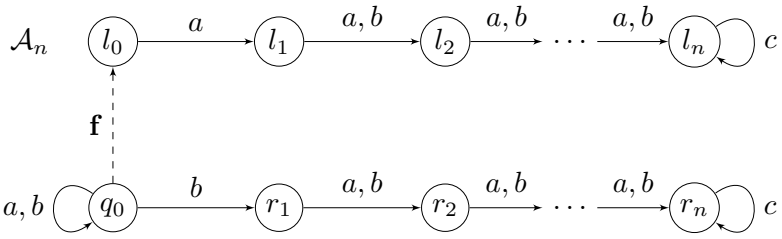

Figure 13 Example of an IF-diagnosable pLTS requiring an IF-diagnoser with exponential size.

Proof. Consider the example of Figure 13 where $\Sigma_o = \{a, b, c\}$ and the initial state is q_0. Consider a finite faulty run including a c event. Its observed sequence belongs to $\mathcal{L} = \{a, b\}^* b \{a, b\}^{n-1} c^+$. Since any finite correct run has an observed sequence belonging to $\mathcal{L}' = \{a, b\}^* \cup \{a, b\}^* a \{a, b\}^{n-1} c^+$ and $\mathcal{L} \cap \mathcal{L}' = \emptyset$, $\mathsf{FAmb}_n \uplus \mathsf{CAmb}_n \subseteq \{\rho \mid \mathcal{P}(\rho) \in \{a, b\}^n\}$. Since $\lim_{n\to\infty} \mathbb{P}(\{\rho \mid \mathcal{P}(\rho) \in \{a, b\}^n\}) = 0$, the pLTS is FA-diagnosable and so IA-diagnosable and IF-diagnosable.
Intuitively, when a c is observed, any IF-diagnoser must have remembered the observable event that happened n steps earlier to know if the run is faulty or not. Thus, it must remember the last n observed events, in case a c event occurs.
More formally, assume there exists a diagnoser $D = (M, \Sigma, m_0, \mathsf{up}, D_{fm})$ with less than 2^n memory states. Then there exist two distinct words $w_1 \in \{a, b\}^n$ and $w_2 \in \{a, b\}^n$ leading to the same memory state: $\mathsf{up}(m_0, w_1) = \mathsf{up}(m_0, w_2)$. The words w_1 and w_2 differ at least from one letter say $w_1[i] = a$ and $w_2[i] = b$. Consider for $k \geq 1$, the signalling correct run $\rho_{1,k}$ corresponding to observed sequence $w_1 a^{i-1} c^k$ whose sequence of visited states is $q_0^i r_1 \dots r_n^{k+1}$ and the signalling faulty run $\rho_{2,k}$ corresponding to observed sequence $w_2 a^{i-1} c^k$ whose sequence of visited states is $q_0^i l_0 l_1 \dots l_n^{k+1}$. They also lead to the same memory state. By soundness, $D(w_1 a^{i-1} c^k) =?$. Thus for all suffix ρ of $\rho_{2,1}$, $D(\rho) =?$ contradicting the reactivity of D. ◀

4.2 FA-diagnoser

FA-diagnosability and IA-diagnosability not only consider the diagnosis of faults but also of correct runs. Different from IF-diagnosers, FA- and IA-diagnosers have three possible verdicts $\top$, related to faulty sequences, $\bot$, linked with correctness, and ? when no information can be derived from the observation.

▶ **Definition 28.** An FA-diagnoser for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, \bot, ?\}$ such that
soundness For every $w \in \Sigma_o^*$
- if $D(w) = \top$, then w is surely faulty;
- if $D(w) = \bot$, then w is surely correct.

reactivity $\mathbb{P}(\{\rho \in \Omega \mid D_{\inf}(\mathcal{P}(\rho)) =?\}) = 0$ where for $w \in \Sigma_o^\omega$, $D_{\inf}(w) = \liminf_{n\to\infty} D(w_{\leq n})$.

▶ **Proposition 29.** *A finite pLTS $\mathcal{A}$ is* FA*-diagnosable if and only if it admits an* FA*-diagnoser.*

Proof. Assume first that there exists an FA-diagnoser D for $\mathcal{A}$. For every $n \in \mathbb{N}$, we define $\mathsf{FD}_n = \{\rho \in \Omega \mid \forall m \geq n, D(\mathcal{P}(\rho_{\downarrow m})) = \top\}$ the set of runs that are diagnosed faulty after n observed events, and symmetrically $\mathsf{CD}_n = \{\rho \in \Omega \mid \forall m \geq n, D(\mathcal{P}(\rho_{\downarrow m})) = \bot\}$ the set of

runs that are persistenly diagnosed correct after n observed events. The sequences $(\mathsf{CD}_n)_{n\in\mathbb{N}}$ and $(\mathsf{FD}_n)_{n\in\mathbb{N}}$ are non-decreasing. As $? \prec \bot$ and $? \prec \top$, for every run $\rho \in \Omega$, $D_{\inf}(\mathcal{P}(\rho)) =?$ is equivalent to $\rho \notin \bigcup_n(\mathsf{FD}_n \cup \mathsf{CD}_n)$. Thus $\bigcup_{n\in\mathbb{N}}(\mathsf{FD}_n \cup \mathsf{CD}_n) = \{\rho \in \Omega \mid D_{\inf}(\mathcal{P}(\rho)) \neq ?\}$. Since D is reactive, $\mathbb{P}(\{\rho \in \Omega \mid D_{\inf}(\mathcal{P}(\rho)) \neq ?\}) = 1$. Moreover, since D is sound, for every $n \in \mathbb{N}$, $\mathsf{FD}_n \subseteq \mathsf{Sf}_n$ and $\mathsf{CD}_n \subseteq \mathsf{C}_n \setminus \mathsf{CAmb}_n$. Thus for every $n \in \mathbb{N}$, $\mathbb{P}(\mathsf{FAmb}_n \cup \mathsf{CAmb}_n) = 1 - \mathbb{P}(\mathsf{Sf}_n \cup \mathsf{C}_n \setminus \mathsf{CAmb}_n) \leq 1 - \mathbb{P}(\mathsf{FD}_n \cup \mathsf{CD}_n)$ and $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \cup \mathsf{CAmb}_n) \leq 1 - \liminf_{n\to\infty} \mathbb{P}(\{\rho \in \mathsf{SR}_n \mid D(\mathcal{P}(\rho)) \neq ?\}) = 0$. This shows that $\mathcal{A}$ is FA-diagnosable.

Assume now that $\mathcal{A}$ is FA-diagnosable. From $\mathsf{FA}(\mathcal{A}) = (Q^*, \Sigma_o, T^*, (\{q_0\}, \emptyset))$ the FA-automaton of $\mathcal{A}$, we define the finite memory diagnoser $D = (M, \Sigma, m_0, \mathsf{up}, D_{fm})$ where $M = Q^*$, $m_0 = (\{q_0\}, \emptyset)$, $\mathsf{up}(m, a) = T^*(m, a)$, $D_{fm}((U, V)) = \top$ iff $U = \emptyset$ and $D_{fm}((U, V)) = \bot$ iff $V = \emptyset$. Let us check that D is an FA-diagnoser, and that its size is at most 2^n if n denotes the number of states of $\mathcal{A}$.

soundness Let $w \in \Sigma_o^*$ be an observation sequence. If (U, V) is the state in $\mathsf{FA}(\mathcal{A})$ reached after reading w, then recall that U (resp. V) is the set of states in $\mathcal{A}$ that can be reached by correct (resp. faulty) signalling runs labeled by w. By construction, if $D(w) = \top$ then w is surely faulty, and if $D(w) = \bot$ then w is surely correct.

reactivity Let ρ be a signalling run such that $D(\mathcal{P}(\rho)) =?$. Due to the characterization of Proposition 10, the SCC of $\mathcal{A}_{\mathsf{FA}}$ that ρ has reached cannot be a BSCC. So given some n, $\mathbb{P}(\{\rho \in \Omega \mid \exists m \geq n \ D(\mathcal{P}(\rho_{\downarrow m}) =?\}) \leq \mathbb{P}(\{\rho \in \Omega \mid \rho_{\downarrow n} \text{ does not reach a BSCC}\})$. Thus $\mathbb{P}(\{\rho \in \Omega \mid D_{\inf}(\mathcal{P}(\rho)) =?\}) = \lim_{n\to\infty} \mathbb{P}(\{\rho \in \Omega \mid \exists m \geq n \ D(\mathcal{P}(\rho_{\downarrow m}) =?\}) \leq \limsup_{n\to\infty} \mathbb{P}(\{\rho \in \Omega \mid \rho_{\downarrow n} \text{ does not reach a BSCC}\}) = 0$.

size D has at most 2^n memory states because every state of $\mathsf{FA}(\mathcal{A})$ consists of a pair (U, V) with $U \subseteq Q_c$ and $V \subseteq Q_f$.

◀

In the latter proof, we explicitely built an FA-diagnoser, yielding an upper bound on the size of FA-diagnosers.

▶ **Proposition 30.** *For every* FA*-diagnosable pLTS $\mathcal{A}$ with n states, one can build an* FA*-diagnoser with at most 2^n memory states.*

As the pLTS of Figure 13 is FA-diagnosable, and since any FA-diagnoser is also an A-diagnoser, using Proposition 27 we obtain the following lower bound for the size of FA-diagnosers.

▶ **Proposition 31.** *There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of* FA*-diagnosable pLTS such that $\mathcal{A}_n$ has $2n + 2$ states and it admits no* FA*-diagnoser with less than 2^n memory states.*

4.3 IA-diagnoser

Last we introduce IA-diagnosers, that mostly differ from FA-diagnosers on the soundness requirement. Intuitively, IA-diagnosers may resolve an ambiguity late, while another one has already been produced, contrary to FA-diagnosers.

▶ **Definition 32.** An IA-diagnoser for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, \bot, ?\}$ such that

soundness For all $w \in \Sigma_o^*$

- if $D(w) = \top$, then w is surely faulty;
- if $D(w) = \bot$, letting $|D(w)|_\bot = |\{0 < n \leq |w| \mid D(w_{\leq n}) = \bot\}|$, then for all signalling run ρ such that $\mathcal{P}(\rho) = w$, $\rho_{\downarrow |D(w)|_\bot}$ is correct.

reactivity $\mathbb{P}(\{\rho \in \Omega \mid D_{\sup}(\mathcal{P}(\rho)) =?\}) = 0$ where for $w \in \Sigma_o^\omega$, $D_{\sup}(w) = \limsup_{n\to\infty} D(w_{\leq n})$.

($D_{\sup}$ is well-defined since once the diagnoser outputs $\top$, it always sticks to this verdict.)

The interpretation of $D(w) = \bot$ is that the diagnoser ensures that any signalling subrun of length $|D(w)|_\bot \leq |w|$ of a signalling run for w is correct. Of course it may deduce this information from the last $|w| - |D(w)|_\bot$ observations. This is illustrated on the example of Figure 14 for which we describe an IA-diagnoser. After observing any sequence $wbaa$, with $w \in \{a, b\}^*$, the diagnoser knows a posteriori that two steps before, that is after the observation of wb, the run was necessarily correct. Indeed, observing the suffix aa is not possible after a fault, yet wba is not surely correct. Let D be defined by: for $w \in \{a, b\}^*(ab \cup aa)$, $D(w) = \bot$, for $w \in \{a, b, c\}^*c$, $D(w) = \top$ and otherwise $D(w) =?$. Then D is an IA-diagnoser.

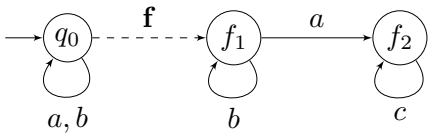

Figure 14 A pLTS which is IA-diagnosable.

▶ **Proposition 33.** *A finite pLTS $\mathcal{A}$ is* IA*-diagnosable if and only if it admits an* IA*-diagnoser.*

Proof. Assume first that there exists an IA-diagnoser D for $\mathcal{A}$, and let ρ be an infinite run. By reactivity, almost surely $D_{\sup}(\mathcal{P}(\rho)) \in \{\top, \bot\}$. If $D_{\sup}(\mathcal{P}(\rho)) = \top$ then there exists some n such that $D(\mathcal{P}(\rho_{\downarrow n})) = \top$. By soundness, $\rho_{\downarrow n}$ is surely faulty and thus ρ is surely faulty. If $D_{\sup}(\mathcal{P}(\rho)) = \bot$, we claim that ρ is surely correct. Observe that the diagnoser infinitely often ouputs $\bot$, so by soundness, for all n, $\mathcal{P}(\rho_{\downarrow n})$ is surely correct and thus in particular $\rho_{\downarrow n}$ is correct. Assume there exists an infinite faulty run ρ' with $\mathcal{P}(\rho') = \mathcal{P}(\rho)$. There exists a n such that for all $m \geq n$, $\rho'_{\downarrow n}$ is faulty. Thus by soundness there can be no more n $\bot$ verdicts for $\mathcal{P}(\rho)$ contradicting the fact that $D_{\sup}(\mathcal{P}(\rho)) = \bot$. Thus with probability 1, an infinite run is unambiguous.

Assume now that $\mathcal{A}$ is IA-diagnosable, and denote $\mathsf{IA}(\mathcal{A})$ its IA-automaton. For any word $w \in \Sigma_o^*$, we denote by (U_w, V_w, W_w) the state in $\mathsf{IA}(\mathcal{A})$ reached after reading w. For every finite signalling run ρ of $\mathcal{A}$, we denote by $(U_\rho, V_\rho, W_\rho) = (U_{\mathcal{P}(\rho)}, V_{\mathcal{P}(\rho)}, W_{\mathcal{P}(\rho)})$. The function D is then defined as follows: $D(w) = \top$ iff $U_w = \emptyset$, $D(w) = \bot$ iff $W_w = \emptyset$ and $U_w \neq \emptyset$, and in all other cases $D(w) =?$. Let us prove that D is indeed an IA-diagnoser for $\mathcal{A}$.

soundness For any word w, if $U_w = \emptyset$, by construction of $\mathsf{IA}(\mathcal{A})$, w is surely faulty. Assume now that $W_w = \emptyset$ and $U_w \neq \emptyset$. Let w' the greatest proper prefix of w such that $W_{w'} = \emptyset$. Let ρ be any signalling run with $\mathcal{P}(\rho) = w$. Assume that $\rho_{\downarrow |w'|}$ is faulty. Thus the states visited by $\rho_{\downarrow n}$ for $|w'| < n \leq |w|$ were always in $W_{\rho_{\downarrow n}}$. Since $W_w = \emptyset$, this not possible and so $\rho_{\downarrow |w'|}$ is correct. Thus every time a state with $W = \emptyset$, the length of the greatest prefix, for which all signalling subruns corresponding to this prefix are correct, is increased. This establishes soundness.

reactivity Let ρ be an infinite run such that $D_{sup}(\mathcal{P}(\rho)) =?$. Due to the characterization of Proposition 12, either (1) the SCC of $\mathcal{A}_{\mathsf{IA}}$ that ρ infinitely often visits is not a BSCC or (2) ρ does not visit infinitely often all the states of this SCC. The probability of such runs is null which establishes the reactivity.

◀

In the latter proof, when $\mathcal{A}$ is IA-diagnosable, an IA-diagnoser is built from the IA-automaton. We derive an upper bound on the size of IA-diagnosers.

▶ **Proposition 34.** *For every pLTS $\mathcal{A}$ with n_c correct states and n_f faulty states which is* IA*-diagnosable, one can build an* IA*-diagnoser with at most $2^{n_c}3^{n_f}$ states.*

The following lower bound can be derived from the proof of Proposition 27, since the pLTS of Figure 13 is IA-diagnosable, and because any IA-diagnoser is also an IF-diagnoser.

▶ **Proposition 35.** *There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of* IA*-diagnosable pLTS such that $\mathcal{A}_n$ has $2n+2$ states and it admits no* IA*-diagnoser with less than 2^n memory states.*

5 Predictability and prediagnosis

In this section we study predictability and introduce prediagnosis, a versatile combination of predictability and diagnosis.

5.1 Predictability

Fault predictability has been first introduced for LTS in [5]: in words, an LTS is predictable (resp. k-predictable) if a fault can be predicted (resp. at least before k observations) whatever the future behavior of the LTS. There are two possible adaptions for pLTS: (1) either one sticks to the original definition and requires that the fault surely occurs or, (2) one relaxes it and only requires that the fault almost surely occurs.

In order to reason about predictability, we introduce some particular prefixes of a run. For a finite run ρ, and $k \in \mathbb{N}$, we define $pre_k(\rho)$, the k-past of ρ, by $pre_k(\rho) = \rho_{\downarrow|\rho|_o - \min(k,|\rho|_o)}$. For example, in the pLTS of Figure 15, $pre_0(q_0bq_1\mathbf{f}q_2) = q_0bq_1$ as $\mathbf{f}$ is unobservable and $pre_1(q_0bq_1\mathbf{f}q_2) = q_0$. In fact for $k \geq 1$, $pre_k(q_0bq_1\mathbf{f}q_2) = q_0$.

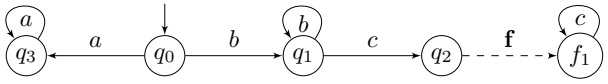

Figure 15 A 0-surely predictable and 1-predictable pLTS.

We also introduce sets of observed sequences defined on their possible future behaviors.

▶ **Definition 36** (ultimately possibly (significantly) correct)**.** Let σ be a finite observed sequence of a pLTS $\mathcal{A}$. Then:

- σ is *ultimately possibly correct* if $\{\rho' \in \Omega \mid \sigma \preceq \mathcal{P}(\rho')\} \cap \mathsf{C}_\infty \neq \emptyset$. The set of ultimately possibly correct observed sequences is denoted UPC.
- σ is *ultimately possibly significantly correct* if $\mathbb{P}(\{\rho' \in \Omega \mid \sigma \preceq \mathcal{P}(\rho')\} \cap \mathsf{C}_\infty) > 0$. The set of ultimately possibly significantly correct observed sequences is denoted UPSC.

▶ **Definition 37** ((sure) predictability)**.** Let $k \in \mathbb{N}$.

- A pLTS $\mathcal{A}$ is *k-surely predictable* if for every run $\rho\mathbf{f}q$ of $\mathcal{A}$, $\mathcal{P}(pre_k(\rho)) \notin$ UPC;
- A pLTS $\mathcal{A}$ is *k-predictable* if for every run $\rho\mathbf{f}q$ of $\mathcal{A}$, $\mathcal{P}(pre_k(\rho)) \notin$ UPSC.

Observe that in the previous definition, one can safely restrict to check the condition on correct runs ρ by considering the first occurrence of a fault in the run $\rho\mathbf{f}q$.

For example, the pLTS of Figure 15 is 0-surely predictable. Every correct run ρ that is followed by $\mathbf{f}$ is such that $\mathcal{P}(\rho) = b^n c$ for some $n \geq 1$. As it is the unique signalling run

with such an observed sequence, the fault can be predicted. It is not 1-surely predictable as the 1-past of $\rho = q_0bq_1cq_2\mathbf{f}f_1$ is $pre_1(\rho) = q_0bq_1$ and the infinite run $\rho' = q_0(bq_1)^\omega$ is correct. However it is 1-predictable as for every signalling run with observed sequence b^n for some $n \geq 1$ (thus ending in q_1) a fault eventually almost surely occurs. Finally it is not 2-predictable since the 2-past of $\rho = q_0bq_1cq_2\mathbf{f}f_1$ is q_0 and the infinite correct run $\rho = q_0(aq_3)^\omega$ has probability $\frac{1}{2}$.

5.1.1 Predictability versus diagnosability

While predictability seems to be a stronger requirement than FA-diagnosability, we show that this is only true for sure predictability. More generally, we compare diagnosability and predictability:

▶ **Theorem 38.** *The following relations between diagnosability and predictability hold.*
- *If a pLTS is k-surely predictable then it is k-predictable.*
- *If a pLTS is 0-surely predictable then it is* FA*-diagnosable.*
- *There exists a pLTS which is* FA*-diagnosable and not 0-predictable.*
- *There exists a pLTS which is k-predictable for every $k \in \mathbb{N}$ and not* IF*-diagnosable.*

Proof. The fact that sure-predictability implies predictability is immediate.
Assume that $\mathcal{A}$ is a 0-surely predictable pLTS. We claim that there is no ambiguous sequences. Let ρ be a finite faulty run decomposed as $\rho = \rho'\mathbf{f}\rho''$ with ρ' a correct prefix. Pick any run ρ^* such that $\mathcal{P}(\rho^*) = \mathcal{P}(\rho)$. Since $\mathcal{P}(\rho') \preceq \mathcal{P}(\rho^*)$, by definition of 0-surely predictability, ρ^* must be faulty. Thus 0-sure predictability implies FA-diagnosability.
The pLTS of Figure 22, on page 33, is not 0-predictable, yet it is FA-diagnosable as we detail later in Section 5.2.3.

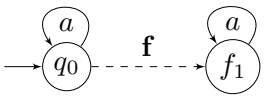

Figure 16 A k-predictable pLTS which is not IF-diagnosable.

Consider now the pLTS of Figure 16. For every $k \in \mathbb{N}$, it is k-predictable, because the probability of correct infinite sequences is zero. However it is not IF-diagnosable since a^ω is ambiguous, and $q_0\mathbf{f}f_1(af_1)^\omega$ has probability $\frac{1}{2}$. ◀

5.1.2 Complexity of predictability

We simultaneously establish a lower bound for the complexity of k-sure predictability and k-predictability problems.

▶ **Proposition 39.** *Let $k \in \mathbb{N}$. Deciding, given $\mathcal{A}$ a pLTS, whether $\mathcal{A}$ is surely-k predictable (resp. k-predictable) is* NLOGSPACE*-hard.*

Proof. We reduce the reachability problem in directed acyclic graphs which is NLOGSPACE-complete [8]. Let $G = (V, E)$ be a directed acyclic graph and $s, t \in V$. As shown in Figure 17, we transform G into a pLTS $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$ where
- $Q = V \uplus \{\mathsf{sink}_f, \mathsf{sink}_c\}$;
- $q_0 = s$
- $\Sigma_o = \{a\}$, $\Sigma_u = \{u, f\}$;

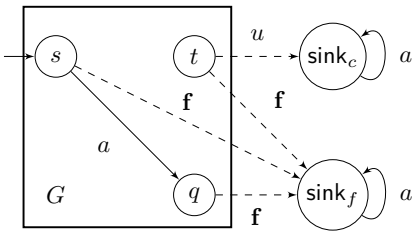

Figure 17 Reduction of the reachability problem.

- $T = \{(q, a, q') | (q, q') \in E\} \cup \{(q, f, \mathsf{sink}_f) \mid q \in V\} \cup \{(t, u, \mathsf{sink}_c)\} \cup \{(\mathsf{sink}_f, a, \mathsf{sink}_f)\} \cup \{(\mathsf{sink}_c, a, \mathsf{sink}_c)\}$;
- $\mathbf{P}$ is the matrix specifying the uniform probability for the transitions outgoing from any state.

This construction can be done in LOGSPACE.

Assume first that t is not reachable from s in G. Then due to acyclicity, all infinite paths end up in sink_f, implying $\mathsf{C}_\infty = \emptyset$. Thus $\mathcal{A}$ is k-surely predictable (and therefore k-predictable) for every $k \in \mathbb{N}$.

Assume now that t is reachable from s in G and let ρ be a correct signalling run that reaches t. Then $\rho\mathbf{f}\mathsf{sink}_f$ is a faulty run and the infinite correct run $\rho u \mathsf{sink}_c a(\mathsf{sink}_c a)^\omega$ has positive probability. Hence $\mathcal{A}$ is not 0-predictable. As a consequence, it is neither k-predictable nor surely k-predictable for every $k \in \mathbb{N}$. ◀

In order to determine an upper-bound for the complexity of k(-sure) predictability, we provide graph-based characterizations. In the sequel, in a graph, a *trivial strongly connected component* consists of a single state without loop.

▶ **Lemma 40.** *Let $\mathcal{A}$ be a pLTS and $k \in \mathbb{N}$. $\mathcal{A}$ is k-surely predictable if and only if there does not exist a pair of runs $q_0 \overset{\rho_0}{\Rightarrow} q_1$ and $q_0 \overset{\rho_0'}{\Rightarrow} q_1'$ such that:*

- $\mathcal{P}(\rho_0) = \mathcal{P}(\rho_0')$;
- $q_1 \overset{\rho_1}{\Rightarrow} q' \xrightarrow{\mathbf{f}} q$ *for some* $q' \in Q_c$ *with* $|\rho_1|_o \leq k$;
- $q_1' \overset{\rho_2}{\Rightarrow} q_2$, *for some* $q_2 \in Q_c$ *belonging to a non trivial SCC of* $\mathcal{A}$.

Proof. Assume such a pair (ρ_0, ρ_0') exists. Let $\rho = q_0 \overset{\rho_0}{\Rightarrow} q_1 \overset{\rho_1}{\Rightarrow} q'$ with $|\rho_1|_o \leq k$ and $\rho^* = q_0 \overset{\rho_0'}{\Rightarrow} q_1' \overset{\rho_2}{\Rightarrow} q_2$. Observe that $pre_k(\rho) \preceq \rho_0$, so that letting ρ_p be the signalling run $\rho^*_{\downarrow |pre_k(\rho)|_o}$, we have $\mathcal{P}(\rho_p) = \mathcal{P}(pre_k(\rho))$. Moreover, since q_2 is correct and belongs to a non trivial SCC, there is an infinite suffix ρ'' of ρ', ρ_p that only visits the (correct) states of this SCC after reaching q_2, and thus ρ'' is correct. This shows that $\mathcal{A}$ is not k-surely-predictable.

Assume now that $\mathcal{A}$ is not k-surely-predictable. Let $\rho\mathbf{f}q$ be a run such that ρ is correct and there exists ρ_p with $\rho_p \in \mathcal{P}^{-1}(\mathcal{P}(pre_k(\rho)))$ and $\rho' \in \Omega$ such that $\rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty$. We let q_1 be the state reached by $pre_k(\rho)$. In particular, $q_1 \in Q_c$, and $q_1 \overset{\rho_1}{\Rightarrow} q' \xrightarrow{\mathbf{f}} q$ where $\rho = pre_k(\rho) \cdot \rho_1$ and thus $q' \in Q_c$ and $|\rho_1|_o \leq k$. We let q_1' be the state reached by ρ_p. As ρ' is infinite, it ends up in a non trivial SCC of correct states. So we choose for q_2 the first state of this SCC reached by ρ' from q_1'. ◀

▶ **Lemma 41.** *$\mathcal{A}$ is k-predictable if and only if there does not exist a pair of runs $q_0 \overset{\rho_0}{\Rightarrow} q_1$ and $q_0 \overset{\rho_0'}{\Rightarrow} q_1'$ such that:*

- $\mathcal{P}(\rho_0) = \mathcal{P}(\rho_0')$;

- $q_1 \stackrel{\rho_1}{\Longrightarrow} q' \xrightarrow{f} q$ *for some* $q' \in Q_c$ *with* $|\rho_1|_o \leq k$;
- $q_1' \stackrel{\rho_2}{\Longrightarrow} q_2$, *for some* $q_2 \in Q_c$ *belonging to a BSCC of* $\mathcal{A}$.

Proof. Assume such a pair (ρ_0, ρ_0') exists. Let ρ be a run of the form $q_0 \stackrel{\rho_0}{\Longrightarrow} q_1 \stackrel{\rho_1}{\Longrightarrow} q'$ with $|\rho_1|_o \leq k$ and $\rho^* = q_0 \stackrel{\rho_0'}{\Longrightarrow} q_1' \stackrel{\rho_2}{\Longrightarrow} q_2$. Observe that $pre_k(\rho) \preceq \rho_0$, so that letting ρ_p be the signalling run $\rho^*_{\downarrow |pre_k(\rho)|_o}$, we have $\mathcal{P}(\rho_p) = \mathcal{P}(pre_k(\rho))$. Moreover, since q_2 is correct and belongs to a BSCC, $\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) \geq \mathbb{P}(C(\rho^*)) > 0$. This shows that $\mathcal{A}$ is not k-predictable.
Assume now that $\mathcal{A}$ is not k-predictable. Let $\rho \mathbf{f} q$ be a run such that ρ is correct and there exists ρ_p with $\rho_p \in \mathcal{P}^{-1}(\mathcal{P}(pre_k(\rho)))$ and $\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) > 0$. We let q_1 be the state reached by $pre_k(\rho)$. In particular, $q_1 \in Q_c$, and $q_1 \stackrel{\rho_1}{\Longrightarrow} q' \xrightarrow{\mathbf{f}} q$ where $\rho = pre_k(\rho) \cdot \rho_1$ and thus $q' \in Q_c$ and $|\rho_1|_o \leq k$. Letting q_1' be the state reached by ρ_p, we decompose the probability $\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\})$ depending on which BSCC runs ρ' almost surely hits:

$$\begin{aligned}&\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) \\ &\qquad = \sum_{\mathcal{C}\ BSCC \text{ reachable from } q_1'} \mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty \wedge \rho' \text{ ends in } \mathcal{C}\})\ .\end{aligned}$$

From $\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) > 0$, we deduce that there exists a BSCC $\mathcal{C}$ such that $\mathbb{P}(\{\rho' \in \Omega \mid \rho_p \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty \wedge \rho' \text{ ends in } \mathcal{C}\}) > 0$. Necessarily, $\mathcal{C} \subseteq Q_c$ and we let q_2 be any state of $\mathcal{C}$ to conclude. ◀

We are now in position to design efficient procedures for sure predictability and predictability decision problems.

▶ **Proposition 42.** *Deciding, given* $\mathcal{A}$ *a pLTS and* $k \in \mathbb{N}$, *whether* $\mathcal{A}$ *is* k*-surely predictable can be done in* NLOGSPACE.

Proof. We design a nondeterministic algorithm operating in logarithmic space to decide whether $\mathcal{A}$ is not k-surely predictable, using characterisation of Lemma 40. This will prove that sure predictability, the complementary problem, is also in NLOGSPACE by Immerman-Szelepscényi's theorem (that is also implicitely used in the rest of the proof).
First guess $q_1, q_1', q_2 \in Q_c$, and check that $q_1' \Rightarrow q_2$, with q_2 belonging to a non trivial SCC of $\mathcal{A}$, and q_1 and q_1' can be reached by runs with equivalent observation. All these checks can be done in NLOGSPACE. More precisely for the last property, one guesses a pair of observationally equivalent paths, using a counter bounded by n^2, with n the number of states of $\mathcal{A}$. In the positive case one guesses a run from q_1 that produces a fault with at most k observable events, using a counter bounded by k. ◀

▶ **Proposition 43.** *Deciding, given* $\mathcal{A}$ *a pLTS and* $k \in \mathbb{N}$, *whether* $\mathcal{A}$ *is* k*-predictable can be done in* NLOGSPACE.

Proof. We design a nondeterministic algorithm operating in logarithmic space to decide whether $\mathcal{A}$ is not k-predictable, using characterisation of Lemma 41.
First guess $q_1, q_1', q_2 \in Q_c$, and check that $q_1' \Rightarrow q_2$, q_2 belongs to a BSCC of $\mathcal{A}$, and q_1 and q_1' can be reached by runs with equivalent observation. All these checks can be done in NLOGSPACE. In the positive case one guesses a run from q_1 that produces a fault with at most k observable events, using a counter bounded by k. ◀

Summarizing the previous results, we get the following theorem.

▶ **Theorem 44.** *Deciding, given $\mathcal{A}$ a pLTS and $k \in \mathbb{N}$, whether $\mathcal{A}$ is k-predictable (resp. surely k-predictable) is an* NLOGSPACE*-complete problem. Moreover, the same complexity applies assuming k is fixed (and not given as input).*

5.1.3 Predictor synthesis

We now define predictors and sure predictors.

▶ **Definition 45.** A *k-predictor* (resp. *k-sure predictor*) is a function $D : \Sigma_o^* \to \{\top, ?\}$ such that:

soundness For every signalling run ρ such that $D(\mathcal{P}(\rho)) = \top$, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) = 0$ (resp. for every $\rho' \in \Omega$ such that $\rho \preceq \rho'$, $\rho' \notin \mathsf{C}_\infty$).

reactivity For every signalling run ρ, if there exists ρ' with $|\rho'|_o \leq k$ and $\rho\rho'$ faulty, then $D(\mathcal{P}(\rho)) = \top$.

The next propositions establish that the existence of k-predictors (resp. k-sure predictors) is equivalent to k-predictability (resp. k-sure predictability). In addition we show that there exist finite memory predictors with optimal exponential size.

▶ **Proposition 46.** *A pLTS $\mathcal{A}$ is k-surely predictable if and only if it admits a k-sure predictor. In the positive case, $\mathcal{A}$ admits a k-sure predictor with at most 2^{n_c} states where n_c is the number of correct states of $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a pLTS and $k \in \mathbb{N}$, and assume there exists a k-sure predictor D for $\mathcal{A}$.
Let $\rho_0 mathbf f q$ be a run of $\mathcal{A}$ such that ρ_0 is correct. The correct prefix ρ_0 decomposes into $\rho_0 = pre_k(\rho_0)\rho_0'$. Since D is reactive and because $pre_k(\rho_0)$ can be extended into $\rho_0 \mathbf{f} q$ which is faulty with $|\rho_0' \mathbf{f} q|_o \leq k$, necessarily $D(\mathcal{P}(pre_k(\rho_0)) = \top$. For every run $\rho_1 \in \mathcal{P}^{-1}(\mathcal{P}(pre_k(\rho_0)))$ observationally equivalent to $pre_k(\rho_0)$, then $D(\mathcal{P}(\rho_1)) = D(\mathcal{P}(\rho_0)) = \top$. By soundness, for every $\rho \in \Omega$ such that $\rho_1 \preceq \rho$ we have: $\rho \notin \mathsf{C}_\infty$. This proves that $\mathcal{A}$ is k-surely predictable.

Assume now that $\mathcal{A}$ is k-surely predictable. Let $\mathsf{IF}(\mathcal{A}) = \{Q^*, \Sigma_o, T^*, \{q_0\}\}$ be the IF-automaton of $\mathcal{A}$, and H_k be the set of correct states q of $\mathcal{A}$ such that there is a run $\rho \mathbf{f} q'$ starting from q in $\mathcal{A}$ with $|\rho|_o \leq k$. We define the finite memory diagnoser $\{Q^*, \Sigma_o, T', \{q_0\}, D_{fm}\}$, with finite memory $Q^* = 2^{Q_c}$, as follows. For every $a \in \Sigma_o$ and every $U \in Q^*$

- If $U \cap H_k = \emptyset$ then $T'(U, a) = T^*(U, a)$ else $T'(U, a) = U$;
- $D_{fm}(U) = \top$ if and only if $U \cap H_k \neq \emptyset$.

Let ρ be a run such that $D(\mathcal{P}(\rho)) = \top$. Since $T'(\{q_0\}, \mathcal{P}(\rho)) \cap H_k \neq \emptyset$, consider the first index i such that $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) \cap H_k \neq \emptyset$ which implies that $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$. By definition of D_{fm} and H_k, there is a run $\rho' \mathbf{f} q'$ starting from $q \in T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ with $|\rho|_o \leq k$ in $\mathcal{A}$. Thus there is a signalling run ρ'' starting from q_0 and reaching q with $\mathcal{P}(\rho'') = \mathcal{P}(\rho_{\downarrow i})$. Considering $\rho'' \cdot \rho' \mathbf{f} q'$, the k-sure predictability of $\mathcal{A}$ implies that every infinite run ρ^* with $\rho_{\downarrow i} \preceq \rho^*$ belongs to C and in particular those that fulfill $\rho \preceq \rho^*$. Therefore D is sound.
Let ρ be a run such that there exists ρ' with $\rho \cdot \rho'$ faulty and $|\rho'|_o \leq k$. By definition of H_k, $T^*(\{q_0\}, \mathcal{P}(\rho)) \cap H_k \neq \emptyset$. Consider the first index i such that $T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) \cap H_k \neq \emptyset$. Then $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ which implies that $T'(\{q_0\}, \mathcal{P}(\rho)) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ and that $D_{fm}(\mathcal{P}(\rho)) = \top$. This shows that D is reactive. ◀

▶ **Proposition 47.** *A pLTS $\mathcal{A}$ is k-predictable if and only if it admits a k-predictor. In the positive case, $\mathcal{A}$ admits a k-predictor with at most 2^{n_c} states where n_c is the number of correct states of $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a pLTS and $k \in \mathbb{N}$, and assume there exists a k-predictor D for $\mathcal{A}$. Let $\rho_0 f q$ be a run of $\mathcal{A}$ such that ρ_0 is correct. ρ_0 can be decomposed in $\rho_0 = pre_k(\rho_0)\rho_0'$. Then, as D is reactive and $pre_k(\rho_0)$ can be extended in $\rho_0 \mathbf{f} q$ which is faulty with $|\rho_0' \mathbf{f} q|_o \leq k$, $D(\mathcal{P}(pre_k(\rho_0)) = \top$. Let $\rho_1 \in \mathcal{P}^{-1}(\mathcal{P}(pre_k(\rho_0)))$, then $D(\mathcal{P}(\rho_1)) = D(\mathcal{P}(\rho_0)) = \top$. By soundness, $\mathbb{P}(\rho \in \Omega \mid \rho_1 \preceq \rho \wedge \rho \in \mathsf{C}_\infty) = 0$. This proves that $\mathcal{A}$ is k-predictable.

Assume now that $\mathcal{A}$ is k-predictable. Let $\mathsf{IF}(\mathcal{A}) = \{Q^*, \Sigma_o, T^*, \{q_0\}\}$ and H_k be the set of correct states q of $\mathcal{A}$ such that there is a run $\rho \mathbf{f} q'$ starting from q in $\mathcal{A}$ with $|\rho|_o \leq k$. We define the diagnoser $\{Q^*, \Sigma_o, T', \{q_0\}, D_{fm}\}$, with finite memory $Q^* = 2^{Q_c}$, as follows. For every $a \in \Sigma_o$ and every $U \in Q^*$

- If $U \cap H_k = \emptyset$ then $T'(U, a) = T^*(U, a)$ else $T'(U, a) = U$;
- $D_{fm}(U) = \top$ if and only if $U \cap H_k \neq \emptyset$.

Let ρ be a run such that $D(\mathcal{P}(\rho)) = \top$. Since $T'(\{q_0\}, \mathcal{P}(\rho)) \cap H_k \neq \emptyset$, consider the first index i such that $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) \cap H_k \neq \emptyset$ which implies that $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$. By definition of D_{fm} and H_k, there is a run $\rho' \mathbf{f} q'$ starting from $q \in T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ with $|\rho|_o \leq k$ in $\mathcal{A}$. Thus there is a signalling run ρ'' starting from q_0 and reaching q with $\mathcal{P}(\rho'') = \mathcal{P}(\rho_{\downarrow i})$. Considering $\rho'' \cdot \rho' \mathbf{f} q'$, the k-predicatibility of $\mathcal{A}$, implies that $\mathbb{P}(\{\rho^* \in \Omega \mid \rho_{\downarrow i} \preceq \rho^* \wedge \rho^* \in \mathsf{C}_\infty\}) = 0$ which then implies that $\mathbb{P}(\{\rho^* \in \Omega \mid \rho \preceq \rho^* \wedge \rho^* \in \mathsf{C}_\infty\}) = 0$. This shows that D is sound.

Let ρ be a run such that there exists ρ' with $\rho \cdot \rho'$ is faulty and $|\rho'|_o \leq k$. By definition of H_k, $T^*(\{q_0\}, \mathcal{P}(\rho)) \cap H_k \neq \emptyset$. Consider the first index i such that $T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) \cap H_k \neq \emptyset$. Then $T'(\{q_0\}, \mathcal{P}(\rho_{\downarrow i})) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ which implies that $T'(\{q_0\}, \mathcal{P}(\rho)) = T^*(\{q_0\}, \mathcal{P}(\rho_{\downarrow i}))$ and that $D_{fm}(\mathcal{P}(\rho)) = \top$. Therefore D is reactive. ◀

As the pLTS of Figure 15 is 0-surely predictable and 1-predictable, one can build with the methods showed in the proof of Proposition 46 and Proposition 47 a correct 0-sure predictor and a 1-predictor. This is done in Figure 18 where the states with double lines correspond to a verdict $\top$.

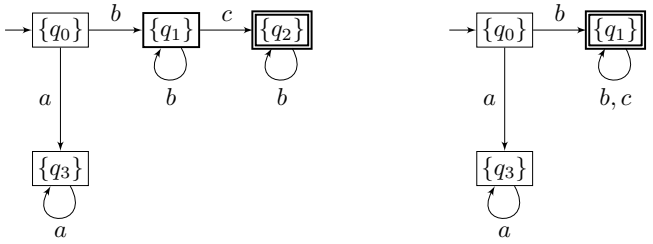

Figure 18 On the left a 0-sure predictor of Figure 15, on the right a 1-predictor.

A lower bound on the size of the two kinds of predictors can be obtained using the family of pLTS of Figure 19.

▶ **Proposition 48.** *There is a family $\{\mathcal{A}_n\}_{n \in \mathbb{N}}$ of 0-surely predictable pLTS where for every n, $\mathcal{A}_n$ has $2n + 2$ correct states, and $\mathcal{A}_n$ admits no 0-sure predictor (resp. no 0-predictor) with less than 2^n states.*

Proof. Consider the pLTS $\mathcal{A}$ of Figure 19 where $\Sigma_o = \{a, b, c\}$ and the initial state is q_0. Any correct run ρ immediately followed by a fault has an observed sequence that belongs to $\{a, b\}^* a \{a, b\}^{n-1} c$ and ρ is the single run with such an observed sequence so that $\mathcal{A}$ is 0-sure predictable.

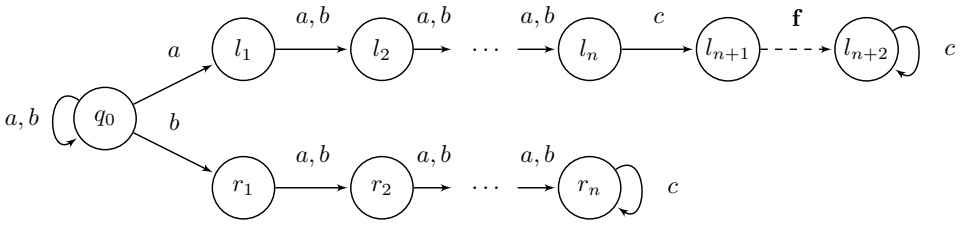

Figure 19 A family of 0-surely predictable pLTS requiring predictors with exponential size.

Intuitively, when a c is observed, any 0-predictor (sure or almost-sure) must have remembered the observable event that happened n steps earlier to know whether the system will commit a fault in the next step or not. Thus, as it does not know when a c will occur, it has to remember the last n letters that were observed as long as no c occurs.
Formally, assume that there exists a predictor D with less than 2^n states. Then there are two distinct words $w_1 \in \{a,b\}^n$ and $w_2 \in \{a,b\}^n$ leading to the same memory state. The words w_1 and w_2 differ at least from one letter say $w_1[i] = a$ and $w_2[i] = b$. Consider the signalling correct run ρ_1 corresponding to observed sequence $w_1 a^{i-1} c$ whose sequence of visited states is $q_0^i l_1 \dots l_n l_{n+1}$ and the signalling correct run ρ_2 corresponding to observed sequence $w_2 a^{i-1} c$ whose sequence of visited states is $q_0^i r_0 r_1 \dots r_n r_n$. They also lead to the same memory state. Since every infinite run extending ρ_2 is correct, by soundness $D(w_2 a^{i-1} c) =?$. But by reactivity, $D(w_1 a^{i-1} c) = \top$. Since $D(w_1 a^{i-1} c) = D(w_2 a^{i-1} c)$, such a predictor cannot exist. ◀

5.2 Prediagnosis

On the one hand, diagnosis is concerned with detection of faults that have occurred: given a sequence of observations a diagnoser tries to detect that a fault has occurred in the *past* of all consistent behaviors. On the other hand, prediction is concerned with anticipation of faults: given a sequence of observations a predictor tries to detect that a fault will eventually occur in the *future* of all consistent behaviors. The notion we introduce now, *prediagnosis*, concerns detection of faults both in the past and in the future.

Let us start by introducing two sets of infinite faulty runs that make prediagnosis impossible. FUPC_∞ is the set of faulty runs that admit for all their finite prefixes a compatible infinite correct run. The condition is strengthened for FUPSC_∞ which gathers the faulty runs that admit for all their finite prefixes, a positive measure of compatible infinite correct runs.

▶ **Definition 49.** Let $\mathcal{A}$ be a pLTS. Then:
- FUPC_∞, the set of *faulty, ultimately possibly correct* runs is defined by:
  $\mathsf{FUPC}_\infty = \{\rho \in \Omega \mid \rho \text{ faulty and } \forall i \in \mathbb{N},\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPC}\}$
- FUPSC_∞, the set of *faulty, ultimately possibly significantly correct* runs is defined by:
  $\mathsf{FUPSC}_\infty = \{\rho \in \Omega \mid \rho \text{ faulty and } \forall i \in \mathbb{N},\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPSC}\}$

The reactivity requirement for prediagnosers will impose that these sets are negligible. The difference between these two sets impacts correctness: relying on FUPC_∞ provides a *sure* correctness while relying on FUPSC_∞ only provides an *almost sure* correctness.

▶ **Definition 50** ((Sure) Prediagnosability)**.** Let $\mathcal{A}$ be a pLTS. Then:

- $\mathcal{A}$ is *surely prediagnosable* if $\mathbb{P}(\mathsf{FUPC}_\infty)=0$;
- $\mathcal{A}$ is *prediagnosable* if $\mathbb{P}(\mathsf{FUPSC}_\infty)=0$.

5.2.1 Prediagnosability versus diagnosability and predictability

Since prediagnosis does not provide information about correct runs, we only compare it with the two notions of fault diagnosability, IF- and FF-diagnosability, and also to (sure) predictability. Surprisingly, sure prediagnosability lies strictly between FF-diagnosability and IF-diagnosability with equivalence for finite branching pLTS. Also (sure) 0-predictability implies (sure) prediagnosability. As expected, the less demanding specification is prediagnosability.

▶ **Theorem 51.** *Let $\mathcal{A}$ be a pLTS. The following relations between prediagnosability, diagnosability and predictability hold.*

- *If $\mathcal{A}$ is* FF*-diagnosable then it is surely prediagnosable. There exists an infinitely branching pLTS which is surely prediagnosable but not* FF*-diagnosable.*
- *If $\mathcal{A}$ is surely prediagnosable then it is* IF*-diagnosable. There exists an infinitely branching pLTS which is* IF*-diagnosable but not surely prediagnosable.*
- *If $\mathcal{A}$ is surely prediagnosable then it is prediagnosable. There exists a finite pLTS which is prediagnosable and not* IF*-diagnosable (and thus not surely prediagnosable).*
- *If $\mathcal{A}$ is* 0*-predictable (resp. surely* 0*-predictable) then it is prediagnosable (resp. surely prediagnosable). There exists a finite pLTS which is surely prediagnosable and not* 0*-predictable.*

Proof. Let us define $\{\mathsf{FUPC}_n\}_{n\in\mathbb{N}}$ by: $\mathsf{FUPC}_n = \{\rho \in \Omega \mid \rho_{\downarrow n}$ faulty and $\forall i \in \mathbb{N}\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPC}\}$. Observe that this sequence of sets is non-decreasing and $\mathsf{FUPC}_\infty = \bigcup_{n\in\mathbb{N}} \mathsf{FUPC}_n$. Therefore, $\mathbb{P}(\mathsf{FUPC}_\infty) = \lim_{n\to\infty} \mathbb{P}(\mathsf{FUPC}_n)$.

Now, observe that for every $n \in \mathbb{N}$, $\mathsf{FUPC}_n \subseteq \mathsf{FAmb}_n$. Thus $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) = 0$ implies $\lim_{n\to\infty} \mathbb{P}(\mathsf{FUPC}_n) = 0$ which shows the first implication.
Consider the infinitely branching pLTS of Figure 20. It contains no infinite correct sequence, so that $\mathsf{UPC} = \emptyset$, and this pLTS is surely prediagnosable. On the other hand, for every $n \geq 1$, $\mathbb{P}(\mathsf{FAmb}_n) = \frac{1}{2}$, so that this pLTS is not FF-diagnosable.

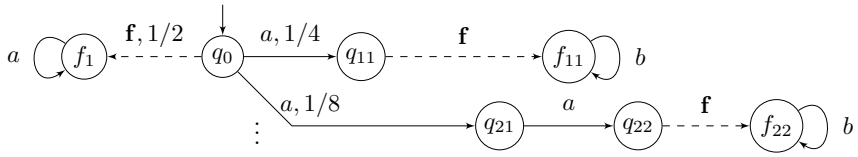

Figure 20 An infinitely branching pLTS that is surely prediagnosable but not FF-diagnosable.

In order to show the second implication, observe that for every n, $\mathsf{FAmb}_\infty \cap \mathsf{FAmb}_n \subseteq \mathsf{FUPC}_n$. Thus $\lim_{n\to\infty} \mathbb{P}(\mathsf{FUPC}_n) = 0$ implies $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \cap \mathsf{FAmb}_n) = 0$. Now, since $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0$ (see Lemma 5), this implies $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$.
Consider the infinitely branching IF-diagnosable pLTS of Figure 2, on page 7. For every $n \geq 1$, $a^n \in \mathsf{UPC}$ due to the infinite run $q_0 a q_{n1} a \dots q_{nn-1} a (q_{nn} b)^\omega$. Thus $\mathbb{P}(\mathsf{FUPC}_n) = \frac{1}{2}$ and this pLTS is not surely prediagnosable.

Since $\mathsf{UPSC} \subseteq \mathsf{UPC}$, the third implication is immediate.

Consider the pLTS of Figure 21. It is not IF-diagnosable since the infinite faulty run $q_0\mathbf{f}(f_1a)^\omega$ is ambiguous and has probability $\frac{1}{2}$. However since the set of infinite correct runs is negligible, $\mathsf{UPSC} = \emptyset$ and so this pLTS is prediagnosable.

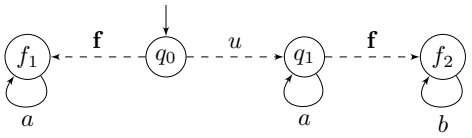

Figure 21 A pLTS which is prediagnosable but not IF-diagnosable.

Assume now that $\mathcal{A}$ is 0-predictable (resp. 0-surely predictable). Let $\rho = \rho'\mathbf{f}\rho''$ be an infinite faulty run. Due to predictability, $\mathcal{P}(\rho') \notin \mathsf{UPSC}$ (resp. $\mathcal{P}(\rho') \notin \mathsf{UPC}$). Since $\mathcal{P}(\rho'') = \mathcal{P}(\rho_{\downarrow i})$ for some i, $\rho \notin \mathsf{FUPSC}_\infty$ (resp. $\rho \notin \mathsf{FUPC}_\infty$). Thus $\mathsf{FUPSC}_\infty = \emptyset$ (resp. $\mathsf{FUPC}_\infty = \emptyset$) which implies that $\mathcal{A}$ is prediagnosable (resp. surely prediagnosable).
Consider the pLTS of Figure 1, on page 6. It is not 0-predictable since the observation of the faulty run $\rho = q_0\mathbf{f}f_1$ is ε, and infinite correct runs have probability $\frac{1}{2} > 0$. On the other hand, let $\rho = \rho' b \rho''$ be a faulty run and let $\sigma = \mathcal{P}(\rho')b$. Then $\{\rho^* \in \Omega \mid \rho^* \text{ correct and } \sigma \preceq \mathcal{P}(\rho^*)\} = \emptyset$. So $\rho \notin \mathsf{FUPC}_\infty$. Since the probability that an infinite run is faulty and does not produce a b is null, this pLTS is surely prediagnosable. ◀

As an immediate consequence of the latter theorem and of Corollary 6, we get:

▶ **Corollary 52.** *A finitely branching pLTS $\mathcal{A}$ is* IF*-diagnosable if and only if it is surely prediagnosable.*

5.2.2 Complexity of prediagnosability

In order to study the complexity of prediagnoability, we introduce some more notations. Given a pLTS $\mathcal{A}$, we let $H_c \subseteq Q_c$ be the subset of correct states from which one cannot reach a BSCC of $\mathcal{A}$ included in Q_c. Observe that for any finite run ρ that ends in H_c, $\mathbb{P}(\{\rho' \in \Omega \mid \rho' \text{ correct and } \rho \preceq \rho'\}) = 0$. The set H_c is the key to a characterization of prediagnosability.

▶ **Lemma 53.** *Let $\mathcal{A}$ be a finite pLTS and $\mathcal{A}_{\mathsf{IF}}$ its* IF*-automaton. Then $\mathcal{A}$ is prediagnosable if and only if there no BSCC of $\mathcal{A}_{\mathsf{IF}}$ contains a state (q, U) such that $q \in Q_f$ and $U \setminus H_c \neq \emptyset$.*

Proof. Assume there exists a state (q, U) of a BSCC $\mathcal{C}$ of $\mathcal{A}_{\mathsf{IF}}$ such that $q \in Q_f$ and $U \setminus H_c \neq \emptyset$. By definition of H_c and the fact that $\mathcal{C}$ is strongly connected, every state $(q', U') \in \mathcal{C}$ fulfills $q' \in Q_f$ and $U' \setminus H_c \neq \emptyset$. Let ρ be an infinite run that ending up $\mathcal{C}$. Pick any index i_0 such that $\rho_{\downarrow i_0}$ reaches $\mathcal{C}$. Thus for all $i \geq i_0$, there is a finite correct run ρ' with $\mathcal{P}(\rho') = \mathcal{P}(\rho_{\downarrow i})$ and ρ' ends in a correct state out of H_c. So there is a finite run ρ'' with $\rho' \preceq \rho''$ and ρ'' ends in a correct BSCC of $\mathcal{A}$. This establishes that $\mathcal{P}(\rho_{\downarrow i})$ belongs to UPSC and thus any infinite run that ends up in $\mathcal{C}$ belongs to FUPSC_∞. Since the probability that an infinite run ends up in $\mathcal{C}$ is positive, $\mathcal{A}$ is not prediagnosable.
Assume there exists no BSCC of $\mathcal{A}_{\mathsf{IF}}$ containing a state (q, U) with $q \in Q_f$ and $U \setminus H_c \neq \emptyset$. Pick an arbitrary BSCC $\mathcal{C}$ of $\mathcal{A}_{\mathsf{IF}}$ and ρ any infinite run that ends up in $\mathcal{C}$. Pick any index i such that $\rho_{\downarrow i}$ reaches $\mathcal{C}$. Thus for every correct signalling run ρ' such that $\mathcal{P}(\rho') = \mathcal{P}(\rho_{\downarrow i})$, ρ' ends in a correct state of H_c which implies that $\mathcal{P}(\rho_{\downarrow i}) \notin \mathsf{UPSC}$ and $\rho \notin \mathsf{FUPSC}_\infty$. Since almost surely infinite runs end up in BSCC, $\mathcal{A}$ is prediagnosable. ◀

Since for pLTS prediagnosability is as demanding as IF-diagnosability, and in particular for the pLTS used to show PSPACE-hardness of IF-diagnosability prediction does not help, one obtains the following complexity result.

▶ **Theorem 54.** *The prediagnosability problem is* PSPACE*-complete.*

Proof. Let us look at $\mathcal{A}$, the pLTS of Figure 11 on page 15, used in the reduction from eventual universality to IF-diagnosability. All correct states reach a correct BSCC of $\mathcal{A}$ (which is one of the BSCC of the NFA). Thus $H_c = \emptyset$ and so $\mathcal{A}$ is IF-diagnosable if and only if it is prediagnosable. This establishes that prediagnosability is PSPACE-hard.
Let us design a non deterministic procedure that checks in polynomial space whether a pLTS is not prediagnosable. First one computes in polynomial time the set H_c. Then one guesses a state (q, U) of $\mathcal{A}_{\mathsf{IF}}$ with $q \in Q_f$ and $U \setminus H_c \neq \emptyset$. Then one non deterministically checks that (q, U) is reachable and belongs to a BSCC. All these subprocedures can be done in polynomial space, and we use Savitch's theorem to conclude. ◀

5.2.3 Prediagnoser synthesis

Prediagnosis allows to announce faults earlier than diagnosis. The corresponding class of monitors, called prediagnosers, is now introduced. As for diagnosers, we will assume that (sure) prediagnosers commit to the $\top$ verdict. Intuitively, sure prediagnosers are IF-diagnosers with the capacity of prediction of sure predictors, while prediagnosers extend IF-diagnosers with the capacity of almost-sure predictors.

▶ **Definition 55.** A *sure prediagnoser* (resp. *prediagnoser*) is a function $D : \Sigma_o^* \to \{\top, ?\}$ such that

soundness For every signalling run ρ, if $D(\mathcal{P}(\rho)) = \top$ then $\{\rho' \in \Omega \mid \rho \preceq \rho'\} \subseteq \mathsf{Sf}_\infty$ (resp. $\mathbb{P}(\rho' \in \Omega \mid \rho \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty) = 0$).

reactivity For every finite faulty run ρ, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) = 0$ where for $w \in \Sigma_o^\omega$, $D(w) = \lim_{n\to\infty} D(w_{\leq n})$.

While sure prediagnosability and IF-diagnosability are equivalent for finitely branching pLTS, there are differences between sure prediagnosers and IF-diagnosers. An IF-diagnoser is a sure prediagnoser, but a sure prediagnoser may output a verdict $\top$ even before a fault. This phenomenon occurs even if the pLTS is non predictable. The non predictable pLTS of Figure 22 points out this difference. A diagnoser can output a $\top$ only after observing two a's, since then surely a fault occurred while a sure prediagnoser can already output a $\top$ after observing the first a. In fact this pLTS is FA-diagnosable since after an occurrence of b, the run is surely correct.

As we wished, the existence of a (sure) prediagnoser is equivalent to the (sure) prediagnosability. Moreover, we provide upper bounds on the size of (sure) prediagnosers.

▶ **Proposition 56.** *A pLTS $\mathcal{A}$ is surely prediagnosable if and only if it admits a sure prediagnoser. In the positive case, $\mathcal{A}$ admits a sure prediagnoser with at most 2^{n_c} states where n_c is the number of correct states of $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a surely prediagnosable pLTS, and $\mathsf{IF}(\mathcal{A}) = (Q^*, \Sigma_o, T^*, \{q_0\})$ its IF-automaton. Consider H_s, the set of correct states q of $\mathcal{A}$ such that there is no run $q\rho q'$ with q' belonging to a correct non trivial SCC of $\mathcal{A}$. We define the finite memory diagnoser $D = (Q^*, \Sigma, T', \{q_0\}, D_{fm})$ where:

- $\forall a \in \Sigma_o, T'(U, a) = U'$ iff $(U, a, U') \in T^*$ and $U \setminus H_s \neq \emptyset$, $T'(U, a) = U$ if $U \setminus H_s = \emptyset$,

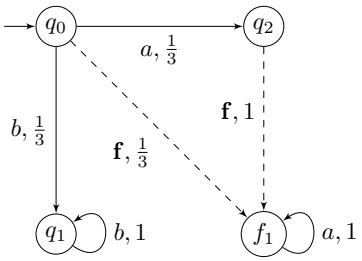

Figure 22 A non-predictable pLTS, for which a sure prediagnoser is quicker than all diagnosers.

- $D_{fm}(U) = \top$ iff $U \setminus H_s = \emptyset$.

D is a finite-memory monitor of size at most 2^{n_c} where n_c is the number of correct states of $\mathcal{A}$. Let us show that D is a sure prediagnoser. It is sound as for every run ρ such that $D(\mathcal{P}(\rho)) = \top$, by definition of H_s, the only reachable non trivial SCC are faulty and, as an infinite run ends in a non trivial SCC, $\{\rho' \in \Omega \mid \rho \preceq \rho'\} \cap \mathsf{C}_\infty = \emptyset$. Let ρ be a finite faulty run and $\rho' \in \Omega$ such that $\rho' \in \{\rho'' \in \Omega \mid \rho \preceq \rho'' \wedge D(\mathcal{P}(\rho'')) = ?\}$. Then for every $i \in \mathbb{N}$, the state U of D reached after observing $\mathcal{P}(\rho'_{\downarrow i})$ satisfies $U \setminus H_s \neq \emptyset$. Thus $\mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPC}$ and $\rho' \in \mathsf{FUPC}_\infty$. Therefore, $\mathbb{P}(\{\rho'' \in \Omega \mid \rho \preceq \rho'' \wedge D(\mathcal{P}(\rho'')) = ?\}) \leq \mathbb{P}(\mathsf{FUPC}_\infty) = 0$ as $\mathcal{A}$ is surely prediagnosable. So D is reactive.

Assume now that $\mathcal{A}$ admits a sure prediagnoser D. Let us recall $\mathsf{FUPC}_\infty = \bigcup_{n\in\mathbb{N}} \mathsf{FUPC}_n$ where $\mathsf{FUPC}_n = \{\rho \in \Omega \mid \rho_{\downarrow n}$ faulty and $\forall i \in \mathbb{N}\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPC}\}$. We claim that for all $n \in \mathbb{N}$, $\mathbb{P}(\mathsf{FUPC}_n) = 0$.

Since D is sound, for any signalling faulty run ρ of observable length n,

$$\{\rho' \in \Omega \mid \rho \preceq \rho' \text{ and } \forall i \in \mathbb{N}\ \mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPC}\} \subseteq \{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) = ?\} \ .$$

Since D is reactive $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) = ?\}) = 0$. Thus:

$$\mathbb{P}(\mathsf{FUPC}_n) = \sum_{\rho \text{ faulty signalling run} \wedge |\rho|_o = n} \mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \text{ and } \forall i \in \mathbb{N}\ \mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPC}\}) = 0 \ .$$

This proves that $\mathcal{A}$ is surely prediagnosable. ◀

Following the same lines, one can show a similar statement for prediagnosability.

▶ **Proposition 57.** *A pLTS $\mathcal{A}$ is prediagnosable if and only if it admits a prediagnoser. In the positive case, $\mathcal{A}$ admits a prediagnoser with at most 2^{n_c} states where n_c is the number of correct states of $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a prediagnosable pLTS, and $\mathsf{IF}(\mathcal{A}) = (Q^*, \Sigma, T^*, \{q_0\})$ its IF-automaton. We define the finite memory diagnoser $D = (Q^*, \Sigma, T', \{q_0\}, D_{fm})$ where:

- $\forall a \in \Sigma, T'(U, a) = U'$ iff $(U, a, U') \in T^*$ and $U \setminus H_c \neq \emptyset$, $T'(U, a) = U$ if $U \setminus H_c = \emptyset$,
- $D_{fm}(U) = \top$ iff $U \setminus H_c = \emptyset$.

D is a finite-memory monitor of size at most 2^{n_c} where n_c is the number of correct states of $\mathcal{A}$. Let us show that D is a prediagnoser. It is sound as for every run ρ such that $D(\mathcal{P}(\rho)) = \top$, by definition of H_c, the only reachable BSCC are faulty and, as runs almost surely end in BSCCs, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge \rho' \in \mathsf{C}_\infty\}) = 0$. Let ρ be a finite faulty run and $\rho' \in \Omega$ such that $\rho' \in \{\rho'' \in \Omega \mid \rho \preceq \rho'' \wedge D(\mathcal{P}(\rho'')) = ?\}$. Then for every $i \in \mathbb{N}$, the state U of D reached after observing $\mathcal{P}(\rho'_{\downarrow i})$ satisfies $U \setminus H_c \neq \emptyset$. Thus $\mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPSC}$ and $\rho' \in \mathsf{FUPSC}_\infty$. Therefore,

$\mathbb{P}(\{\rho'' \in \Omega \mid \rho \preceq \rho'' \wedge D(\mathcal{P}(\rho'')) =?\}) \leq \mathbb{P}(\mathsf{FUPSC}_\infty) = 0$ as $\mathcal{A}$ is prediagnosable. So D is reactive.
Assume now that $\mathcal{A}$ admits a prediagnoser D. Let us define $\mathsf{FUPSC}_n = \{\rho \in \Omega \mid \rho_{\downarrow n}$ faulty and $\forall i \in \mathbb{N}\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPSC}\}$. Observe that $\mathsf{FUPSC}_\infty = \bigcup_{n\in\mathbb{N}} \mathsf{FUPSC}_n$. We claim that for all $n \in \mathbb{N}$, $\mathbb{P}(\mathsf{FUPSC}_n) = 0$.
Since D is sound, for any signalling faulty run ρ of observable length n,

$$\{\rho' \in \Omega \mid \rho \preceq \rho' \text{ and } \forall i \in \mathbb{N}\ \mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPSC}\} \subseteq \{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}\ .$$

Since D is reactive $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\mathcal{P}(\rho')) =?\}) = 0$. Thus:

$$\mathbb{P}(\mathsf{FUPSC}_n) = \sum_{\rho \text{ faulty signalling run} \wedge |\rho|_o = n} \mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \text{ and } \forall i \in \mathbb{N}\, \mathcal{P}(\rho'_{\downarrow i}) \in \mathsf{UPSC}\}) = 0\ .$$

This proves that $\mathcal{A}$ is prediagnosable.

◀

Let us illustrate the construction of prediagnosers on the pLTS of Figure 23. It is

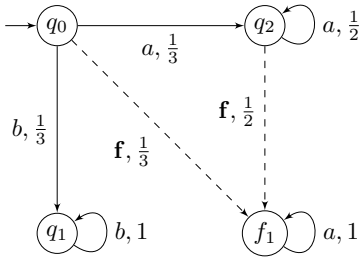

Figure 23 A non-predictable, non diagnosable yet prediagnosable pLTS.

prediagnosable, and its associated prediagnoser (as defined in the proof of Proposition 57) is depicted in Figure 24. A fault is announced when the monitor reaches its belief state $\{q_2\}$.

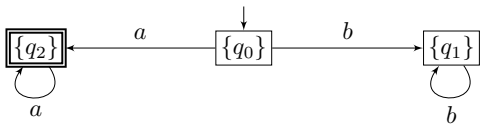

Figure 24 A prediagnoser for the pLTS of Figure 23.

To obtain a lower bound on the size of prediagnosers, observe that predictivity does not help on the example of Figure 13, on page 20. Therefore, for this family of pLTS, any (sure) prediagnoser is also an IF-diagnoser. Relying on the proof of Proposition 27 we thus obtain:

▶ **Proposition 58.** *There is a family $(\mathcal{A}_n)$ of (surely) prediagnosable pLTS, such that $\mathcal{A}_n$ has $n+1$ correct states and $\mathcal{A}_n$ admits no (sure) prediagnoser with less than 2^n states.*

Prediagnosers, can be thought of as monitors that emit verdicts as soon as possible, while preserving soundness. The prediagnosers built in the proofs of Propositions 56 and 57 are indeed optimal in that sense.

6 Conclusion

In this work, we settled the foundations of diagnosability and predictability for partially observed stochastic systems. In particular, we investigated semantical issues and provided several meaningful definitions for diagnosability and predictability in a probabilistic context. We also introduced prediagnosability, that combines the advantages of diagnosability and predictability. Beyond providing relations between these notions, we obtained tight complexity bounds using graph-based characterizations on the product of the system under scrutiny and an appropriate monitor. The complexity ranges from NLOGSPACE-completeness for predictability to PSPACE-completeness for diagnosability and prediagnosability, as summarized on Figure 25. Last, we proved exponential almost matching lower and upper bounds for the diagnosers, predictors, and prediagnosers synthesis problems.

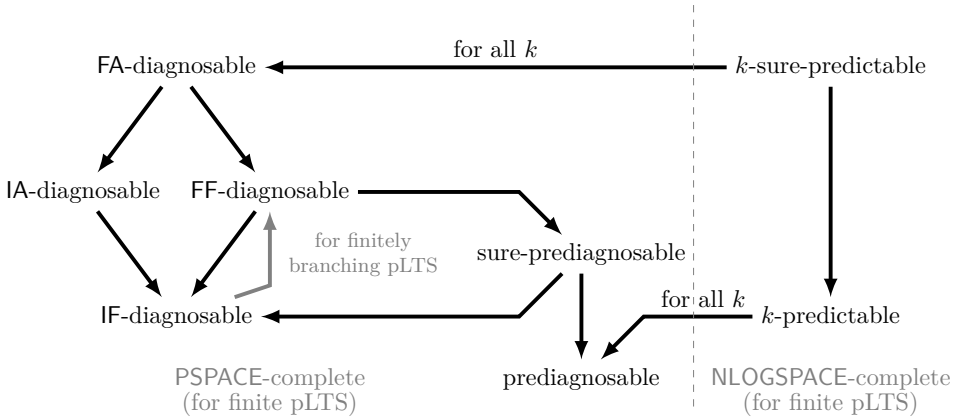

Figure 25 Summarizing relations between specifications, and associated complexities.

The present contribution opens several interesting research perspectives. First of all, the decidability status (and in the positive case, the precise complexity) of the approximate diagnosability (AA-diagnosability) introduced in [13] is still open since we only proved the algorithm from [3] to be erroneous. Second, beyond diagnosability and its variants (predictability and prediagnosability), we wish to conduct a systematic study of other paradigms related to partial observability, such as opacity or detectability, in a probabilistic context. Last, we plan to move to more quantitative versions of diagnosis including optimization issues. The objective would be to minimize the observational capacities of the monitor, either spatially or timely by restricting either the observable actions, or the observation time instants, while preserving diagnosability.

References

1 N. Bertrand, E. Fabre, S. Haar, S. Haddad, and L. Hélouët. Active diagnosis for probabilistic systems. In *Proceedings of FoSSaCS'14*, volume 8412 of *Lecture Notes in Computer Science*, pages 29–42. Springer, 2014.

2 B.G. Buchanan and E.H. Shortliffe. *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley, 1984.

3 J. Chen and R. Kumar. Polynomial test for stochastic diagnosability of discrete-event systems. *IEEE Transactions on Automation Science and Engineering*, 10(4):969–979, 2013.

4 L. Doyen, T. A. Henzinger, and J-F. Raskin. Equivalence of labeled markov chains. *International Journal of Foundations of Computer Science*, 19(3):549–563, 2008.

5 S. Genc and S. Lafortune. Predictability of event occurrences in partially-observed discrete-event systems. *Automatica*, 45(2):301–311, 2009.

6 S. Haar, S. Haddad, T. Melliti, and S. Schwoon. Optimal constructions for active diagnosis. In *Proceedings of FSTTCS'13*, volume 24 of *LIPIcs*, pages 527–539. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.

7 S. Jiang, Z. Huang, V. Chandra, and R. Kumar. A polynomial algorithm for testing diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 46(8):1318–1321, 2001.

8 N.D. Jones. Space-bounded reducibility among combinatorial problems. *Journal of Computer and System Sciences*, 11(1):68–85, 1975.

9 A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *Proceedings of the 13th Annual Symposium on Switching and Automata Theory (Swat 1972)*, SWAT '72, pages 125–129, Washington, DC, USA, 1972. IEEE Computer Society.

10 N. Rampersad, J. Shallit, and Z. Xu. The computational complexity of universality problems for prefixes, suffixes, factors, and subwords of regular languages. *Fundam. Inf.*, 116(1-4):223–236, January 2012.

11 M. Sampath, S. Lafortune, and D. Teneketzis. Active diagnosis of discrete-event systems. *IEEE Transactions on Automatic Control*, 43(7):908–929, 1998.

12 M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *IEEE Trans. Aut. Cont.*, 40(9):1555–1575, 1995.

13 D. Thorsley and D. Teneketzis. Diagnosability of stochastic discrete-event systems. *IEEE Transactions on Automatic Control*, 50(4):476–492, 2005.

A An erroneous PTIME procedure for A-diagnosability

This algorithm of [3] is an adaptation of the one in [7] which decides diagnosability for deterministic discrete-event systems in PTIME. In [3], the correct behaviour is specified by a deterministic automaton and a fault consists in producing a word not in the language. Once the synchronized product between the system and the specification is performed, this framework boils down to ours. So we describe how the procedure proposed in [3] checks the A-diagnosability in our framework. First it builds a *testing* sub-pLTS (which means that the probability of the outgoing edges from a state is less or equal than 1) as follows.

- A state is a pair (q_1, q_2) where $q_1 \in Q$ and $q_2 \in Q_c$. The initial state is (q_0, q_0) where q_0 is the initial state of the pLTS.
- Transitions are labelled by observable events.
- There is a transition labelled by a from (q_1, q_2) to (q_1', q_2') with probability $p_1 p_2 > 0$ if:
  1. the sum of the probabilities of signalling runs from q_1 to q_1' and labelled by a is equal to p_1;
  2. assuming that the set of correct signalling runs from q_2 and labelled by a is non empty, then the sum of the probabilities of (correct) signalling paths from q_2 to q_2' and labelled by a, conditionned over all correct signalling paths from q_2 and labelled by a, is equal to p_2.

Then the pLTS is A-diagnosable if there is no recurrent class (i.e. a bottom strongly connected component with for all states the sum of outgoing probabilities being equal to 1) in the testing sub-pLTS where the first components of the states belong to Q_f.

We have illustrated in Figure 26 the testing sub-pLTS associated with the A-diagnosable pLTS of Figure 1. From q_0 by observing a, one reaches q_1, f_1 or f_2. Thus from (q_0, q_0) we have three outgoing transitions whose first components are q_1, f_1 or f_2 and second component is q_1, the only non faulty reachable faulty state. From q_1, one stays in q_1 observing a which entails a loop around (q_1, q_1). From f_1 reading a leads either to itself or f_2 which entails transition toward (f_1, q_1) and another toward (f_2, q_1). From f_2, only b is observable which cannot observed from q_1. Thus (f_2, q_1) has no outgoing state. In this sub-pLTS, the single recurring class is $\{(q_1, q_1)\}$ whose first component is correct. Thus the algorithm soundly returns that the pLTS is A-diagnosable.

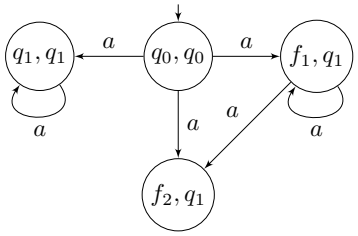

Figure 26 The testing sub-pLTS for the pLTS of Fig. 1

Let us look at the pLTS of Figure 27. There is a single faulty run $q_0 f_1^\omega$ and it has a positive probability. This run can be mimicked by the correct run $q_0 q_1^\omega$. So this pLTS is not A-diagnosable.

We have illustrated in Figure 28 the testing sub-pLTS associated with the pLTS of Figure 27. The single recurring class is $\{(q_2, q_2)\}$. So the algorithm incorrectly returns that this pLTS is diagnosable.

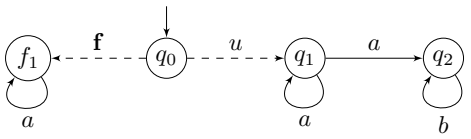

Figure 27 A pLTS which is not A-diagnosable

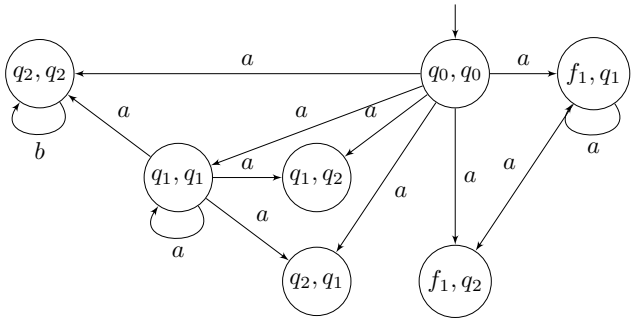

Figure 28 The testing sub-pLTS for the pLTS of Fig. 27

Since the proof of correctness is provided in [3], let us point out where is the error. The error is located in the proof of sufficiency. Equation (2) in page 973 claims that given a faulty run s the probability of a continuing run t in the LTS such that $\mathcal{P}(st)$ is ambiguous is less or equal than the sum of the probability of continuing runs $(\mathcal{P}(t), \mathcal{P}(t'))$ following some $(\mathcal{P}(s), \mathcal{P}(s'))$ in the testing sub-pLTS. Now consider the run $s = q_0 \mathbf{f} f_1 a f_1$ and the continuing run $t = f_1 a f_1 a f_1$. Its probability is equal to 1. The corresponding runs in the sub-pLTS are $(f_1, q_1)a(f_1, q_1)a(f_1, q_2)$ and $(f_1, q_1)a(f_1, q_1)a(f_1, q_1)$ with probability $\frac{1}{2} = \frac{1}{4} + \frac{1}{4}$. So Equation (2) does not hold.

B An erroneous PTIME procedure for AA-diagnosability

In [3] a polynomial time algorithm is also provided that deals with another kind of diagnosticability the authors called SS-diagnosticability (which was first introduced in [13] as AA-diagnosticability). Unfortunately this algorithm is also wrong as we establish below. So to the best of our knowledge, the complexity of AA-diagnosability remains an open problem.

▶ **Definition 59** ([3]). A pLTS $\mathcal{A}$ is AA-diagnosable if and only if for all $\varepsilon > 0, \tau > 0$ there exists $n_{\varepsilon,\tau} \in \mathbb{N}$ such that for every faulty run ρ and all $n \geq n_{\varepsilon,\tau}$ we have:

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) > \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho'\})} < \tau$$

$$\text{where } \mathbb{P}_{amb}(\sigma) = \frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{|\sigma|} \mid \mathcal{P}(\rho') = \sigma \wedge \rho' \in \mathsf{C}_n\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{|\sigma|} \mid \mathcal{P}(\rho') = \sigma\})}$$

We now introduce an equivalence between (fully observed) pLTS that is used by the algorithm we are going to detail.

▶ **Definition 60.** Let $\mathcal{A}_i = \langle Q_i, \pi_{0,i}, \Sigma_o, T_i, \mathbf{P}_i \rangle$ for $i \in \{1,2\}$ be two pLTS (where $\pi_{0,i}$ denotes the initial distribution). Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are p-equivalent if for all word $w \in \Sigma_o^*$,

$$\mathbb{P}_1(\rho \mid \sigma_\rho = w) = \mathbb{P}_2(\rho \mid \sigma_\rho = w)$$

with $\mathbb{P}_i$ the probability induced by the pLTS $\mathcal{A}_i$.

As done in [4], by adapting a classical procedure on probabilistic finite automata, p-equivalence can be checked in polynomial time.

Let us describe how the procedure proposed in [3] checks the AA-diagnosability. First it builds the *testing* sub-pLTS used in the (erroneous) algorithm for A-diagnosability. It then enriches the label of every transition by a second probability defined as follows. Assuming there is a transition labelled by a from (q_1, q_2) to (q_1', q_2'), the second probability is $p_1 p_2$ where:

1. the sum of the probabilities of signalling runs from q_2 to q_2' labelled by a is equal to p_2;
2. the sum of the probabilities of signalling runs from q_1 to q_1' and labelled by a, conditionned over all signalling runs from q_1 and labelled by a, is equal to p_1.

Then one looks at the BSCCs such that for all states the sum of outgoing probabilities are equal to 1 whatever the component of the pair of probabilities is chosen (such BSCC are called *bi-closed*) and such that the first component of the state is faulty. For every such BSCC $\mathcal{C}$ of the enriched testing sub-pLTS one computes the pLTS $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) over the states of $\mathcal{C}$ obtained by considering the first (resp. the second) probability of the label. Finally the procedure checks whether $\mathcal{A}_1$ and $\mathcal{A}_2$, with initial distribution equal to their stationnary distribution, are p-equivalent. If at least one such BSCC yields p-equivalence then the pLTS is not AA-diagnosable.

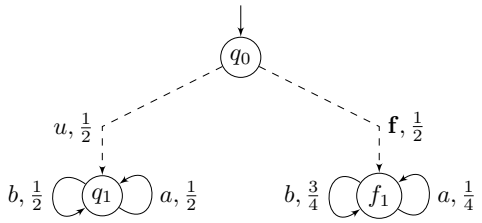

Figure 29 A pLTS which is not diagnosable w.r.t. Definition 59 but diagnosable w.r.t. Definition 62.

Let us look at the pLTS of Figure 29. It is an example taken from [3] where it is claimed to be AA-diagnosable. Figure 30 represents the graph that the algorithm associates with the pLTS of Figure 29. The only bi-closed BSCC is (f_1, q_1) and it is not p-equivalent. So the check performed by the testing sub-pLTS returns that the pLTS is AA-diagnosable.

Unfortunately, this claim is false.

▶ **Proposition 61.** *The pLTS of Figure 29 is not AA-diagnosable according to Definition 59.*

Proof. Assume that $\mathcal{A}$, the pLTS of Figure 29, is AA-diagnosable. Then let $\varepsilon > 0, 1 > \tau > 0$, there exists $n_0 \in \mathbb{N}$ such that for every faulty run ρ and all $n \geq n_0$ we have:

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) > \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho'\})} < \tau.$$

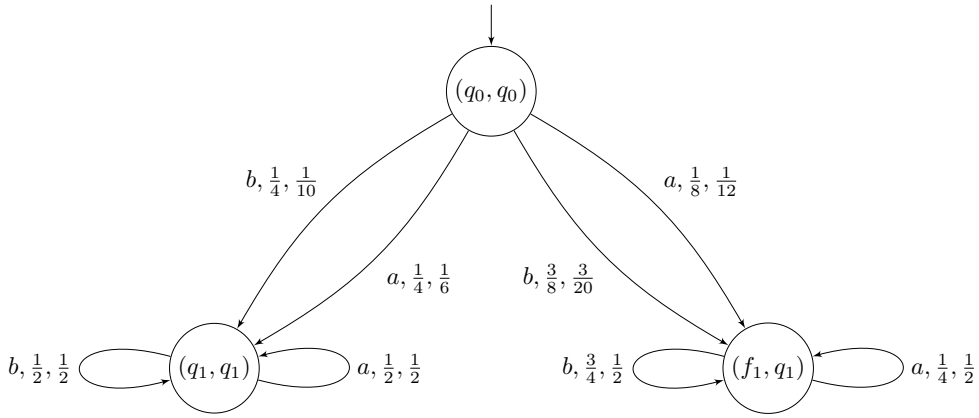

Figure 30 Construction of Chen and Kumar for the pLTS of Figure 29

Let ρ be the faulty run $q_0\mathbf{f}(f_1a)^{n_1}f_1$ with $n_1 > \log_2(\frac{\varepsilon}{1-\varepsilon}) + n_0(\log_2(3) - 1)$.
Then for $\rho' \in \mathsf{SR}_{n_0+|\rho|_o}$ such that $\rho \preceq \rho'$ one gets:

$$\mathbb{P}_{amb}(\mathcal{P}(\rho')) \geq \frac{\frac{1}{2^{n_0+n_1}}}{\frac{1}{2^{n_0+n_1}} + \frac{1}{4^{n_1}}(\frac{3}{4})^{n_0}} = \frac{1}{1 + \frac{3^{n_0}}{2^{n_0+n_1}}} > \varepsilon.$$

Thus, $\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} | \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) > \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} | \rho \preceq \rho'\})} = 1 < \tau$ which contradicts $\tau < 1$.
Therefore $\mathcal{A}$ is not AA-diagnosable. ◀

As seen in the proof of the previous proposition, the requirement that the integer $n_{\varepsilon,\tau}$ does not depend on the faulty run could explain the problem of this algorithm. So let us introduce an alternative definition of AA-diagnosability.

▶ **Definition 62** (AA-diagnosability revisited)**.** A pLTS $\mathcal{A}$ is AA-diagnosable if and only if for all $\varepsilon > 0$, for every faulty run ρ we have:

$$\lim_{n\to\infty} \frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) > \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho'\})} = 0$$

$$\text{where } \mathbb{P}_{amb}(\sigma) = \frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_{|\sigma|} \mid \mathcal{P}(\rho') = \sigma \wedge \rho' \in \mathsf{C}_n\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_{|\sigma|} \mid \mathcal{P}(\rho') = \sigma\})}$$

▶ **Proposition 63.** *The pLTS of Figure 29 is AA-diagnosable according to Definition 62.*

Proof. Let $\varepsilon > 0, \tau > 0$.
Pick $\lambda > 0$ with $\lambda < \frac{3}{4} - \ln(2)$.
Due to the choice of λ, there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0, \frac{1}{1+\frac{3^{n(\frac{3}{4}-\lambda)}}{2^n}} < \varepsilon$.
Consider an observed sequence σ with $|\sigma| = n \geq n_0$ such that $|\frac{|\sigma|_a}{n} - \frac{1}{4}| < \lambda$. There is a single correct signalling run and a single faulty signalling run corresponding σ. Thus:

$$\mathbb{P}_{amb}(\sigma) = \frac{\frac{1}{2^n}}{\frac{1}{2^n} + \frac{3^{n-|\sigma|_a}}{4^n}} \leq \frac{1}{1 + \frac{3^{n(\frac{3}{4}-\lambda)}}{2^n}} < \varepsilon$$

Let $\rho = q_0 \mathbf{f} f_1 x_1 f_1 \dots x_k f_1$ be an arbitrary faulty run where $x_i \in \{a, b\}$.
According to the law of large numbers, as the expected value for a to occur in f_1 is $\frac{1}{4}$:

$$\exists n_1 \in \mathbb{N}, \forall n \geq n_1, \frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \left|\frac{|\rho'|_a}{n} - \frac{1}{4}\right| \geq \lambda\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})} < \tau$$

(observe that this integer n_1 depends on ρ).
Let $m = \max(n_1, n_0)$. Combining the previous results for all $n \geq m$,

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) \geq \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})} < \tau$$

Thus the pLTS is AA-diagnosable. ◀

Unfortunately even with Definition 62, the algorithm of Chan and Kumar is erroneous. Let us look at the pLTS of Figure 31. In the corresponding construction depicted in Figure 32, there is a single bi-closed SCC to consider: $\{(f_1, q_1), (f_2, q_2)\}$. Whatever the component, the stationary distribution of this SCC is equidistributed. So one may exchange (f_1, q_1) an (f_2, q_2) in one of the components and then one observes that the two pLTS are identical and so p-equivalent. So the algorithm of Chan and Kumar returns that the original pLTS is not AA-diagnosable. The next proposition establishes that this is not the case with Definition 62.

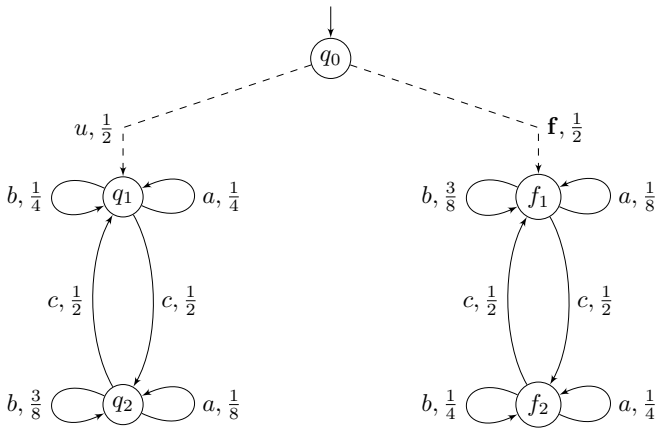

Figure 31 Another pLTS which is AA-diagnosable w.r.t. Definition 62.

▶ **Proposition 64.** *The pLTS of Figure 29 is AA-diagnosable according to Definition 62.*

Proof. Intuitively, the pLTS of Figure 31 is AA-diagnosable since in a faulty run depending on the parity of the number of c observed, the probability that b occurs is greater than the one of a or their occurrence probability are equal.
Let $\varepsilon > 0, \tau > 0$. Pick $\lambda > 0$ with $\lambda < \frac{3}{4} - \ln(2)$.
Due to the choice of λ, there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0, \frac{1}{1+\frac{3^{n(\frac{3}{4}-\lambda)}}{2^n}} < \varepsilon$.

We inductively define for $x \in \{a, b\}$, the function $even_x$ from Σ_o^* to $\mathbb{N}$ by:

- $even_x(\varepsilon) = 0$;
- If $|\sigma|_c$ is even then $even_x(\sigma x) = even_x(\sigma) + 1$ and $even_x(\sigma y) = even_x(\sigma)$ for $y \neq x$;

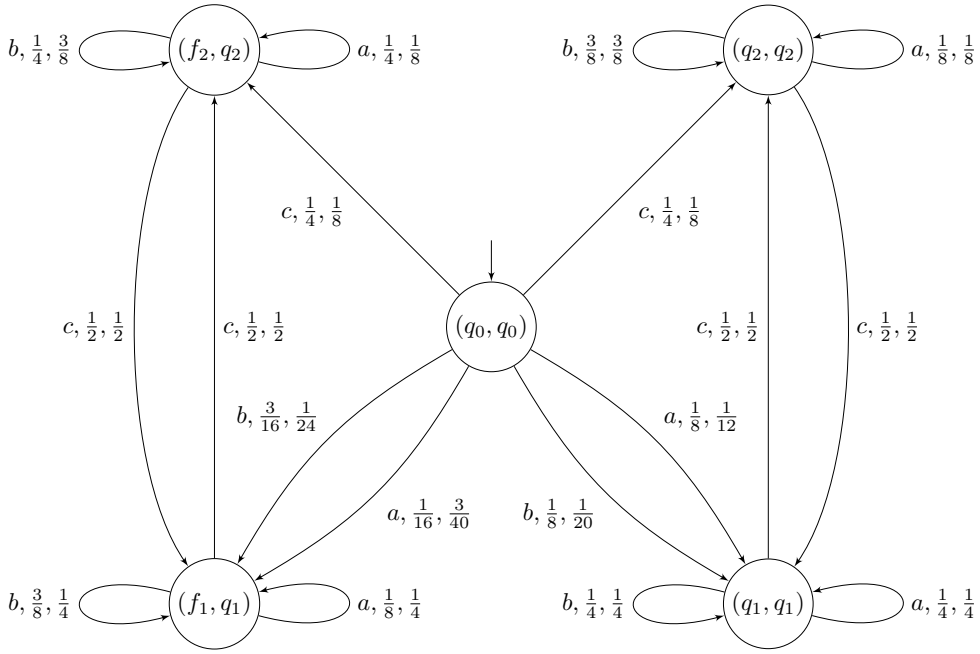

Figure 32 Construction of Chen and Kumar for the pLTS of Figure 31

- If $|\sigma|_c$ is odd then $even_x(\sigma y) = even_x(\sigma)$ for all y.

Consider an observed sequence σ with $even_a(\sigma)+even_b(\sigma) = n \geq n_0$ such that $|\frac{even_a(\sigma)}{n} - \frac{1}{4}| < \lambda$. There is a single correct signalling run and a single faulty signalling run corresponding σ. Thus:

$$\mathbb{P}_{amb}(\sigma) = \frac{\frac{1}{2^n}}{\frac{1}{2^n} + \frac{3^{n-even_a(\sigma)}}{4^n}} \leq \frac{1}{1 + \frac{3^{n(\frac{3}{4}-\lambda)}}{2^n}} < \varepsilon$$

Let ρ be an arbitrary faulty run. Consider a random variable S_n with binomial distribution $B(n, \frac{1}{4})$. According to the law of large numbers:

$$\exists n_1 \in \mathbb{N}, \forall n \geq n_1, \mathbb{P}(\left|\frac{S_n + even_a(\rho)}{n + even_a(\rho) + even_b(\rho)} - \frac{1}{4}\right| \geq \lambda) < \frac{\tau}{2}$$

In the pLTS, $\mathbb{P}(\{\rho' \in \Omega \mid even_a(\mathcal{P}(\rho')) + even_b(\mathcal{P}(\rho')) = \infty\}) = 1$. Thus there exists n_2 such that for $n \geq n_2$:

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge even_a(\mathcal{P}(\rho')) + even_b(\mathcal{P}(\rho')) < \max(n_0, n_1)\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})} < \frac{\tau}{2}$$

Combining the previous results for all $n \geq n_2$,

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) \geq \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})} \leq$$

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge even_a(\mathcal{P}(\rho')) + even_b(\mathcal{P}(\rho')) < \max(n_0, n_1)\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})} +$$

$$\frac{\sum_{x\geq \max(n_0,n_1)} \mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge even_a(\mathcal{P}(\rho')) + even_b(\mathcal{P}(\rho')) = x \wedge \mathbb{P}_{amb}(\mathcal{P}(\rho')) \geq \varepsilon\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})}$$
$$< \frac{\tau}{2} + \frac{\sum_{x\geq \max(n_0,n_1)} \frac{\tau}{2}\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge even_a(\mathcal{P}(\rho')) + even_b(\mathcal{P}(\rho')) = x\})}{\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho'\})}$$
$$< \tau$$

Thus the pLTS is AA-diagnosable.

◀