2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Abstraction-Based Synthesis of Controllers for Approximate Opacity

Junyao Hou, Siyuan Liu, Xiang Yin and Majid Zamani

Abstract— **Opacity is an important information-flow security property which characterizes the plausible deniability of certain "secret behaviors" in dynamical systems. In this paper, we study the problem of synthesizing controllers enforcing a notion of opacity over discrete-time control systems with continuous state sets. In this paper, we develop an *abstraction-based* approach to tackle the controller synthesis problem. Specifically, we adopt a notion of approximate opacity which is suitable for continuous-space control systems. We propose a notion of *approximate initial-state opacity preserving alternating simulation relation* which characterizes the closeness between two systems in terms of opacity preservation. We show that, based on this new notion of system relation, one can synthesize an opacity-enforcing controller for the abstract system which is finite and then refine it back to enforce opacity over the original control system. Finally, we present a method for constructing opacity-preserving finite abstractions for discrete-time control systems under some stability properties. Our results are illustrated on a two-room temperature control problem.**

I. INTRODUCTION

With the advancements of cyber-physical systems (CPS) such as smart manufacturing, and smart cities, information security and privacy issues are becoming increasingly important for design considerations due to large information exchanges in real-time. For dynamical systems, an important aspect of security is to analyze what crucial information can be released through its *information flow*. In this work, we consider an important class of information-flow security properties called *opacity* [6]. Roughly speaking, opacity captures the system's plausible deniability of its "secret" such that its secret and non-secret behaviors should be indistinguishable for an intruder (passive eavesdropper).

In the last decades, a wide range of results on the analysis of opacity have been developed in the context of discrete event systems (DES). Depending on the secret requirements and the information structure of the system, different notions of opacity were proposed in the literature [15], [18], [20]. Among the various notions, initial-state opacity requires that the intruder can never determine for sure that the system was initiated from a secret state. When the original system is not opaque, different approaches have also been proposed to enforce opacity. Among them, one known approach is to use the supervisory control theory, where a controller is used to restrict the behavior of the system such that the closed-loop system under control is opaque [2], [19].

This work was supported by the National Natural Science Foundation of China (62061136004, 62173226, 61833012), the German Research Foundation under Grant ZA 873/7-1, the Shanghai Jiao Tong University Scientific and Technological Innovation Funds, and the National science Foundation under Grant ECCS-2015403.

J. Hou, X. Yin are with Department of Automation and Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: {houjunyao,yinxiang}@sjtu.edu.cn.

S. Liu is with Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: siyliu@kth.se.

M. Zamani is with Department of Computer Science, University of Colorado Boulder, CO 80309, USA, and Department of Computer Science, LMU Munich, 80538 Munich, Germany. E-mail: majid.zamani@colorado.edu

The aforementioned results on the verification or synthesis of opacity mainly deal with DES with discrete-state sets and event-triggered dynamics. However, many real-world systems are hybrid involving both continuous state sets and time-driven dynamics. More recently, notions of opacity have been further extended from DES to general CPS with continuous state sets; see, e.g., [8], [12], [13]. Particularly, in a recent result [21], notions of *approximate opacity* have been proposed, which generalize the opacity concepts from DES to metric systems. Compared with notions of opacity in DES literature, approximate opacity takes into account the imprecise measurements which are typical in real-world applications, and thus are more suitable for CPS with continuous state sets.

Related work. Since the state sets for continuous systems are uncountable, the verification or synthesis are *undecidable* in general. To address this issue, a promising approach is to use *abstraction-based techniques* [16]. In this context, one needs to build a *finite abstraction (a.k.a. symbolic model)* of the original concrete system such that these two systems have certain *relations* under which the analysis or synthesis results over the finite systems can be refined and carried over to the original ones. Abstraction-based techniques have been developed only recently to tackle security properties including the results in [4], [5], [10], [21], [22]. For the purpose of verifying approximate opacity for general control systems, opacity-preserving simulation relations together with the corresponding abstraction algorithms are first developed in [7], [10], [21], [22]. However, all these works are only dealing with verification rather than controller synthesis of opacity. In the context of synthesizing opacity-enforcing controllers, the results in [5] provide a notion of opacity-preserving alternating simulation relations that allows controller refinement with respect to opacity. However, the results in [5] are only applicable to systems with finite state sets and under the assumption of precise observations, which are not appropriate for general CPS.

Our contribution. In this work, we propose a novel approach for synthesizing controllers that enforce approximate initial-state opacity for CPS with continuous state sets. To this end, we first propose a new system relation called *approximate initial-state opacity-preserving (AInit-SOP) alternating simulation relation*. We show that this new

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

system relation preserves approximate initial-state opacity between the abstract and the concrete system in terms of controller synthesis. In particular, one can synthesize opacity-enforcing controllers directly by applying existing synthesis algorithms to the finite abstractions that simulate the concrete systems via AInitSOP alternating simulation relations. We further propose an effective approach to construct finite abstractions which preserve the proposed system relation for a class of discrete-time control systems under some stability assumptions. To the best of our knowledge, this paper is the first to provide directly a controller synthesis approach to enforce opacity for continuous-space control systems using abstraction-based techniques.

II. Preliminaries

A. Notation

Given a vector $x \in \mathbb{R}^n$, we denote by $\|x\|$ the infinity norm of x. A set $B \subseteq \mathbb{R}^m$ is called a *box* if $B = \prod_{i=1}^{m}[c_i, d_i]$, where $c_i, d_i \in \mathbb{R}$ with $c_i < d_i$ for each $i \in \{1, \ldots, m\}$. For any set $A = \bigcup_{j=1}^{M} A_j$ of the form of finite union of boxes, where $A_j = \prod_{i=1}^{n}[c_i^j, d_i^j]$, we define $span(A) = \min\{span(A_j) \mid j = 1, \ldots, M\}$, where $span(A_j) = \min\{|d_i^j - c_i^j| \mid i = 1, \ldots, m\}$. For any $\mu \leq span(A)$, define $[A]_\mu = \bigcup_{j=1}^{M}[A_j]_\mu$, where $[A_j]_\mu = [\mathbb{R}^m]_\mu \cap A_j$ and $[\mathbb{R}^m]_\mu = \{a \in \mathbb{R}^m \mid a_i = k_i\mu, k_i \in \mathbb{Z}, i = 1, \ldots, m\}$. We denote the different classes of comparison functions by $\mathcal{K}$, $\mathcal{K}_\infty$, and $\mathcal{KL}$, where $\mathcal{K} = \{\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} : \gamma$ is continuous, strictly increasing and $\gamma(0) = 0\}$; $\mathcal{K}_\infty = \{\gamma \in \mathcal{K} : \lim_{r\to\infty} \gamma(r) = \infty\}$; $\mathcal{KL} = \{\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} :$ for each fixed s, the map $\beta(r, s)$ belongs to class $\mathcal{K}$ with respect to r and, for each fixed nonzero r, the map $\beta(r, s)$ is decreasing with respect to s and $\beta(r, s) \to 0$ as $s \to \infty\}$.

B. System

In this paper, we employ a notion of "system" introduced in [16] as the underlying model of systems describing both continuous-space and finite control systems, which is modeled by the 6-tuple

$$T = (X, X_0, U, \longrightarrow, Y, H),$$

where X is a (possibly infinite) set of states, $X_0 \subseteq X$ is the set of initial states, U is a (possibly infinite) set of inputs, $\longrightarrow \subseteq X \times U \times X$ is a transition relation, Y is a (possibly infinite) set of outputs, and $H : X \to Y$ is an output map. For simplicity, we also denote a transition $(x, u, x') \in \longrightarrow$ by $x \xrightarrow{u} x'$, where we say that x' is a u-successor, or simply successor, of x. For each state $x \in X$, we denote by $U(x)$ the set of all inputs defined at x, i.e., $U(x) = \{u \in U : \exists x' \in X \text{ s.t. } x \xrightarrow{u} x'\}$, and by $U_u^{post}(x)$ the set of u-successors of state x. A system T is said to be

- *metric*, if the output set Y is equipped with a metric $\mathbf{d} : Y \times Y \to \mathbb{R}_{\geq 0}$;
- *finite* (or *symbolic*), if X and U are finite sets;
- *deterministic*, if for any state $x \in X$ and any input $u \in U$, $|U_u^{post}(x)| \leq 1$ and *nondeterministic* otherwise.

A finite state run is an internal behavior of a system S generated from an initial state $x_0 \in X_0$ under an input sequence $u_1 \cdots u_n$, and is a sequence of transitions $x_0 \xrightarrow{u_1} x_1 \xrightarrow{u_2} \cdots \xrightarrow{u_n} x_n$ such that $x_i \xrightarrow{u_{i+1}} x_{i+1}$ for all $0 \leq i \leq n-1$. The corresponding output run (external behavior) is a sequence of outputs $H(x_0)H(x_1)\cdots H(x_n)$.

Let $T_a = (X_a, X_{a0}, U_a, \underset{a}{\longrightarrow}, Y, H_a)$ and $T_b = (X_b, X_{b0}, U_b, \underset{b}{\longrightarrow}, Y, H_b)$ be two metric systems with the same output set and metric $\mathbf{d}$. Let $\mathcal{I} \subseteq X_a \times X_b \times U_a \times U_b$ be an ε*-approximate interconnection relation* [16] such that $\forall (x_a, x_b) \in \pi_X(\mathcal{I}) : \mathbf{d}(H_a(x_a), H_b(x_b)) \leq \varepsilon$, where $\pi_X(\cdot)$ denotes the projection to $X_a \times X_b$. The composition of T_a and T_b with the interconnection relation $\mathcal{I}$ is a new system

$$T_a \times_{\mathcal{I}}^{\varepsilon} T_b = (X_{ab}, X_{ab0}, U_{ab}, \underset{ab}{\longrightarrow}, Y, H_{ab}),$$

where $X_{ab} = \pi_X(\mathcal{I})$, $X_{ab0} = X_{ab} \cap (X_{a0} \times X_{b0})$, $U_{ab} = U_a \times U_b$, $H_{ab}((x_a, x_b)) = \frac{1}{2}(H_a(x_a) + H_b(x_b))$ and $(x_a, x_b) \xrightarrow[ab]{(u_a, u_b)} (x'_a, x'_b)$ if (i) $x_a \xrightarrow[a]{u_a} x'_a$; (ii) $x_b \xrightarrow[b]{u_b} x'_b$; and (iii) $(x_a, x_b, u_a, u_b) \in \mathcal{I}$. The subscript $\mathcal{I}$ will be dropped when it is clear from the context.

C. Approximate Opacity

In this paper, we consider internal behaviors as the information available to the system, i.e., state information, while external behaviors are considered as the information available to the outside(for example, to an intruder). The information of the system is released by the output mapping $H : X \to Y$. Besides, the system model and its dynamics are also known by the outside intruders.

In many realistic CPS applications, the system might have some "secret" that does not want to be revealed to the outside world via the external behavior. Specifically, we assume that $S \subseteq X$ is a set of *secret states*, and hereafter, we write a system in the form of $T = (X, X_0, S, U, \longrightarrow, Y, H)$ by incorporating the secret state set. The notion of opacity captures the plausible deniability of the system's secret under the information leakage. Note that for metric systems whose outputs are physical signals, due to the imperfect measurement precision of outside observers, it is very difficult to distinguish two observations if their distance is very small. Therefore, in this paper, we adopt a type of opacity called δ-approximate initial-state opacity [21] which quantifies the measurement precision of the intruder, and thus is more applicable to metric systems.

Definition 1: Consider a system $T = (X, X_0, S, U, \longrightarrow, Y, H)$. We say that T is *δ-approximate initial-state opaque* if for any $x_0 \in X_0 \cap S$ and any finite state run $x_0 \xrightarrow{u_1} x_1 \xrightarrow{u_2} \cdots \xrightarrow{u_n} x_n$, there exist $x'_0 \in X_0 \setminus S$ and a finite state run $x'_0 \xrightarrow{u'_1} x'_1 \xrightarrow{u'_2} \cdots \xrightarrow{u'_n} x'_n$ such that

$$\max_{i \in \{0,\ldots,n\}} \mathbf{d}(H(x_i), H(x'_i)) \leq \delta.$$

Intuitively, approximate initial-state opacity requires that an intruder with imperfect measurement can never know that the system was initiated from a secret state. The following example illustrates this notion.

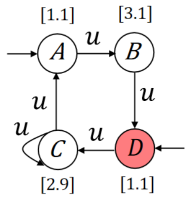


Fig. 1. Example for δ-approximate initial-state opacity on system T_1.

Example 1: Consider system $T_1 = (X_1, X_{1,0}, S_1, U_1, \longrightarrow, Y_1, H_1)$ as shown in Figure 1, where $X_1 = \{A, B, C, D\}$, $X_{1,0} = \{A, D\}$, $S_1 = \{D\}$, $U_1 = \{u\}$, $Y_1 = \{1.1, 2.9, 3.1\} \subseteq \mathbb{R}$ equipped with metric $\mathbf{d}$ defined by $\mathbf{d}(y_1, y_2) = |y_1 - y_2|$, $\forall y_1, y_2 \in Y_1$. We mark all secret states by red, and the output of each state is specified by a value associated to it. First, one can check that T_1 is not 0-approximate/exact initial-state opaque since we know immediately that the system is at secret state when a finite path $D \xrightarrow{u} C$ which generates output path [1.1][2.9] is observed. Next, consider an intruder with measurement precision $\delta = 0.2$. One can observe that T_1 is not 0.2-approximate initial-state opaque due to existence of a self-loop behavior at state C. For example, consider a secret-starting finite path $D \xrightarrow{u} C \xrightarrow{u} C$ which generates output path [1.1][2.9][2.9]. The intruder can infer for sure that the system started from a secret state since there is no path which started from a non-secret state $x_0 \notin S_1$ generating an equivalent output path which is close to [1.1][2.9][2.9] up to precision $\delta = 0.2$. However, once the self-loop is removed from state C, we can readily see that the new system is 0.2-approximate initial-state opaque since for every path starting from a secret state, there always exists a path that starts from a non-secret state with δ-close observations. $\diamond$

III. Abstraction-Based Controller Synthesis

In this section, we discuss how to leverage abstraction-based technique to synthesize controllers that enforce opacity of systems as defined in Subsection II-B.

A. Feedback Composition

When a system T does not satisfy some desired property, e.g., opacity, we can synthesize a controller for T such that the closed-loop system meets the specification. There are several (equivalent) definitions for controllers in the literature. In this paper, we adopt the definition in [16], in which a controller is considered also as a system that is composable to the original one through *approximate alternating simulation relation* defined as follows.

Definition 2: (Approximate Alternating Simulation Relation) Let $T_a = (X_a, X_{a0}, U_a, \underset{a}{\longrightarrow}, Y, H_a)$ and $T_b = (X_b, X_{b0}, U_b, \underset{b}{\longrightarrow}, Y, H_b)$ be two systems with the same output set. A relation $R \subseteq X_a \times X_b$ is said to be an approximate alternating simulation relation from T_a to T_b if the following conditions hold:

1) $\forall x_{a0} \in X_{a0}, \exists x_{b0} \in X_{b0} : (x_{a0}, x_{b0}) \in R$;
2) $\forall (x_a, x_b) \in R : \mathbf{d}(H_a(x_a), H_b(x_b)) \leq \varepsilon$;
3) $\forall (x_a, x_b) \in R, \forall u_a \in U_a(x_a), \exists u_b \in U_b(x_b)$ such that $\forall x_b \xrightarrow{u_b} x'_b, \exists x_a \xrightarrow{u_a} x'_a : (x'_a, x'_b) \in R$.

We say that T_a is ε-approximate alternatingly simulated by T_b (or T_b ε-approximate alternatingly simulates T_a denoted by $T_a \preceq^{\varepsilon}_{AS} T_b$, if there exists an ε-approximate alternating simulation relation from T_a to T_b.

An alternating simulation relation $R \subseteq X_a \times X_b$ from T_a to T_b can also be extended to an interconnection relation $R^e \subseteq X_a \times X_b \times U_a \times U_b$ defined by: $(x_a, x_b, u_a, u_b) \in R^e$ if

(i) $(x_a, x_b) \in R$;
(ii) $u_a \in U_a(x_a), u_b \in U_b(x_b)$; and
(iii) $\forall x_b \xrightarrow{u_b} x'_b, \exists x_a \xrightarrow{u_a} x'_a : (x'_a, x'_b) \in R$.

Intuitively, R^e explicitly specifies which inputs we need to choose to maintain the alternating simulation relation.

The detailed control mechanism of (approximate) alternating simulation relation is explained for finite systems, e.g., [5], [16]. We recall this mechanism succinctly as follows. Consider two systems T_a and T_b under the above defined (approximate) alternating simulation relation, i.e., $T_a \preceq^{\varepsilon}_{AS} T_b$. Then, T_a can be a controller that offers an input $u_a \in U_a(x_a)$; this input is then transferred to T_b as a matching input $u_b \in U_b(x_b)$ via the interconnection relation R^e. Due to the non-determinism, T_b may go to any successor of u_b. Once T_b measures the successor state, T_a will update its state by matching the successor in T_b, and then offer a new input, and this continues. Note that controller T_a can also be non-deterministic as $u_a \in U_a(x_a)$ may not be unique. The above discussion is summarized by the following definition.

Definition 3: (Approximate Feedback Composition) A system T_c is said to be *ε-approximately feedback composable* with a system T_1 if there exists an ε-approximate alternating simulation relation R from T_c to T_1. When T_c is *ε-approximately feedback composable* with T_1, the feedback composition of T_c and T_1 is given by

$$T_c \times^{\varepsilon}_{\mathcal{F}} T_1 = (X_c \times X_1, X_{c0} \times X_{1,0}, U_c \times U_1, \underset{\mathcal{F}}{\longrightarrow}, Y, H_{c1}),$$

where the interconnection relation $\mathcal{F} = R^e$ is an extended ε-approximate alternating simulation relation as in Definition 2. For the sake of simplicity, the subscript $\mathcal{F}$ will be dropped when it is clear from the context.

Based on the above definition, we refer to T_c as a *controller* for system T if it is approximately feedback composable with T.

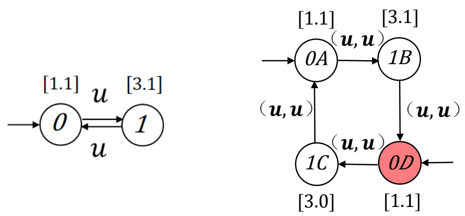
B. Opacity-Enforcing Control Problem

In this paper, in contrast to the existing results on verification of opacity [9], [21], our main goal is to tackle the *opacity-enforcing control problem* which requires to synthesize a controller T_c for system T such that it enforces approximate initial-state opacity on the composed system $T_c \times^{\varepsilon}_{\mathcal{F}} T$. More specifically, we say that T_c *enforces δ-approximate initial-state opacity* over T if for any $(x_{c0}, x_0) \in X_{c0} \times (X_0 \cap S)$ and any sequence

$$(x_{c0}, x_0) \xrightarrow[\mathcal{F}]{(u_{c1}, u_1)} (x_{c1}, x_1) \xrightarrow[\mathcal{F}]{(u_{c2}, u_2)} \cdots \xrightarrow[\mathcal{F}]{(u_{cn}, u_n)} (x_{cn}, x_n),$$

there exist $x'_0 \in X_0 \setminus S$ and a sequence

$$x'_0 \xrightarrow{u'_1} x'_1 \xrightarrow{u'_2} \cdots \xrightarrow{u'_n} x'_n$$



(a) Controller T_c (b) Closed-loop system $T_c \times_{\mathcal{R}_{c1}}^{\varepsilon_{c1}} T_1$

Fig. 2. Example to illustrate the opacity-enforcing control problem.

such that $\max_{i\in\{0,\ldots,n\}} \mathbf{d}(H(x_i), H(x_i')) \le \delta$. Note that in this paper, we assume that $T_c \times_{\mathcal{F}}^{\varepsilon} T$ is non-blocking, i.e., $\forall x \in X : U(x) \neq \emptyset$, which is a conventional assumption in symbolic control. Below, we illustrate the above-mentioned opacity-enforcing control problem on a simple finite system.

Example 2: Let us still consider system T_1 shown in Figure 1. To illustrate the opacity-enforcing control problem on T_1, we assume that there exists a controller $T_c = (X_c, X_{c0}, U_c, \underset{c}{\longrightarrow}, Y_c, H_c)$ shown in Figure 2 (a), where $X_c = \{0, 1\}$, $X_{c0} = \{0\}$, $U_c = \{u\}$, $Y_c = \{1.1, 3.1\} \subseteq \mathbb{R}$ equipped with metric $\mathbf{d}$ defined by $\mathbf{d}(y_1, y_2) = |y_1 - y_2|$, $\forall y_1, y_2 \in Y_c$. The output of each state is specified by a value associated to it. One can readily check that $T_c \preceq_{AS}^{\varepsilon_{c1}} T_1$ with $\varepsilon_{c1} = 0.2$ through the approximate alternating simulation relation $R_{c1} = \{(0, A), (1, B), (1, C), (0, D)\}$. By Definition 3, the closed-loop system $T_c \times_{R_{c1}}^{\varepsilon_{c1}} T_1 = (X_c \times X_1, X_{c0} \times X_{1,0}, U_c \times U_1, \underset{\mathcal{R}_{c1}}{\longrightarrow}, Y, H_{c1})$ is shown in Figure 2(b). One can readily verify that for any path that started from a secret initial state in $(x_{c0}, x_0) \in X_{c0} \times (X_{1,0} \cap S_1)$ in the closed-loop system $T_c \times_{R_{c1}}^{\varepsilon_{c1}} T_1$, there exists an output-equivalent path initiated from a non-secret state $x \in X_{1,0} \setminus S_1$ in system T_1. For example, for $(0, D) \in X_{c0} \times (X_{1,0} \cap S_1)$ and a finite path $(0, D) \xrightarrow[\mathcal{R}_{1c}]{(u,u)} (1, C)$, there exist $A \in X_{1,0} \setminus S_1$ and a finite path $A \xrightarrow{u} B$ such that $|(H_{c1}(0, D) - H_1(A)| = 0 \le 0.2$ and $|H_{c1}(1, C) - H_1(B)| = 0.1 \le 0.2$. Therefore, we can conclude that T_c is a controller that enforces 0.2-approximate initial-state opacity over T_1. ⋄

Note that parameters δ and ε in this paper specify two different types of precision. Parameter δ is used to specify the measurement precision of outside intruder under which we can guarantee the approximate opacity of a single system, while the parameters ε in the definition of approximate alternating simulation relation is used to describe the "distance" between two systems.

Note that the opacity-enforcing control problem is known to be undecidable for continuous-space systems. To this end, a promising approach is to leverage abstraction-based approaches as a bridge for the purpose of controller synthesis [16]. In this context, one first needs to build a finite abstraction of the concrete continuous-space control system, then synthesize a discrete controller based on the finite abstraction, and finally, refine the synthesized discrete controller back as a hybrid one to the original concrete system. The key to abstraction-based approach is to find appropriate relations between concrete systems and their finite abstractions such that properties of interest can be preserved under controller refinement. The abstraction-based controller refinement scheme is formalized in the following subsection.

C. Abstraction and Controller Refinement

Although approximate alternating simulation relations have shown to be useful [16] for controller refinement of properties such as ω-regular properties, unfortunately, they *do not preserve* security properties including opacity [1], [22]; check [22] for some counterexamples. Therefore, we introduce a new notion of *opacity-preserving* approximate alternating simulation relation, so that it can be applied to the abstraction-based opacity-enforcing control problem for continuous-space control systems.

Here, we propose a notion of so-called *approximate initial-state opacity preserving* (AInitSOP) alternating simulation relation. Specifically, this new notion of system relation from T_1 to T_2 is required to satisfy the following requirements: (i) it is still an alternating simulation relation; (ii) enforcing opacity for T_1 implies the enforcement of opacity for T_2 after the controller refinement. The proposed notion of AInitSOP alternating simulation relation is introduced in the following definition.

Definition 4: (Approximate Initial-State Opacity Preserving Alternating Simulation Relation) Let T_1, T_2 be two systems, where $T_i = (X_i, X_{i,0}, S_i, U_i, \underset{i}{\longrightarrow}, Y, H_i), i = 1, 2$. A relation $R \subseteq X_1 \times X_2$ is said to be an ε-approximate initial-state opacity preserving (AInitSOP) alternating simulation relation from T_1 to T_2 if

1) a) $\forall x_{1,0} \in X_{1,0}, \exists x_{2,0} \in X_{2,0} : (x_{1,0}, x_{2,0}) \in R$;
   b) $\forall x_{1,0} \in X_{1,0} \setminus S_1, \exists x_{2,0} \in X_{2,0} \setminus S_2 : (x_{1,0}, x_{2,0}) \in R$;
2) $\forall (x_1, x_2) \in R : \mathbf{d}(H_1(x_1), H_2(x_2)) \le \varepsilon$
3) $\forall (x_1, x_2) \in R$, we have
   a) $\forall u_1 \in U_1(x_1), \exists u_2 \in U_2(x_2), \forall x_2 \xrightarrow{u_2} x_2', \exists x_1 \xrightarrow{u_1} x_1'$ such that $(x_1', x_2') \in R$;
   b) $\forall x_1 \xrightarrow{u_1} x_1', \exists x_2 \xrightarrow{u_2} x_2'$ such that $(x_1', x_2') \in R$.

We say that T_1 is AInitSOP alternatingly simulated by T_2 (or T_2 AInitSOP alternatingly simulates T_1), denoted by $T_1 \preceq_{AIAS}^{\varepsilon} T_2$, if there exists an AInitSOP alternating simulation relation from T_1 to T_2.

If $T_1 \preceq_{AIAS}^{\varepsilon} T_2$, we say that T_1 is an *abstraction* of T_2. In the sequel, we denote the original system by T_2 and the abstract system by T_1.

Note that an AInitSOP alternating simulation relation is still an alternating simulation relation, which makes the controller refinement procedure still possible. Next, we present the first main result of our paper which shows how to use the above-defined AInitSOP alternating simulation relation for the purpose of opacity-enforcing controller synthesis.

Theorem 1: Consider two systems T_1 and T_2, where $T_i = (X_i, X_{i,0}, S_i, U_i, \underset{i}{\longrightarrow}, Y, H_i), i = 1, 2$, and suppose that $T_1 \preceq_{AIAS}^{\varepsilon_{12}} T_2$. Then for any controller T_c that enforces δ-approximate initial-state opacity for the abstract system T_1 with $T_c \preceq_{AS}^{\varepsilon_{c1}} T_1$, the refined controller $T_{ref} = T_c \times_{\mathcal{F}_{c1}}^{\varepsilon_{c1}} T_1$ also enforces $\mathbf{max}\{(\frac{1}{2}\varepsilon_{c1} + \frac{3}{2}\varepsilon_{12} + \delta), (\varepsilon_{c1} + \frac{3}{2}\varepsilon_{12})\}$-approximate initial-state opacity for the original system T_2.

In essence, the role of AInitSOP alternating simulation relation is to build a "bridge" between the original system and the controller of the abstract system. Based on this theorem, one can design a controller that enforces opacity of the finite abstract system, and then refine the controller back to enforce opacity over the original control system.

Note that in symbolic control, the controllers synthesized for abstract systems (with finite state set) are often precise, i.e., $\varepsilon_{c1} = 0$. In this case, we get a more succinct result as presented in the following corollary.

Corollary 1: Consider two systems T_1 and T_2, where $T_i = (X_i, X_{i,0}, S_i, U_i, \underset{i}{\longrightarrow}, Y, H_i), i = 1, 2$, and suppose that $T_1 \preceq^{\varepsilon_{12}}_{AIAS} T_2$. Then for any controller T_c that enforces δ-approximate initial-state opacity for the abstract system T_1 where $T_c \preceq^{0}_{AS} T_1$, the refined controller $T_{ref} := T_c \times^{\varepsilon_{c1}}_{\mathcal{F}_{c1}} T_1$ enforces $(\frac{3}{2}\varepsilon_{12} + \delta)$-approximate initial-state opacity over the original system T_2.

IV. Approximate Opacity-Preserving Finite Abstractions

In the previous section, we introduced a notion of approximate initial-state opacity preserving alternating simulation relations. Naturally, the next question is how to construct an opacity-preserving finite abstraction for a concrete control system so that it can be used for the sake of opacity-enforcing controller synthesis. In general, the approach to construct finite abstractions is system-dependent, and not all systems admit finite abstractions. Next, we show that a class of discrete-time control systems admits opacity-preserving finite abstractions under certain stability assumptions.

A. Discrete-time Control Systems

In this section, we consider a class of discrete-time control systems of the following form.

Definition 5: A discrete-time control system Σ is defined by the tuple $\Sigma = (\mathbb{X}, \mathbb{S}, \mathbb{U}, f, \mathbb{Y}, h)$, where $\mathbb{X}$, $\mathbb{U}$, and $\mathbb{Y}$ are the state, input, and output sets, respectively, and are subsets of normed vector spaces with appropriate dimensions. Set $\mathbb{S} \subseteq \mathbb{X}$ is a set of secret states. The map $f : \mathbb{X} \times \mathbb{U} \to \mathbb{X}$ is called the transition function, and $h : \mathbb{X} \to \mathbb{Y}$ is the output map and assumed to satisfy the following Lipschitz condition: $\|h(x) - h(x')\| \leq \alpha(\|x - x'\|)$ for some $\alpha \in \mathcal{K}_\infty$ and all $x, x' \in \mathbb{X}$. The discrete-time control system Σ is described by difference equations of the form

$$\Sigma : \begin{cases} \xi(k+1) = f(\xi(k), \upsilon(k)), \\ \zeta(k) = h(\xi(k)), \end{cases} \tag{1}$$

where $\xi : \mathbb{N} \to \mathbb{X}$, $\zeta : \mathbb{N} \to \mathbb{Y}$, and $\upsilon : \mathbb{N} \to \mathbb{U}$ are the state, output, and input signals, respectively.

We denote by $\xi_{x\upsilon}(k)$ the point reached at time k under the input signal υ from initial condition $x = \xi_{x\upsilon}(0)$. Similarly, let $\zeta_{x\upsilon}(k)$ denote the output corresponding to state $\xi_{x\upsilon}(k)$, i.e. $\zeta_{x\upsilon}(k) = h(\xi_{x\upsilon}(k))$. Note that we implicitly assumed that $\mathbb{X}$ is positively invariant.

B. Construction of Finite Abstractions

Next, we present how to construct finite abstractions for a class of discrete-time control systems. Specifically, the finite abstraction is built under the assumption that the concrete discrete-time control system is *incrementally input-to-state stable* as defined in [17] and recalled below.

Definition 6: System $\Sigma = (\mathbb{X}, \mathbb{S}, \mathbb{U}, f, \mathbb{Y}, h)$ is called incrementally input-to-state stable (δ-ISS) if there exist functions $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}_\infty$ such that $\forall x, x' \in \mathbb{X}$ and $\forall \upsilon, \upsilon' \in \mathbb{N} \to \mathbb{U}$, the following holds for any $k \in \mathbb{N}$:

$$\|\xi_{x\upsilon}(k) - \xi_{x'\upsilon'}(k)\| \leq \beta(\|x - x'\|, k) + \gamma(\|\upsilon - \upsilon'\|). \tag{2}$$

Next, in order to construct approximate initial-state opacity preserving finite abstractions for a control system $\Sigma = (\mathbb{X}, \mathbb{S}, \mathbb{U}, f, \mathbb{Y}, h)$ in Definition 5, we define an associated metric system $T(\Sigma) = (X, X_0, X_S, U, \longrightarrow, Y, H)$, where $X = \mathbb{X}$, $X_0 = \mathbb{X}$, $X_S = \mathbb{S}$, $U = \mathbb{U}$, $Y = \mathbb{Y}$, $H = h$, and $x \overset{u}{\longrightarrow} x'$ if and only if $x' = f(x, u)$. In the sequel, we will use $T(\Sigma)$ to denote the concrete control systems interchangeably.

Next, we introduce a symbolic system for the control system $\Sigma = (\mathbb{X}, \mathbb{S}, \mathbb{U}, f, \mathbb{Y}, h)$. To do so, in the rest of the paper, we assume that sets $\mathbb{X}$, $\mathbb{S}$ and $\mathbb{U}$ are of the form of finite union of boxes. Consider a concrete control system Σ and a tuple $\mathsf{q} = (\eta, \mu)$ of parameters, where $0 < \eta \leq \min\{\mathit{span}(\mathbb{S}), \mathit{span}(\mathbb{X} \setminus \mathbb{S})\}$ is the state set quantization, and $0 < \mu \leq \mathit{span}(\mathbb{U})$ is the input set quantization. Now let us introduce the symbolic system

$$T_{\mathsf{q}}(\Sigma) = (X_{\mathsf{q}}, X_{\mathsf{q}0}, X_{\mathsf{q}S}, U_{\mathsf{q}}, \underset{\mathsf{q}}{\longrightarrow}, Y_{\mathsf{q}}, H_{\mathsf{q}}), \tag{3}$$

where $X_{\mathsf{q}} = X_{\mathsf{q}0} = [\mathbb{X}]_\eta$, $X_{\mathsf{q}S} = [\mathbb{S}]_\eta$, $U_{\mathsf{q}} = [\mathbb{U}]_\mu$, $Y_{\mathsf{q}} = \{h(x_{\mathsf{q}}) \mid x_{\mathsf{q}} \in X_{\mathsf{q}}\}$, $H_{\mathsf{q}}(x_{\mathsf{q}}) = h(x_{\mathsf{q}})$, $\forall x_{\mathsf{q}} \in X_{\mathsf{q}}$, and

- $x_{\mathsf{q}} \underset{\mathsf{q}}{\overset{u_{\mathsf{q}}}{\longrightarrow}} x'_{\mathsf{q}}$ if and only if $\|x'_{\mathsf{q}} - f(x_{\mathsf{q}}, u_{\mathsf{q}})\| \leq \frac{1}{2}\eta$.

Now, we are ready to present the main result of this section, which shows that under some condition over the quantization parameters η and μ, the finite abstraction $T_{\mathsf{q}}(\Sigma)$ constructed in (3) indeed simulates our concrete control system $T(\Sigma)$ through the proposed relation.

Theorem 2: Let $\Sigma = (\mathbb{X}, \mathbb{S}, \mathbb{U}, f, \mathbb{Y}, h)$ be a δ-ISS control system. For any desired precision $\varepsilon > 0$, and any tuple $\mathsf{q} = (\eta, \mu)$ of parameters satisfying

$$\beta\left(\alpha^{-1}(\varepsilon), 1\right) + \frac{1}{2}\eta \leq \alpha^{-1}(\varepsilon), \tag{4}$$

we have $T_{\mathsf{q}}(\Sigma) \preceq^{\varepsilon}_{AIAS} T(\Sigma)$.

Intuitively, this theorem shows that under certain conditions over the quantization parameter η, one can construct a finite abstraction as in (3) which is related to the original control system through the proposed AInitSOP alternating simulation relation. Let us recall that such an abstraction is a crucial bridge to the opacity-enforcing controller synthesis of continuous-space control systems. To be specific, one can first design symbolic controllers for the finite abstractions, and then leverage the results proposed in Theorem 1 to refine controllers to hybrid ones that render opacity over the original systems. Note that the design of symbolic controllers for finite abstractions is out of the scope of this paper. However, since the abstractions are finite, one can readily utilize the existing works and computational tools in the DES literature (e.g., [3]) to design controllers that enforce opacity over the abstractions.

We should mention that one can always find quantization parameters η such that (4) holds as long as $\beta\left(\alpha^{-1}(\varepsilon), 1\right) \leq \alpha^{-1}(\varepsilon)$. This inequality can be ensured by regarding the discrete-time control system as a sampled-data version of an original continuous-time system with large-enough sampling time (see [21, Remark VI.8]).

V. Case Study

In this section, we demonstrate the proposed abstraction-based controller synthesis approach on a two-room temperature control problem, where each room is equipped with a heater. This model is borrowed from [11]. The temperature evolution of two rooms is:

$$\Sigma : \begin{cases} \boldsymbol{x}(k+1) = A\boldsymbol{x}(k) + \alpha_h x_h \boldsymbol{u}(k) + \alpha_e \boldsymbol{x}_e, \\ \boldsymbol{y}(k) = h(\boldsymbol{x}(k)), \end{cases} \tag{5}$$

where $\boldsymbol{x}(k)$=[$\boldsymbol{x}_1(k)$; $\boldsymbol{x}_2(k)$], where $\boldsymbol{x}_i(k)$, $i \in \{1; 2\}$, represents the temperature of each room at time k, $\boldsymbol{u}(k)$=[$\boldsymbol{u}_1(k)$; $\boldsymbol{u}_2(k)$], where $\boldsymbol{u}_i(k) \in [0, 1]$, $\forall i \in [1; 2]$, represents the ratio of the heater valve being open in room i, $A \in \mathbb{R}^{2\times 2}$ is a heat exchange matrix for this model with elements $\{A\}_{11} = \{A\}_{22} = \alpha$, $\{A\}_{12} = 1 - 2\alpha - \alpha_e - \alpha_h c_1$, $\{A\}_{21} = 1 - 2\alpha - \alpha_e - \alpha_h c_2$. The parameters $\alpha = 0.1$, $\alpha_h = 0.5$, $\alpha_e = 0.1$, $c_1 = 0.4$ and $c_2 = 3$ are heat exchange coefficients of this model, $\boldsymbol{x}_e = [x_{e1}; x_{e2}] = [5\,^{\circ}C; 5\,^{\circ}C]$ represents the environment temperature and $x_h = 50\,^{\circ}C$ represents the heater temperature. The output of this system is assumed to be the temperature of the second room, i.e., $\boldsymbol{y}(k) = h(\boldsymbol{x}(k)) = \boldsymbol{x}_2(k)$. In this example, our region of interest is as follows: $X = [0, 40] \times [0, 40]$, $X_0 = [20, 25] \times [20]$, $X_s = [23.5, 25] \times [20]$.

It is assumed that the secret of the system is whether the initial temperature of room 1 is higher than $23.5\,^{\circ}C$, as this could mean that there are sensitive devices running or people are gathering in this room. We also assume that there is a malicious intruder interested in reasoning about the initial temperature of the first room by knowing the dynamics of the system and the output of the model. It is worth mentioning that due to the imperfect precision, the intruder cannot accurately obtain the output values of the system. Correspondingly, the measurement precision of the intruder is assumed to be $\delta_1 = 3.5$. Note that this can be captured as an δ_1-approximate initial-state opacity property of the system. By the verification approach of [9], this system is not 3.5-approximate initial-state opaque. Now, we apply our proposed abstraction-based framework to synthesize a controller to enforce approximate initial-state opacity on Σ. To do this, let us first build a finite abstraction of Σ using the approach presented in Subsection IV-B with a desired precision $\varepsilon = 1$. One can readily check that the system Σ is incrementally input-to-state stable. Hence, by leveraging Theorem 2 and based on inequality (4), we simply choose

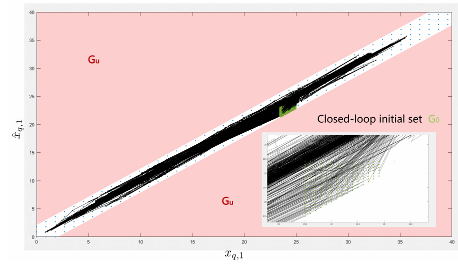


Fig. 3. Trajectories of the augmented closed-loop abstract system projected on the first-room coordinate starting from initial region G_0 (represented by the green area) under symbolic control. The black lines denote the state trajectories of the augmented abstract system. The red regions constitute the unsafe set G_u.

the state quantization parameter to be $\eta = 0.9$ and the input quantization parameter $\mu = 0.5$. Then, following the approach presented in Subsection IV-B, one can obtain a finite abstraction $T_{\mathsf{q}}(\Sigma)$ such that $T_{\mathsf{q}}(\Sigma) \preceq^{\varepsilon}_{AIAS} T(\Sigma)$ holds.

Next, we proceed with the opacity-enforcing controller synthesis by leveraging the result in Corollary 1. Specifically, in order to enforce the original system Σ to be 3.5-approximate initial-state opaque, we can design a 2.0-approximate initial-state opacity-enforcing controller for the abstract system T_{q}, and then refine it back to a controller that enforces 3.5-approximate initial-state opacity on the original system Σ.

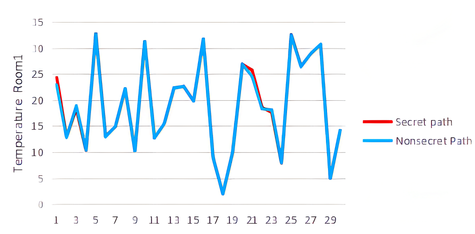


Fig. 4. First-room temperature trajectories initiated from different initial states (one from a secret state $x = [24.3; 20]$ and the other one from a non-secret state $x' = [22.9; 20]$.

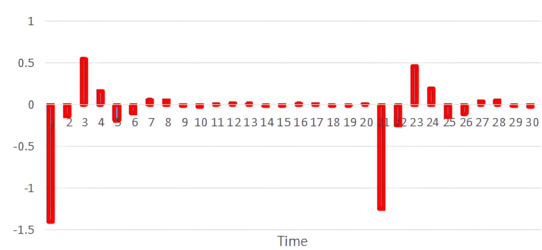


Fig. 5. Distance between the output trajectories corresponding to the two state trajectories depicted in Figure 4.

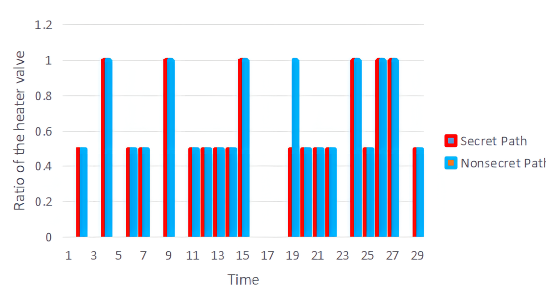


Fig. 6. The input runs corresponding to the state trajectories in Figure 4.

For the sake of completeness of the example, we briefly discuss our symbolic controller design process with the help of SCOTS [14] together with the ideas proposed in [9]. In order to utilize SCOTS to design an opacity-enforcing symbolic controller, we resort to an approach developed in [9] which essentially converts the opacity property of a single control system to a safety property of an augmented system which can be seen as the product of a control system and itself. We refer interested readers to [9] for more details on the translation of opacity property to a safety one. Here, we briefly recall some of the notations that are used in this example: Given a single system T_{q}, an augmented system is defined as $T_{\mathsf{q}} \times T_{\mathsf{q}} = (X_q \times X_q, X_{q0} \times X_{q0}, X_{qS} \times X_{qS}, U_q \times U_q, f_q \times f_q, Y_q \times Y_q, H_q \times H_q)$. We use $G = X_q \times X_q$ to denote the augmented symbolic state set. Recall that a safety property essentially requires that any trajectory starting from a certain initial region should never reach an unsafe region. In this example, the initial and unsafe region for the obtained safety property is as follows: the initial region is $G_0 = \{(x_q, \hat{x}_q) \in (X_{q0} \cap X_{qS}) \times (X_{q0} \setminus X_{qS}) | \; \|H(x_q) - H(\hat{x}_q)\| \leq \delta_2\}$, the unsafe region is $G_u = \{(x_q, \hat{x}_q) \in X_q \times X_q) | \; \|H(x_q) - H(\hat{x}_q)\| > \delta_2\}$, where $\delta_2 = 2$. Then, the safety controller synthesis problem is solved using SCOTS. In Figure 3, we show the state trajectories of the augmented closed-loop abstract system projected on the first-room coordinate under the controller provided by SCOTS. It can be readily seen that the safety property is satisfied on the augmented system, which implies that the individual closed-loop abstract system is 2.0-approximate initial-state opaque.

So far, we have obtained a controller that enforces opacity on the abstract system with the closed-loop system denoted by $T_{ref} = T_c \times^{0}_{\mathcal{F}_{c1}} T_q(\Sigma)$. Then, according to Corollary 1, let T_{ref} be the refined controller for the original system. We have the guarantee that the closed-loop control system $T_{ref} \times^{1}_{\mathcal{F}_{12}} T(\Sigma)$ is δ_1-approximate initial-state opaque, where $\delta_1 = (\frac{3}{2}\varepsilon + \delta_2) = 3.5$, and the AInitSOP alternating simulation relation is as follows: $R_{12} = \{(x_{q,1}, x_1) \in X_{q,1} \times X_1 | \|x_1 - x_{\mathsf{q},1}\| \leq 1.0 \wedge (x_c, x_{q,1}) \in R_{c1}\}$. Figure 4 shows the simulation results of our implementation, which illustrates δ_1-approximate initial-state opacity of the closed-loop control system. In particular, two trajectories are depicted in this figure, where one is initiated from a secret state $[24.3; 20]$ while the other started from a non-secret state $x' = [22.9; 20]$. The distance between the corresponding output trajectories of these two state runs is depicted in Figure 5. The input runs of the trajectories are shown in Figure 6.

VI. Conclusion

In this work, we developed an abstraction-based approach for synthesizing controllers that enforce approximate initial-state opacity over continuous-space control systems. To this end, we proposed a notion of approximate initial-state opacity-preserving alternating simulation relation, which can be used to capture the distance between a concrete control system and its finite abstraction. Under this system relation, an opacity-enforcing controller designed for the finite abstraction can be refined back to the original control system. We further showed that under an incremental input-to-state stability assumption, a finite abstraction can be readily computed for a control system through the proposed system relation. Finally, we used a two-room temperature control example to illustrate our proposed abstraction-based controller synthesis framework.

References

[1] R. Alur, P. Černỳ, and S. Zdancewic. Preserving secrecy under refinement. In *International Colloquium on Automata, Languages, and Programming*, pages 107–118. Springer, 2006.

[2] J. Dubreil, P. Darondeau, and H. Marchand. Supervisory control for opacity. *IEEE Transactions on Automatic Control*, 55(5):1089–1100, 2010.

[3] Y. Falcone and H. Marchand. Enforcement and validation (at runtime) of various notions of opacity. *Discrete Event Dynamic Systems*, 25(4):531–570, 2015.

[4] J. Hou, S. Liu, X. Yin, and M. Zamani. Abstraction-based verification of approximate preopacity for control systems. *IEEE Control Systems Letters*, 7:1087–1092, 2022.

[5] J. Hou, X. Yin, S. Li, and M. Zamani. Abstraction-based synthesis of opacity-enforcing controllers using alternating simulation relations. In *IEEE Conference on Decision and Control*, pages 7653–7658, 2019.

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

[7] S. Liu, A. Swikir, and M. Zamani. Verification of approximate opacity for switched systems: A compositional approach. *Nonlinear Analysis: Hybrid Systems*, 42:101084, 2021.

[8] S. Liu, A. Trivedi, X. Yin, and M. Zamani. Secure-by-construction synthesis of cyber-physical systems. *Annual Reviews in Control*, 53:30–50, 2022.

[9] S. Liu and M. Zamani. Verification of approximate opacity via barrier certificates. *IEEE Control Systems Letters*, 5(4):1369–1374, 2020.

[10] S. Liu and M. Zamani. Compositional synthesis of opacity-preserving finite abstractions for interconnected systems. *Automatica*, 131:109745, 2021.

[11] P.-J. Meyer, A. Girard, and E. Witrant. Compositional abstraction and safety synthesis using overlapping symbolic models. *IEEE Transactions on Automatic Control*, 63(6):1835–1841, 2017.

[12] M. Mizoguchi and T. Ushio. Abstraction-based control under quantized observation with approximate opacity using symbolic control barrier functions. *IEEE Control Systems Letters*, 2021.

[13] B. Ramasubramanian, W.R. Cleaveland, and S. Marcus. Notions of centralized and decentralized opacity in linear systems. *IEEE Transactions on Automatic Control*, 265(4):1442–1455, 2020.

[14] M. Rungger and M. Zamani. Scots: A tool for the synthesis of symbolic controllers. In *HSCC*, pages 99–104, 2016.

[15] A. Saboori and C.N. Hadjicostis. Verification of infinite-step opacity and complexity considerations. *IEEE Transaction Automatic Control*, 57(5):1265–1269, 2012.

[16] P. Tabuada. *Verification and Control of Hybrid Systems: A Symbolic Approach*. Springer Science & Business Media, 2009.

[17] DN Tran. *Advances in stability analysis for nonlinear discrete-time dynamical systems*. PhD thesis, 2018.

[18] A. Wintenberg, M. Blischke, S. Lafortune, and N. Ozay. A general language-based framework for specifying and verifying notions of opacity. *Discrete Event Dynamic Systems*, pages 1–37, 2022.

[19] X. Yin and S. Lafortune. A uniform approach for synthesizing property-enforcing supervisors for partially-observed discrete-event systems. *IEEE Transactions on Automatic Control*, 61:2140–2154, 08 2016.

[20] X. Yin and S. Lafortune. A new approach for the verification of infinite-step and K-step opacity using two-way observers. *Automatica*, 80:162–171, 2017.

[21] X. Yin, M. Zamani, and S. Liu. On approximate opacity of cyber-physical systems. *IEEE Transactions on Automatic Control*, 66(4):1630–1645, 2021.

[22] K. Zhang, X. Yin, and M. Zamani. Opacity of nondeterministic transition systems: A (bi) simulation relation approach. *IEEE Transactions on Automatic Control*, 64(12):5116–5123, 2019.