An SMT-based Approach to Secure State Estimation Under Sensor and Actuator Attacks

Mehrdad Showkatbakhsh, Yasser Shoukry, Robert H. Chen, Suhas Diggavi and Paulo Tabuada

Abstract— **This paper addresses the problem of state estimation of a linear time-invariant system when some of the sensors or/and actuators are under adversarial attack. In our set-up, the adversarial agent attacks a sensor (actuator) by manipulating its measurement (input), and we impose no constraint on how the measurements (inputs) are corrupted. We introduce the notion of "sparse strong observability" to characterize systems for which the state estimation is possible, given bounds on the number of attacked sensors and actuators. Furthermore, we develop a secure state estimator based on Satisfiability Modulo Theory (SMT) solvers.**

I. Introduction

Cyber-Physical Systems (CPSs) are characterized by the tight interconnection of cyber and physical components. Security of CPS is no longer restricted to the cyber domain, and recent attacks such as the StuxNet malware [1] and the security flaws reported on modern cars [2], [3] motivated the recent interest in security of CPS by the control community, (see for example, [4], [5], [6], [7] and references therein).

Several different security problems have been investigated in the literature, *e.g.,* denial-of-service [8], [9], [10], [11], reply attack [12], man-in-the-middle [13], false data injection [14] etc. In this paper, we investigate the problem of state estimation when some of the *sensors and actuators* are under adversarial attacks. In the rest of this paper, we broadly refer to the problem of state estimation under adversarial attacks as "secure state estimation". Our attack model is quite general and the adversary can alter sensors measurements and actuators inputs arbitrarily, we do not impose any restriction on the magnitude, statistical properties and temporal characteristics of the adversarial signals throughout this work.

In [15], the problem of control and estimation under sensor attack is investigated and the authors derived necessary and sufficent conditions under which estimation and stabilization are possible. Shoukry et. al. further refined this property and called it "sparse observability" [16]. Chong et. al. independently derived a similar condition in [17] for continuous-time systems and called it "observability under attack". Nakahira et. al. relaxed the sparse observability condition to sparse detectability [18]. Mishra et. al. investigated the noisy version of this problem and derived the optimal solution for Gaussian noise [19]. In our set-up, the adversarial model not only covers sensor attacks, but also includes actuator attacks. Shoukry et. al. proposed a novel secure state estimator using Satisfiability Modulo Theory (SMT) paradigm, called IMHOTEP-SMT [20]. In this paper we address the more general problem of sensor *and* actuator attacks and build an SMT-based estimator that can resist against both attacks.

In another line of work, problem of secure state estimation has been investigated when the model of the system is not known exactly [21], [22]. Tiwari et. al. proposed a method that does not rely on the model of the underlying system and builds so-called "safety envelopes" as it receives attack-free data [23]. Showkatbakhsh et. al. considered the model identification under sensors attacks [24], [25]. In all of these works, the adversarial power is limited to the sensors and all actuators are supposed to be safe.

Fault tolerant and fault detection filters are closely related to secure state estimation. The classical fault tolerant filters can detect faults on actuators and sensors, however, they are not adequate for the purpose of security. Some of these filters assume a priori knowledge (statistical or temporal) of the fault signals [26], an assumption that does not hold in the secutiry framework. The classical fault detection filters [27] do not guarantee identification of all possible adversarial signals and zero-dynamics attacks remain stealthy. Therefore, the state estimate is not guaranteed to be correct. In contrast, the method proposed in this paper is guaranteed to construct the correct estimate of the state in spite of attacks on sensors and/or actuators. In a recent work [28], Harirchi et. al. proposed a sound and complete fault detection approach using techniques from model invalidation. The authors pursued a worst-case scenario approach and therefore their framework is suitable for security. However, necessary and sufficient conditions for state estimation in a general adversarial setting were not investigated in [28]. In this paper, we precisely characterize the class of systems, by providing necessary and sufficient conditions, for which state reconstruction is possible despite sensor and/or actuator attacks.

In [29], the problem of attack detection and identification is considered. The authors related the "undetectable" and "unidentifiable" attacks to the zero-dynamics of the underlying system. The proposed identification monitor consists of a number of fault detection filters that grows exponentially with the size of the attacks, and therefore hinders scalability. In another work [30], the authors investigated detectibility and identifiability of attacks in the presence of disturbances and the concept of security index is generalized to dynamical systems. The proposed method suffers from the problem of scalability. In this paper, by leveraging the SMT paradigm

M. Showkatbakhsh, S. Diggavi and P. Tabuada are with the UCLA Electrical Engineering Department, Los Angeles , CA { mehrdadsh, suhas, tabuada } @ucla.edu

R. H. Chen is with NG Next, Redondo Beach, CA Robert.Chen@ngc.com

Y. Shoukry is with the UC Berkeley Electrical Engineering and Computer Science Department, Berkeley, CA and the UCLA Electrical Engineering Departments, Los Angeles , CA yshoukry@eecs.berkeley.edu

we design a state estimator that substantially outperforms the brute-force approach.

Our contributions are as follows:

- By drawing inspiration from [15], [16], we introduce the notion of "sparse strong observability" that generalizes "sparse observability" to the scenario where both sensors and actuators are susceptible to adversarial attacks.
- We develop a secure state estimator by leveraging the SMT paradigm, building on [20]. Furthermore, we propose methods to improve the running time.

This paper is organized as follows. In Section II, we precisely formulate the problem after introducing some notation. Section III gives the main theoretical contribution of this paper. In Section IV, we develop an SMT-based estimator followed by a discussion on improving the running time. Section IV ends with experimental results. Section V concludes the paper.

II. Problem Definition

A. Notation

We represent the sets of real, natural and binary numbers by $\mathbb{R}$, $\mathbb{N}$ and $\mathbb{B}$. Given a vector $x \in \mathbb{R}^n$ and a set $O \subseteq \{1,\dots,n\}$, we use $x|_O$ to denote the vector obtained from x by removing all elements except those indexed by the set O. Similarly, for a matrix $C \in \mathbb{R}^{n_1 \times n_2}$ we use $C|_{(O_1,O_2)}$ to denote the matrix obtained from C by eliminating all rows and columns except the ones indexed by O_1 and O_2, respectively, where $O_i \subseteq \{1,\dots,n_i\}$ with $n_i \in \mathbb{N}$ for $i \in \{1,2\}$. In order to simplify the notation, we use $C|_{(.,O_2)} := C|_{(\{1,\dots,n_1\},O_2)}$ and $C|_{(O_1,.)} := C|_{(O_1,\{1,\dots,n_2\})}$. We denote the complement of O by $\bar{O} := \{1,\dots,n\} \setminus O$. We use the notation $\{x(t)\}_{t=0}^{T-1}$ to denote the sequence $x(0),\dots,x(T-1)$, we drop the sub(super)scripts whenever it is clear from the context.

A Linear Time Invariant (LTI) system is characterized by the following equations:

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \tag{1}$$

where $u(t) \in \mathbb{R}^m$, $x(t) \in \mathbb{R}^n$ and $y(t) \in \mathbb{R}^p$ are the input, state and output variables, respectively, $t \in \mathbb{N}_0$ denotes time, A, B, C and D are system matrices with appropriate dimensions. The order of an LTI system is defined as the dimension of its state space. A trajectory of the system consists of an input sequence with its corresponding output sequence. For an LTI system,

$$\mathscr{O}_{(A,C)} := \begin{bmatrix} C^T & A^TC^T & \dots & (A^T)^{n-1}C^T \end{bmatrix}^T, \tag{2}$$

$$\mathscr{N}_{(A,B,C,D)} := \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & & \ddots & \\ CA^{n-2}B & CA^{n-3}B & \dots & D \end{bmatrix}, \tag{3}$$

are the *observability* and *invertibility* matrices, respectively, where n is the order of the underlying system. In this paper, we often work with subsets of sensors and actuators. For a subset of sensors $K \subseteq \{1,\dots,p\}$, we use the notation $\mathscr{O}_K := \mathscr{O}_{(A,C|_{(K,.)})}$ to denote the observability matrix of sensors in the set K. For a set of actuators $R \subseteq \{1,\dots,m\}$, we denote the invertibility matrix relating R to K by $\mathscr{N}_{R\to K} := \mathscr{N}_{(A,B_{(.,R)},C_{(K,.)},D_{(K,R)})}$. For $x \in \mathbb{R}^n$, we define its support set as the set of indices of its non-zero components, denoted by $\text{supp}(x)$. Similarity we define the support of the sequence as $\text{supp}(\{x(t)\}) := \cup_t \text{supp}(x(t))$.

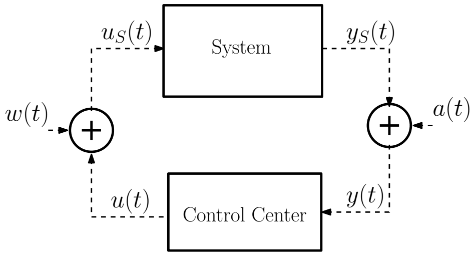

Fig. 1. The generic attack model considering in this paper.

B. System and Attack model

This work is concerned with the problem of attack detection and identification of LTI systems. We consider the scenario in which the sensors and actuators are both susceptible to attacks. The ultimate goal is to reconstruct the state despite these attacks.

The system S, is described by the following equations:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu_S(t), \\ y_S(t) &= Cx(t) + Du_S(t). \end{aligned} \tag{4}$$

Without loss of generality we assume $\begin{bmatrix} B^T & D^T \end{bmatrix}^T$ to be of full rank.

In this set up, the adversary can attack sensors and/or actuators. We model these attacks by additive terms and by imposing a sparsity constraint on them, i.e.,

$$\begin{cases} u_S(t) &= u(t) + w(t), \\ y(t) &= y_S(t) + a(t), \end{cases} \tag{5}$$

where $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ are the controller-designed input and the observed output, respectively, and $w(t) \in \mathbb{R}^m$ and $a(t) \in \mathbb{R}^p$ are signals injected by the malicious agent into the actuators and sensors. In the rest of this paper, we refer to these signals as the attack strategy of the malicious agent. We use the subscript S for signals that directly come from/to the system. The controller can only observe $y(t)$ and compute the input $u(t)$.

When the adversary attacks a sensors (actuator) it can changes its measurement (input) to any arbitrary value, we model this attack by an additive term in (5) without any restriction (statistical or otherwise). The only limitation that we impose on the power of the malicious agent is the number of sensors and actuators under attack. We assume that an upper bound on the number of attacked sensors and actuators is given by a pair (s,r). Therefore, the malicious agent can attack a subset of sensors and actuators denoted by $K \subseteq \{1,\dots,p\}$ and $R \subseteq \{1,\dots,m\}$, respectively, with $|K| \leq s$ and $|R| \leq r$, such that $\text{supp}(\{w(t)\}) \subseteq R$ and $\text{supp}(\{a(t)\}) \subseteq K$. Note that these sets are not known in the controller side and only upper bounds on their cardinality are given. Once the adversary chooses these sets, other sensors and actuators remain unattacked.

Assumption 1: The number of sensors and actuators under attack is bounded by s and r, respectively.

Problem statement: For the linear system (4) under the attack model (5), we seek solutions to the following problems:

- Under which conditions the state reconstruction is possible despite attacks on both sensors and actuators?
- How can we design such an efficient estimator?

III. Necessary and Sufficient Conditions for Secure State Estimation

In this section, we look at this problem from a theoretical perspective. We seek conditions on the underlying system under which the state estimation (possibly with delay) is possible despite attacks on both sensors and actuators.

In some applications, the state of the system is to be estimated but not all the inputs are available or known. To address this problem, the notion of "strong observability" has been introduced in the literature [16]. For strongly observable systems, one can still estimate the state of the system without the knowledge of inputs. The following definition formalizes this concept.

Definition 1 (Strong observability): The system described by the equations (1) is called strongly observable if for any initial state $x(0)$ and any input sequence $u(0), u(1), \ldots$ there exists an integer L such that $x(0)$ can be uniquely recovered from $y(0), y(1), \ldots, y(L)$.

Note that L is always upper-bounded by the order of the system. Linearity of LTI systems directly implies the following corollary.

Lemma 1: System described by the equations (1) is strongly observable if and only if $y(t) = 0$ for $t \in \mathbb{N}_0$ implies that $x(0) = 0$.

Proof: We first prove the sufficiency part. For the sake of contradiction, suppose that the underlying system is not strongly observable but the property of Corollary 1 is true. If the underlying system (4) is not strongly observable, it means there exist two initial conditions, denoted by $x^1(0)$ and $x^2(0)$ possibly with different input sequences denoted by $\{u^1(t)\}$ and $\{u^2(t)\}$, respectively, that correspond to the same output sequence $\{y(t)\}$. The underlying system is linear, therefore the nonzero initial condition of $x^1(0) - x^2(0)$ under the input sequence $\{u^1(t) - u^2(t)\}$ produces the zero output sequence which contradicts the property given in Corollary 1. The necessity can be concluded using the similar argument. For the sake of contradiction let us assume this property does not hold, i.e., there exists a non zero initial state $x(0) \neq 0$ that corresponds to the zero output sequence. This contradicts the strong observability since the zero output sequence can be generated from both zero and $x(0)$ as initial conditions under (possibly different) input sequences. ■

Similarly, it is straightforward to conclude the following corollary.

Corollary 1: System (4) is not strongly observable if and only of there exist a non-zero intial state and an input sequence such that $y(t) = 0$ for $t \in \mathbb{N}_0$

Proof: Follows directly from Lemma 1. ■

The following proposition will be used multiple times in the next section.

Proposition 1 (See section III-B of [31]): The system defined by (1) is strongly observable if and only if the following equality holds:

$$\operatorname{rank}\begin{bmatrix} \mathscr{O}_{(A,C)} & \mathscr{N}_{(A,B,C,D)} \\ I_{n\times n} & 0 \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathscr{O}_{(A,C)} & \mathscr{N}_{(A,B,C,D)} \end{bmatrix}, \quad (6)$$

where 0 is a zero matrix with appropriate dimensions.

In analogy to sparse observability [16], we define the notion of (s,r)-sparse strong observability as follows:

Definition 2 ((s,r)-sparse strong observability): System (4) is (s,r)-sparse strongly observable if for any $\Gamma_y \subseteq \{1,\ldots,p\}$ and $\Gamma_u \subseteq \{1,\ldots,m\}$ with $|\Gamma_y| \leq s$ and $|\Gamma_u| \leq r$, the system $(A, B_{(.,\Gamma_u)}, C_{(\bar{\Gamma_y},.)}, D_{(\bar{\Gamma_y},\Gamma_u)})$ is strongly observable.

Note that in Definition 2, the value of s and r are upper bounded by the number of sensors and actuators, respectively. This modified notion of strong observability is the key for formalizing redundancy across sensors. We show that a necessary and sufficient condition for secure state estimation can be stated using this property. Note that $(s,0)$-sparse strong observability is equivalent to the notion of s-sparse observability that was introduced before in the literature [16]. The following theorem is the main theoretical result in this paper.

Theorem 1: Let the number of attacked sensors and actuators be bounded by s and r, respectively. Under this attack model (5), the state can be reconstructed (possibly with delay) if and only if the underlying system S, is $(2s,2r)$-sparse strongly observable.

Remark 1: It is worth mentioning that the maximum number of attacked sensors, s, cannot be greater than $\lfloor \frac{p}{2} \rfloor$ and it is an inherent limitation of LTI systems with p sensors [16]. However the maximum number of attacked actuators is not inherently restricted by $\lfloor \frac{m}{2} \rfloor$ and can take values up to m, depending on the specific system under the consideration.

Proof: First we show that $(2s,2r)$-sparse strong observability is a sufficient condition for estimating the correct state. For the sake of the contradiction, assume that the state cannot be estimated, i.e., there exist two different (initial) states, denoted by x^1 and x^2, that cannot be distinguished under this attack model. More precisely, there exist two attack strategies that will lead to the exactly same (corrupted) outputs. We denote the adversarial additive terms by $\{w^1(t)\}, \{a^1(t)\}$ and $\{w^2(t)\}, \{a^2(t)\}$ for the first and second attack strategy, respectively. We represent the corresponding inputs and outputs of the system by $\{u_S^1(t)\}, \{y_S^1(t)\}$ and $\{u_S^2(t)\}, \{y_S^2(t)\}$, and the common (corrupted) measured output and the controller input sequences are denoted by $\{y(t)\}$ and $\{u(t)\}$, respectively. We denote the signals related to the first and second scenarios by superscripts ^{1} and ^{2}.

By the assumption of the attack model (5), there exist Γ_u^i, Γ_y^i for $i \in \{1,2\}$ with bounded cardinality such that

$$\operatorname{supp}(\{w^i(t)\}) \subseteq \Gamma_u^i, \operatorname{supp}(\{a^i(t)\}) \subseteq \Gamma_y^i \quad \text{for} \quad i \in \{1,2\}. \tag{7}$$

Note that

$$\forall t: \begin{cases} u_S^1(t) = u(t) + w^1(t) \\ u_S^2(t) = u(t) + w^2(t) \end{cases}, \tag{8}$$

where $u(t)$ is the controller designed input. Therefore

$$u_S^1(t) - w^1(t) = u_S^2(t) - w^2(t) \implies \tag{9}$$

$$\operatorname{supp}(\{u_S^1(t) - u_S^2(t)\}) = \operatorname{supp}(\{w^1(t) - w^1(t)\}) \tag{10}$$
$$\subseteq \Gamma_u^1 \cup \Gamma_u^2.$$

Similarly, it is straightforward to conclude that $\operatorname{supp}(\{y_S^1(t) - y_S^2(t)\}) \subseteq \Gamma_y^1 \cup \Gamma_y^2$. Now, we are ready to reach the contradiction. The underlying system is LTI, thus the input sequence $\{u_S^1(t) - u_S^2(t)\}$ with the initial state $x^1 - x^2$ generates the output sequence $\{y_S^1(t) - y_S^2(t)\}$. The underlying system is $(2s,2r)$-sparse strongly observable so the sub-system $(A, B_{(.,\Gamma_u)}, C_{(\bar{\Gamma}_y,.)}, D_{(\bar{\Gamma}_y,\Gamma_u)})$ is strongly observable for any $|\Gamma_u| = 2r$ and $|\Gamma_y| = 2s$. Let us choose Γ_u and Γ_y as any set of $2r$ inputs and $2s$ outputs containing $\Gamma_u^1 \cup \Gamma_u^2$ and $\Gamma_y^1 \cup \Gamma_y^2$, respectively. Note that $\{y_S^1(t) - y_S^2(t)\}_{\overline{\Gamma_y}}$ is a zero sequence, hence by Lemma 1 we conclude that the corresponding initial state $(x^1 - x^2)$ is zero, which contradicts the assumption of $x^1 \neq x^2$.

Now we prove that $(2s,2r)$-sparse strongly observability is a necessary condition. For the sake of contradiction, suppose that the system described by (4) is not $(2s,2r)$-sparse strongly observable, however, reconstructing the state (possibly with delays) is still possible. We construct two system trajectories with different (initial) states that have exactly the same input and output sequences under suitable attack strategies (additive terms). This implies that estimating the correct state is indeed impossible.

By the assumption of the contradiction, the underlying system is not $(2s,2r)$-sparse strongly observable, so there exist subsets of inputs and outputs denoted by Γ_u and Γ_y, respectively, such that $(A, B_{(.,\Gamma_u)}, C_{(\bar{\Gamma}_y,.)}, D_{(\bar{\Gamma}_y,\Gamma_u)})$ is not strongly observable. Corollary 1 implies that there exist an initial condition Δx and an input sequence $\{\Delta u(t)\}$ (with its support lying inside Γ_u) that generates an output sequence $\{\Delta y(t)\}$ with $\operatorname{supp}(\{\Delta y(t)\}) \subseteq \Gamma_y$. One can rewrite $\Delta u(t)$ and $\Delta y(t)$ as sum of two sparse signals, more precisely:

$$\Delta y(t) = \Delta^1 y(t) + \Delta^2 y(t), \Delta u(t) = \Delta^1 u(t) + \Delta^2 u(t), \tag{11}$$

where cardinality of $\operatorname{supp}(\{\Delta^i y(t)\})$ and $\operatorname{supp}(\{\Delta^i y(t)\})$ are upper-bounded by s and r for $i \in \{1,2\}$, respectively. (for example, we can rewrite $\Gamma_y = \Gamma_y^1 \cup \Gamma_y^2$ where $|\Gamma_y^i| \leq s$ for $i \in \{1,2\}$. Then we define

$$\begin{cases} \Delta^i y(t)|_{\Gamma^i} & := \Delta y(t)|_{\Gamma^i} \\ \Delta^i y(t)|_{\overline{\Gamma^i}} & := 0 \end{cases}, \quad \text{for} \quad i \in \{1,2\}.$$

Now consider the following two different trajectories of the system

$$\begin{cases} \big(u_S^1(t), y_S^1(t)\big) & = \big(\Delta u(t), \Delta y(t)\big) \\ \big(u_S^2(t), y_S^2(t)\big) & = (0,0) \end{cases}, \quad \forall t \tag{12}$$

, with their initial states

$$\begin{cases} x^1(0) & = \Delta x \\ x^2(0) & = 0 \end{cases}, \tag{13}$$

and their corresponding attack strategies,

$$\begin{cases} \big(w^1(t), a^1(t)\big) = \big(-\Delta^1 u(t), -\Delta^1 y(t)\big) \\ \big(w^2(t), a^2(t)\big) = \big(\Delta^2 u(t), \Delta^2 y(t)\big) \end{cases}, \quad \forall t \tag{14}$$

It is straightforward to verify that $\{y^1(t)\} = \{y^2(t)\}$ and $\{u^1(t)\} = \{u^2(t)\}$, i.e., under the attack model (5) the observed outputs and controlled inputs are exactly the same, and the proof is complete. ■

IV. An SMT-based estimator

In this section, we propose an algorithm that estimates the state despite attacks on actuators and sensors, followed by designing an estimator. We work with batches of data in order to estimate the state. We use capital letters to represent these batches,

$$Y^n(t) := \begin{bmatrix} y(t-n+1)^T & \dots & y(t)^T \end{bmatrix}^T, \tag{15}$$

$$W^n(t) := \begin{bmatrix} w(t-n+1)^T & \dots & w(t)^T \end{bmatrix}^T. \tag{16}$$

Whenever n is the order of the underlying system, we may drop the superscript for ease of notation. For a subset of sensors (actuators), denoted by $K \subseteq \{1,\dots,p\}$ ($R \subseteq \{1,\dots,m\}$), we use the notation $Y^n|_K(t)$ ($W^n|_R(t)$) for the batches of length n that only consists of sensors (actuators) in the set K (R). Based on the attack model (5), the input to the system is decomposed into two additive terms, controller-designed input $u(t)$ and the adversarial input $w(t)$. The underlying system (4) is linear and therefore we can easily exclude the effect of the controller-designed input from the output by subtracting its effect. Hence, without loss of generality we assume that the true $u(t)$ is zero. The proposed algorithm is based on the following proposition.

Proposition 2: Suppose that the underlying system is $(2s,2r)$-sparse strongly observable, and the number of attacked sensors and actuators are bounded by s and r, respectively. Given any subset of sensors and actuators denoted by K and R with $|K| \geq p - s$ and $|R| \leq r$, the first statement below implies the second:

1) There exist $\hat{U} \in \mathbb{R}^{n|R|}$ and $\hat{x} \in \mathbb{R}^n$ such that
$$Y|_K(t) = \mathscr{O}_K \hat{x} + \mathscr{N}_{R \to K} \hat{U}. \tag{17}$$
2) The estimated state $\hat{x}$, is equal to the actual state of the system at time $t-n+1$, $x(t-n+1)$, where n is the order of the underlying system.

Remark 2: The underlying system is $(2s,2r)$-sparse strongly observable therefore $(A, B_{(.,R)}, C_{(K,.)}, D_{(K,R)})$ is strongly observable. If (17) has a solution, then $\hat{x}$ would be the unique solution for x (see section III-B of [31]).

Proof: Let us denote the set of attack-free sensors and under-attack actuators by K_{s} and R_{a}. At most s sensors are under attack, therefore there is a subset K with at least $p-2s$ attack-free sensors, we denote it by K_s. Note that $Y|_{K_s}$ can be written as follows:

$$Y|_{K_s} = O_{K_s}x(t-n+1) + \mathscr{N}_{R\to K_s}W|_R + \mathscr{N}_{R_a\setminus R\to K_s}W|_{R_a\setminus R}. \tag{18}$$

On the other hand, we can rewrite (17) by taking only sensors in K_s,

$$Y|_{K_s} = O_{K_s}\hat{x} + \mathscr{N}_{R\to K_s}\hat{U} + \mathscr{N}_{R_a\setminus R\to K_s}0, \tag{19}$$

where 0 is a zero matrix with appropriate dimensions. The underlying system is $(2s,2r)$-sparse strongly observable, therefore the sub-system $\hat{S} := (A, B_{(.,R\cup R_a)}, C_{(K_s,.)}, D_{(K_s,R\cup R_a)})$ is strongly observable. One can reinterpret both equations as two (possibly different) valid trajectories of the system $\hat{S}$ that share the same output sequence. Strong observability of $\hat{S}$ implies that $\hat{x} = x(t-n+1)$ and completes the proof. ■

The main algorithm in this paper builds up on this proposition. Basically we are searching for a set of sensors and actuators that satisfies (17), i.e., we check if there exist $\hat{U}$ and $\hat{x}$ that make the equality (17) hold. For each sensor (actuator), we assign a binary variable $b_i \in \mathbb{B}$ ($c_i \in \mathbb{B}$) that indicates if the corresponding sensor (actuator) is under attack or not, i.e., $b_i = 1$ ($c_i = 1$) if the i^{th} sensor (actuator) is under attack. This task is combinatorial in nature and in order to efficiently decide which set of sensors and actuators satisfies the test, we construct a detection algorithm using lazy SMT paradigm [32].

A. Architecture

As in IMHOTEP-SMT [20], our solver consists of two blocks that interact with each other, a SAT solver and a Theory solver. The former reasons about the combination of Boolean and pseudo-Boolean constraints and produces a feasible instance of $b \in \mathbb{B}^p$ and $c \in \mathbb{B}^m$, based on the current status of the SAT solver. The initial pseudo-Boolean constraint only bounds the number of attacked sensors and actuators, i.e.,

$$\Phi_B := (\sum_{i=1}^{p} b_i \leq s) \bigwedge (\sum_{j=1}^{m} c_i \leq r). \tag{20}$$

The theory solver checks the equality (17) for $K := \overline{\text{supp}(b)}$ and $R := \text{supp}(c)$. If the equality is satisfied, then the algorithm terminates and returns the (delayed) estimate of the state. Otherwise, the Theory solver outputs UNSAT and generates a reason for the conflict, a certificate, or a counterexample that is denoted by Φ_{cert}. This counterexample encodes the inconsistency among the chosen attack-free actuators and sensors. The following always constitutes a naive certificate.

$$\Phi_{\text{naive-cert}} := \sum_{i\in\overline{\text{supp}(b)}} b_i + \sum_{j\in\overline{\text{supp}(d)}} c_j \geq 1. \tag{21}$$

In the rest of this section, we show how we can build smaller certificates and hence improve the run time. On the next iteration, the SAT solver updates the constraint by conjoining Φ_{cert} to Φ_B, and generates another feasible assignment for b and c. This procedure is repeated until the theory solver returns SAT as illustrated in Algorithm 1.

Algorithm 1: Pseudo-code of the proposed algorithm.

```
input  : $A, B, C, D$, $Y$ (output), $s, r$ ;
output:  $(x, b, c)$ ;
status ← UNSAT ;
$\Phi_{\text{cert}} \leftarrow$ True ;
$\Phi_B \leftarrow (\sum_{i\in\{1,\ldots,p\}} b_i \leq s) \bigwedge (\sum_{i\in\{1,\ldots,m\}} c_i \leq r)$ ;
while status == UNSAT do
    $\Phi_B \leftarrow \Phi_B \bigwedge \Phi_{\text{cert}}$;
    (b,c) ← SAT-solver($\Phi_B$) ;
    (status, x) ← T-solver.check($\overline{\text{supp}(b)}, \text{supp}(c)$);
    $\Phi_{\text{cert}} \leftarrow$ T-solver.certificate($\overline{\text{supp}(b)}, \text{supp}(c)$);
end
return $(x, b, c)$
```

Note that Proposition 2 implies that SAT solver eventually produces an assignment that satisfies the test and therefore Algorithm 1 always terminates. The size of the certificate plays an important role in the overall execution time of the algorithm. In the rest of this section, we focus on constructing shorter counterexamples.

Algorithm 2: T-solver.check

```
Solve: $(x, U) = \text{argmin}_{x,U} \|Y|_K - \mathscr{O}_K x - \mathscr{N}_{R\to K}U\|$ ;
if $\|Y|_K - \mathscr{O}_K x - \mathscr{N}_{R\to K}U\| == 0$ then
    status = SAT ;
else
    status = UNSAT ;
end
return (status, $x$)
```

B. Shortening the SAT certificate

In this section, we improve the efficiency of Algorithm 1 by constructing a shorter certificate. As it was discussed before, the naive certificate contains at least $p+m-s-r$ of assigned boolean variables. This certificate only excludes the current assignment of b and c from the search space of the SAT solver, however, by exploiting the structure of the underlying system, we show that we can further decrease the size of the certificate and therefore prune the search space more efficiently. In order to generate a shorter certificate, we look for a subset of sensors and actuators that cannot be all attack-free. One of the main results of this paper is to show that we can always find a shorter conflicting subset of sensors and actuators. In this paper, we propose two different methods for generating shorter certificates. The first method guarantees a counterexample with a size of at most $p+m-2s-2r+2$, we explain this method in detail and give a formal proof of the existence of such shorter certificate. The second approach generates two shorter certificates at each iteration.

Let us assume that the SAT solver hypothesized K and R as the set of attack-free sensors and under-attack actuators, respectively. We aim to shorten the size of the counterexample, i.e., we look for a $K_{\text{temp}} \subseteq K$ and $R_{\text{temp}} \supseteq R$ that would not satisfy the equality (17). For the first method, we do this in two steps. In the first step, we increase the size of conflicting (supposedly under attack) actuators and afterwards we decrease the size of the supposedly safe sensors. We begin by arbitrarily adding actuators to R to get a subset of size $\max(m,2r)$ denoted by R_{temp}. If T-solver.check(K,R_{temp}) returns UNSAT, we pick R_{temp} as the augmented set of conflicting actuators. Otherwise, we search for an actuator $j \in R_{\text{temp}} \setminus R$ such that T-solver.check$(K,R_{\text{temp}} \setminus \{j\})$ returns UNSAT. The following lemma guarantees existence of such an actuator.

Algorithm 3: T-solver.certificate 1

```
input  : $K, R, x$ ;
output: $\Phi^1_{\text{cert}}$ ;
step 1: Conduct a linear search in the actuator set ;
pick a set of size max($m, 2r$): $R_{\text{temp}} \supseteq R$ ;
$\tilde{R} \leftarrow R_{\text{temp}}$ ;
(status,$x$) ← T-Solver.check($K, \tilde{R}$) ;
while status == SAT do
    (status,$x$) ← T-Solver.check($K, \tilde{R}$) ;
    pick another actuator index $j \in R_{\text{temp}} \setminus R$ ;
    $\tilde{R} \leftarrow R_{\text{temp}} \setminus \{j\}$ ;
end
step 2: Conduct a linear search in the sensor set ;
pick a set of size $p-2s$: $K_{\text{temp}} \subseteq K$ ;
$\tilde{K} \leftarrow K_{\text{temp}}$ ;
(status,$x$) ← T-Solver.check($\tilde{K}, \tilde{R}$) ;
while status == SAT do
    (status,$x$) ← T-Solver.check($\tilde{K}, \tilde{R}$) ;
    pick another sensor index $i \in K_{\text{temp}} \setminus K$ ;
    $\tilde{K} \leftarrow K_{\text{temp}} \setminus \{i\}$ ;
end
$\Phi^1_{\text{cert}} \leftarrow \sum_{i \in \tilde{K}} b_i + \sum_{j \in \overline{\tilde{R}}} c_j \geq 1$ ;
return $\Phi_{\text{cert}}$
```

Lemma 2: Suppose that the system S is $(2s,2r)$-sparse strongly observable, and the number of attacked sensors and actuators are bounded by s and r, respectively. Pick any subset of sensors and actuators denoted by K and R with $|K| \geq p-s$ and $|R| \leq r$, that do not satisfy the equality (17). Given any subset of at most $\max(2r,m)$ actuators denoted by $R_{\text{temp}} \supseteq R$, one of the following is true:

1) T-solver.check(K,R_{temp}) returns UNSAT,
2) There exists a $j \in R_{\text{temp}} \setminus R$ such that T-solver.check$(K,R_{\text{temp}} \setminus \{j\})$ returns UNSAT.

Proof: We prove it by contradiction. The idea behind the proof is to show if none of the above are true then K and R should satisfy the equality which contradicts the assumption. For the sake of contradiction, let us assume none of the above statements are true, i.e., not only (K,R_{temp}) satisfies the equality but also for any $j \in R_{\text{temp}} \setminus R$ T-solver.check$(K,R_{\text{temp}} \setminus \{j\})$ returns SAT. Let us fix j:

$$Y|_K = \mathscr{O}_K \hat{x}^0 + \mathscr{N}_{R_{\text{temp}\setminus\{j\}} \to K} \hat{U}^0|_{R_{\text{temp}}\setminus\{j\}} + \mathscr{N}_{\{j\} \to K} \hat{U}^0|_j, \tag{22}$$

$$Y|_K = \mathscr{O}_K \hat{x}^j + \mathscr{N}_{R_{\text{temp}}\setminus\{j\} \to K} \hat{U}^j + \mathscr{N}_{\{j\} \to K_s} 0, \tag{23}$$

where $\hat{x}^0, \hat{x}^j \in \mathbb{R}^n$ are states that T-solver.check returns, $\hat{U}^0, \hat{U}^j$ are matrices with appropriate dimensions that satisfy the equality, and 0 is a zero matrix with appropriate dimensions. Note that the underlying system is $(2s,2r)$-sparse strongly observable, $|R_{\text{temp}}| \leq 2r$ and $|K| \geq p-s$ therefore $\hat{S} := (A, B_{(.,R_{\text{temp}})}, C_{(K,.)}, D_{(K,R_{\text{temp}})})$ is strongly observable. One can reinterpret (22) and (23) as two (possibly different) valid trajectories of a strongly observable system $\hat{S}$ with identical output sequences. Strong observability implies that the state can be uniquely determined from the output with a delay bounded by n, therefore $\hat{x}^0 = \hat{x}^j$. Furthermore, the equality of right hand sides of (22) and (23) implies that $\mathscr{N}_{\{j\} \to K} \hat{U}^0|_j = 0$, more precisely there exists a solution to the equality (17) for (K,R_{temp}) with the corresponding $\hat{U}^0|_j$ equals to zero. If (23) holds for all $j \in R_{\text{temp}\setminus T}$, it follows that $\mathscr{N}_{R_{\text{temp}} \setminus R \to K} \hat{U}^0|_{R_{\text{temp}} \setminus R} = 0$, and therefore we have:

$$Y|_K = \mathscr{O}_K \hat{x}^0 + \mathscr{N}_{R \to K_s} \hat{U}^0|_R, \tag{24}$$

$$\tag{25}$$

that contradicts the assumption that T-solver.check(K,R) return UNSAT and the proof is complete. It should be clear that $\hat{x}^0$ and $\hat{x}^j$ are not necessarily equal to the state of the underlying system. ∎

Let us denote the new set of actuators by $\tilde{R}$ which consists of at least $2r-1$ actuators. At the second step, we shrink the set of conflicting sensors in order to further shorten the size of the counterexample, let us denoted an arbitrary subset of K of size $p-2s$ by K_{temp}. Similar to Lemma 2, the following lemma shows we can reduce the size of conflicting sensors at least by $s-1$.

Lemma 3: Suppose that the system S is $(2s,2r)$-sparse strongly observable, and the number of attacked sensors and actuators are bounded by s and r, respectively. Pick any subset of sensors and actuators denoted by K and $\tilde{R}$ with $|K| \geq p-s$ and $|\tilde{R}| \leq 2r$, that do not satisfy the equality (17). Given any subset of at most $p-2s$ sensors denoted by $K_{\text{temp}} \subseteq K$, one of the following is true:

1) T-solver.check$(K_{\text{temp}}, \tilde{R})$ returns UNSAT,
2) There exists a $i \in K \setminus K_{\text{temp}}$ such that T-solver.check$(K_{\text{temp}} \cup \{i\}, \hat{R})$ returns UNSAT.

Proof: The proof is quite similar to the proof of Lemma 2. For the sake of the contradiction, we assume both of statements are not true, therefore we have the following for any $i \in K \setminus K_{\text{temp}}$,

$$Y|_{K_{\text{temp}}} = \mathscr{O}_{K_{\text{temp}}}\hat{x}^0 + \mathscr{N}_{\tilde{R}\to K_{\text{temp}}}\hat{U}^0, \tag{26}$$

$$Y|_{K_{\text{temp}}\cup\{i\}} = \mathscr{O}_{K_{\text{temp}}\cup\{i\}}\hat{x}^i + \mathscr{N}_{\tilde{R}\to K_{\text{temp}}\cup\{i\}}\hat{U}^i, \tag{27}$$

where $\hat{x}^0, \hat{x}^i \in \mathbb{R}^n$ are states that T-solver.check returns, $\hat{U}^0, \hat{U}^i$ are matrices with appropriate dimensions that satisfy the equality. Using same arguments and by strong observability of $(A, B_{(.,\tilde{R})}, C_{(K,.)}, D_{(K,\tilde{R})})$, we conclude that $\hat{x}^0 = \hat{x}^i$ and furthermore $\hat{U}^0 = \hat{U}^i$. Therefore we can rewrite $Y|_K$ as follows

$$Y|_K = \mathscr{O}_K\hat{x}^0 + \mathscr{N}_{\tilde{R}\to K}\hat{U}^0, \tag{28}$$

which contradicts the assumption that T-solver.check$(K,\tilde{R})$ returns UNSAT. ■

We denote this smaller set of conflicting sensors by $\tilde{K}$. Lemma 2 and 3 give formal guarantees of the existence of shorter certificates which hold no matter how the subsets of sensors and actuators are chosen. In practice, we choose these subsets based on heuristics that have for objective a decrease in the overall algorithms running time. We assign slack variables to sensors and actuators similarly to [20] and sort them based on the structure of the system. The summary of the above procedure of shortening certificates is illustrated in Algorithm 3.

As it was noted before, we propose two different approaches for shortening the counterexample and therefore improving the running time. The second method generates two certificates by reducing the size of "supposedly" safe sensors and actuators separately. Therefore the theory solver produces two counterexamples at each iteration. Suppose that the SAT solver hypothesized K and R as the set of attack-free sensors and under-attack actuators. In the first step, the theory solver looks for a larger subset of actuators $\tilde{R} \supseteq R$ for which T-solver.check$(K,\tilde{R})$ returns UNSAT, Lemma 2 guarantees the existence of such set. The first counterexample consists of sensors in K and actuators in $\tilde{R}$. The second certificate is constructed by reducing the size of sensors and keeping the same set of "supposedly" under-attack actuators, K. Algorithm 4 illustrates this procedure.

C. Simulation Results

We implemented our SMT-based estimator in MATLAB while interfacing with the SAT solver SAT4J [33]. In this subsection, we assess the performance of our algorithm by using both the certificates. We randomly generate systems with a fixed state dimension ($n = 20$) and increase the number of sensors and actuators. In each experiment, twenty percent of sensors and actuators are under adversarial attacks, and we randomly generate the support set for the adversarial signals. All the systems under experiment satisfy a suitable sparse strong observability condition. Figures 2 and 3 report the results of the simulations. All the experiments run on an Intel Core i5 2.7GHz processor with 16GB of RAM. We verify the run-time improvement by using the shorter certificates, Φ^1_{cert} and Φ^2_{cert}, compared to the theoretical upper-bound of the brute-force approach in Figure 2. For instance, consider the scenario with $p = 24$ and $m = 10$ in Figures 2 and 3. In the brute-force approach, we require to check all $\binom{24}{4}$ × $\binom{10}{2}$ $\approx 4.8 * 10^5$ different combinations of sensors and actuators, however, by exploiting either Φ^1_{cert} or Φ^2_{cert} we observe a substantial improvement. It is worth mentioning the Φ^2_{cert} gives better performance than Φ^1_{cert} for this set of experiments.

Algorithm 4: T-solver.certificate 2

```
input : $K, R, x$ ;
output: $\Phi^2_{\text{cert}}$ ;
step 1: Conduct a linear search in the actuator set ;
pick a set of size max$(2r, m)$: $R_{\text{temp}} \supseteq R$ ;
$\tilde{R} \leftarrow R_{\text{temp}}$ ;
$(\text{status}, x) \leftarrow$ T-Solver.check$(K, \tilde{R})$ ;
while status == SAT do
    $(\text{status}, x) \leftarrow$ T-Solver.check$(K, \tilde{R})$ ;
    pick another actuator index $j \in R_{\text{temp}} \setminus R$ ;
    $\tilde{R} \leftarrow R_{\text{temp}} \setminus \{j\}$ ;
end
$\Phi^2_{\text{cert}} \leftarrow \sum_{i\in K} b_i + \sum_{j\in\overline{T}} c_j \geq 1$ ;
step 2: Conduct a linear search in the sensor set ;
pick a set of size $p - 2s$: $K_{\text{temp}} \subseteq K$ ;
$\tilde{K} \leftarrow K_{\text{temp}}$ ;
$(\text{status}, x) \leftarrow$ T-Solver.check$(\tilde{K}, R)$ ;
while status == SAT do
    $(\text{status}, x) \leftarrow$ T-Solver.check$(\tilde{K}, R)$ ;
    pick another sensor index $i \in K_{\text{temp}} \setminus K$ ;
    $\tilde{K} \leftarrow K_{\text{temp}} \setminus \{i\}$ ;
end
$\Phi^2_{\text{cert}} \leftarrow \Phi^2_{\text{cert}} \wedge \left(\sum_{i\in\tilde{K}} b_i + \sum_{j\in\overline{T}} c_j \geq 1\right)$ ;
return $\Phi_{\text{cert}}$
```

V. Conclusion

In this paper, we considered the problem of secure state estimation when sensors and/or actuators are under adversarial attacks. In this set-up, there is no restriction on how the adversary manipulates sensors and actuators. By introducing the notion of "sparse strong observability", we derived necessary and sufficient conditions under which the state estimation is possible given bounds on the number of attacked sensors and actuators. Furthermore, we supported the theory by developing an SMT-based estimator that reconstruct the state.

References

[1] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[2] A. Greenberg, "Hackers remotely kill a jeep on the highway, with me in it," *[online] http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway*, 2015.

[3] L. Kelion, "Nissan leaf electric cars hack vulnerability disclosed," *[online] http://www.bbc.com/news/technology-35642749*, 2016.

[4] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems.," in *HotSec*, 2008.

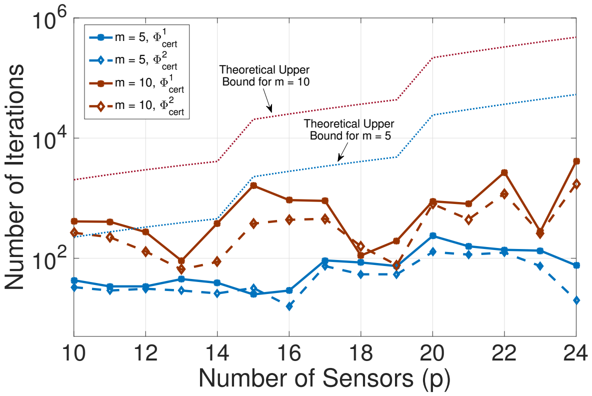

Fig. 2. Number of iterations in Algorithm 1 using Φ_{cert^1} and Φ_{cert^2} versus the number of sensors (p) and actuators (m). The dotted lines are the theoretical upper-bounds for the number of iterations in the brute force approach.

[5] S. Sundaram, M. Pajic, C. N. Hadjicostis, R. Mangharam, and G. J. Pappas, "The wireless control network: monitoring for malicious behavior," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 5979–5984, 2010.

[6] S. Amin, G. A. Schwartz, and A. Hussain, "In quest of benchmarking security risks to cyber-physical systems," *IEEE Network*, vol. 27, no. 1, pp. 19–24, 2013.

[7] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[8] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 804–808, 2014.

[9] C. De Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2930–2944, 2015.

[10] D. Senejohnny, P. Tesi, and C. De Persis, "A jamming-resilient algorithm for self-triggered network coordination," *arXiv preprint arXiv:1603.02563*, 2016.

[11] A. Gupta, C. Langbort, and T. Basar, "Optimal control in the presence of an intelligent jammer with limited actions," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 1096–1101, 2010.

[12] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 911–918, IEEE, 2009.

[13] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *Control Systems Magazine, IEEE*, vol. 35, no. 1, pp. 82–92, 2015.

[14] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 5967–5972, 2010.

[15] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[16] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2016.

[17] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *American Control Conference (ACC)*, pp. 2439–2444, 2015.

[18] Y. Nakahira and Y. Mo, "Dynamic state estimation in the presence of compromised sensory data," in *54th Annual Conference on Decision and Control (CDC)*, pp. 5808–5813, IEEE, 2015.

[19] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," *preprint arXiv:1504.05566*, 2015, accepted to IEEE Transactions on Networked Control Systems, 2016.

[20] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. PP, no. 99, pp. 1–1, 2017.

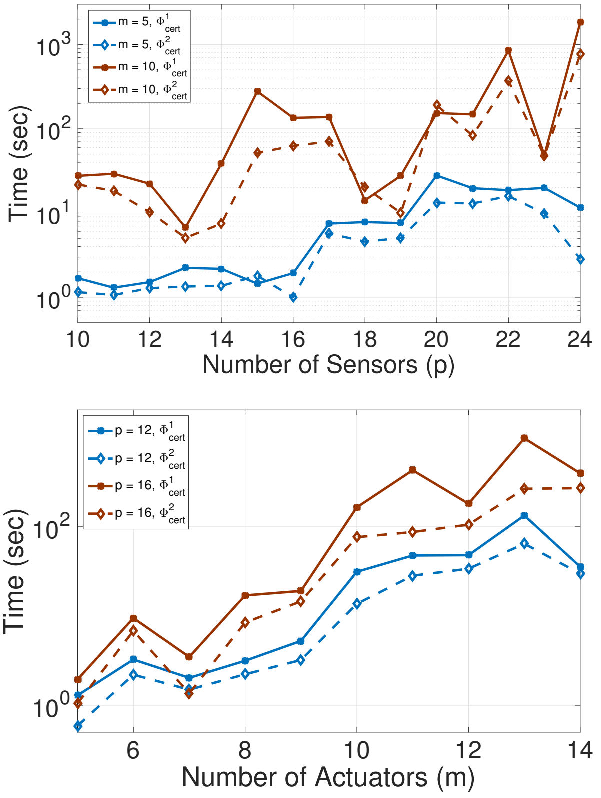

Fig. 3. Execution time of Algorithm 1 using Φ_{cert^1} and Φ_{cert^2} versus the number of sensors (p) and actuators (m).

[21] S. Z. Yong, M. Q. Foo, and E. Frazzoli, "Robust and resilient estimation for cyber-physical systems under adversarial attacks," in *American Control Conference (ACC), 2016*, pp. 308–315, IEEE, 2016.

[22] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ICCPS'14: ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014)*, pp. 163–174, 2014.

[23] A. Tiwari, B. Dutertre, D. Jovanović, T. de Candia, P. D. Lincoln, J. Rushby, D. Sadigh, and S. Seshia, "Safety envelope for security," in *ACM Proceedings of the 3rd international conference on High confidence networked systems*, pp. 85–94, 2014.

[24] M. Showkatbakhsh, P. Tabuada, and S. Diggavi, "System identification in the presence of adversarial outputs," in *Decision and Control (CDC), IEEE 55th Conference on*, pp. 7177–7182, IEEE, 2016.

[25] M. Showkatbakhsh, P. Tabuada, and S. Diggavi, "Secure system identification," in *54th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1137–1141, IEEE, 2016.

[26] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder, *Diagnosis and fault-tolerant control*, vol. 691. Springer, 2006.

[27] H. L. Jones, *Failure detection in linear systems.* PhD thesis, Massachusetts Institute of Technology, 1973.

[28] F. Harirchi and N. Ozay, "Guaranteed model-based fault detection in cyber-physical systems: A model invalidation approach," *arXiv preprint arXiv:1609.05921*, 2016.

[29] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[30] H. Sandberg and A. M. Teixeira, "From control system security indices to attack identifiability," in *Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*, pp. 1–6, IEEE, 2016.

[31] T. Yoshikawa and S. Bhattacharyya, "Partial uniqueness: Observability and input identifiability," *IEEE Transactions on Automatic Control*, vol. 20, no. 5, pp. 713–714, 1975.

[32] C. W. Barrett, R. Sebastiani, S. A. Seshia, and C. Tinelli, "Satisfiability

modulo theories.," *Handbook of satisfiability*, vol. 185, pp. 825–885, 2009.

[33] D. Le Berre and A. Parrain, "The sat4j library, release 2.2, system description," *Journal on Satisfiability, Boolean Modeling and Computation*, vol. 7, pp. 59–64, 2010.