Fifty-fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 27 - 30, 2016

Secure System Identification

Mehrdad Showkatbakhsh, Paulo Tabuada and Suhas Diggavi

Abstract— **This work is concerned with the identification of linear time-invariant systems in the presence of an adversarial agent that attacks sensor measurements. The attacker is omniscient and we impose no restrictions (statistical or otherwise) on how the adversary alters the sensor measurements. We work in a noisy scenario where, in addition to the attacks, the sensor measurements are also affected by additive noise. Given a bound on the number of attacked sensors, and under a certain observability condition, we show that we can still construct a model that is useful for stabilization. Furthermore, we show that this model is closely related to the original system.**

I. Introduction

Cyber-attacks used to be restricted to the cyber domain, however, recent events such as StuxNet malware [1] and the recently reported vulnerabilities on modern cars [2], [3] motivated the recent interest in security of cyber-physical systems 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], man-in-the-middle [12], false data injection [13] etc.

System Identification is of great importance in control theory and is crucial for designing controllers. The previously mentioned attacks on cyber-physical systems motivate the need for new methods of system identification in order to defend against such attacks. In this short paper we focus on the identification of systems under attacks on the sensor measurements.

Among a variety of security problems, our work is the closest to the line of research on the secure state estimation [14], [15], [16], [17], [18], [19] and [20] are a few of many works on this topic. The results in [21], [22], [18], [23] and [24] all rely on the knowledge of the system dynamics to defend against attacks. However, our goal is to identify the underlying model. In another line of work, Tiwari et. al. [25] considered the problem of sensor spoofing attack detection without the prior knowledge of the system dynamics. The proposed method relies on the availability of attack-free streams of data. In contrast, our method can be applied directly to the corrupted data and does not rely on the existence of attack-free data.

The problem of system identification under adversarial attacks is an ill-posed problem. Clearly, one should not expect to recover the correct measurement of an attacked sensor without any prior knowledge of the underlying system. Furthermore, detection of attacked sensors is not always possible. For instance consider the system in Figure 1 labeled "attack free" and its attacked version labeled "under attack". The attack consists in changing the output of the p^{th} sensor from $c_p x$ to $c'_p x$. Since the resulting system is still LTI, we cannot expect to distinguish the attacked system from the un-attacked system in the bottom of Figure 1 solely based on the (corrupted) measured data.

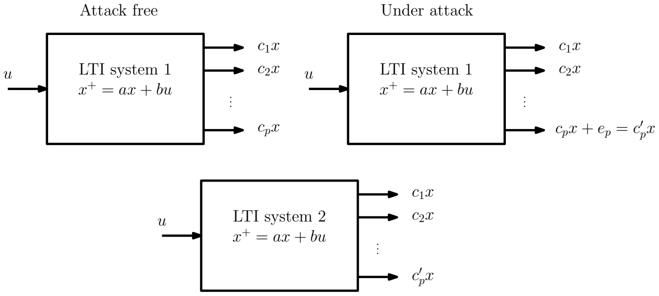

Fig. 1: An example that illustrates the impossibility of exact system identification. Under a proper attack strategy, LTI systems 1 and 2 are indistinguishable after one sensor attack, even though they satisfy proper sparse observability conditions.

In a recent paper [26] we have shown that under suitable observability conditions, it is still possible to construct a meaningful model using only the corrupted data. Furthermore, this model can be used to stabilize the underlying systems. In [26], sensor measurements are either exact or corrupted by the adversary. In this paper, we extend the result to the noisy scenario where the sensor measurements are also affected by additive noise.

The main contribution of this paper is as follows. We show that even in the noisy scenario, one can construct a meaningful model that is closely related to the underlying system. Furthermore, this model is indeed useful for one prominent control purpose, stabilization.

This paper is organized as follows. In Section II the problem formulation is precisely introduced after establishing the notation. We present the main theorem in Section III. Section IV concludes the paper.

II. Preliminaries and Problem Definition

A. Notation and Definitions

We represent real numbers and vectors by lower case letters, matrices with capital letters. The sets of natural and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively. 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

This work was supported in part by NSF awards 1136174 and 1321120.

The authors are with the UCLA Electrical Engineering Department, University of California at Los Angeles, CA 90095-1594 { `mehrdadsh, tabuada, suhas` } `@ucla.edu`

978-1-5090-4550-1/16/$31.00 ©2016 IEEE

ones indexed by the set O. Similarly, for a matrix C with n columns we use the notation $C|_O$ to denote a matrix obtained from C by eliminating all columns except ones indexed by O.

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

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

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

$$\mathscr{O} := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}, \quad \mathscr{C} := \begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix}, \quad (2)$$

are the observability and controllability matrices, respectively.

A system is called "observable" if the state sequence can be uniquely determined given the knowledge of input and output sequences. It is well-known [27] that an LTI system is observable if and only if its observability matrix is injective. Observability of an LTI system depends solely on the pair of matrices (A,C), therefore we may refer to a pair of matrices as an observable pair if the corresponding observability matrix is injective, i.e., any LTI system with this pair of matrices is observable. A system is called "controllable" if there exists an input sequence that transfers the state $x(t)$ from x_0 to the origin, for any $x_0 \in \mathbb{R}^n$.

Throughout this paper, we always assume the underlying system to be observable and controllable. We denote the set of all such systems with $\mathbb{L}$. Our goal is system identification and since we only work with input and output variables, the unobservable part of the state is of no importance to us. Systems that are not controllable cannot be identified using only the knowledge of the input and output sequences (see for example [28] for a through discussion on identifiability). We say a system S explains the data sequence if the sequence is a valid trajectory of the system.

In the rest of this section, we briefly discuss the concepts that are needed for developing our method. The notion of s-sparse observability was first introduced by Shoukry et. al. [20] for the purpose of secure state estimation. In Section III we will show that s-sparse observability is also useful for the identification of systems under adversarial attacks.

Definition 1: **(s-sparse observable)** A system $S := (A,B,C,D)$ is s-sparse observable if by removing any s outputs, the system remains observable, i.e., for any $\Gamma \subseteq \{1,\dots,p\}$ with $|\Gamma| = p - s$, (A, C_Γ) is an observable pair.

Two LTI systems $S_i : (A_i,B_i,C_i,D_i), i \in \{1,2\}$ are called "similar" if there exists a linear change of coordinates P such that $A_2 = PA_1P^{-1}$, $B_2 = PB_1$, $C_2 = C_1P^{-1}$ and $D_2 = D_1$ [27]. Two similar systems share the same input-output characteristics, i.e., they have the same transfer function. In the context of system identification, one can only identify a system up to this equivalence class. In [26], authors introduced the notion of similarity modulo outputs to characterize the identifiablity of systems under attacked outputs. This concept is crucial in order to understand the main results. We define it here using only state-space notions.

Definition 2: **(Similar modulo outputs)** Two LTI systems $S_i : (A_i,B_i,C_i,D_i), i \in \{1,2\}$ are similar modulo outputs if:

1) Both have the same order, n, and the same input space, $\mathbb{R}^r$.
2) There exists an n dimensional subspace that is invariant under the dynamics of $S_p := (A_p,B_p,C_p,D_p)$ (parallel composition), where

$$A_p := \mathrm{diag}(A_1,A_2), \qquad B_p := \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad (3)$$

$$C_p := \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \qquad D_p := \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}. \quad (4)$$

Similarity modulo outputs is an equivalence relation [26] and we denote it by $\sim$. The following proposition is useful in understanding this notion.

Proposition 1 (see Lemma 1 in [26]): Two systems S_1 and S_2 are similar modulo outputs if and only if there exists a change of coordinate P such that $A_2 = PA_1P^{-1}$ and $B_2 = PB_1$.

Basically, two systems are similar modulo outputs if they share the same internal dynamical structure.

B. Problem formulation

We consider the problem of system identification of a noisy LTI system when some of sensor measurements are under adversarial attacks. Assume our noisy system, S, can be described by the following equations:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y_S(t) &= Cx(t) + Du(t) + \varepsilon(t), \end{aligned} \quad (5)$$

where $u(t) \in \mathbb{R}^r$, $x(t) \in \mathbb{R}^n$ and $y_S(t) \in \mathbb{R}^p$ are input, state and output variables, respectively. The random variable $\varepsilon(t)$ represents additive noise, which is assumed to be Independent and Identically Distributed (i.i.d.) with zero mean.

We assume that identification is possible when there is no attack on the sensors.

Assumption 1 (Identifiablity given the input sequence): System (5) is identifiable from the attack-free data.

A variety of identification algorithms guarantee exact identification of an LTI system up to a similarity transformation under mild conditions on the input sequence and the dynamics of the system (see for example [28]). However, when the sensor measurements are corrupted by additive noise, identification task becomes more challenging and requires further consideration in order to remove the effect of noise. In the literature of system identification, there exist several methods that guarantee exact identification when the length of the training sequence tends to infinity (see for example [29], [30]). Subspace identification algorithms are one of the most prominent such methods. It is known that under mild conditions on the input sequence and an upper-bound on the order of system, it is still possible to identify the system, see for example [29]. Our approach does not rely on a specific identification algorithm, instead we transform the secure system identification problem to a problem that can be solved with any of the existing algorithms for noisy system identification.

We impose no restriction on how the adversary manipulates the measurements of attacked sensors, however, we assume an upper-bound on the number of attacked sensors is given.

Assumption 2 (Bound on the number of attacked sensors): We assume that an upper bound s on the number of attacked sensors is given.

More specifically, we assume the attacker can choose a subset of sensors K and can only attack sensors in this set. This is a realistic assumption since the time it takes for the adversarial agent to attack other sensors is large compared to the time scale of the modern control systems. Although the upper-bound on the cardinality of K is known, we do not know K itself.

The sensor measurements are given by, $y(t) = y_S(t) + y_{\text{attack}}(t)$ for $t \in \{0,\dots,T-1\}$, where $y_{\text{attack}}(t)$ is the signal injected by the adversary at time t and T is the length of the sequence. Note that $y_{\text{attack}}(t)$ is a sparse signal, i.e., $y_{\text{attack}}(t)|_{\{1,\dots,p\}\setminus K} = 0$ and we do not impose any other restriction on $y_{\text{attack}}(t)|_K$.

Problem statement: Given the input sequence $\{u(t)\}_{t=0}^{T-1}$, and the corrupted output sequence $\{y(t)\}_{t=0}^{T-1}$, we seek answers to the following problems:

1) Determine a model that can explain the attack-free sensor measurements $\{y_{\text{attack}}(t)|_{\{1,\dots,p\}\setminus K}\}_{t=0}^{T-1}$, which is closely related to the underlying system.
2) Stabilize System (5) using the identified model.

III. Secure system identification

In this section we show that $2s$-sparse observability is the key notion required to solve the secure identification problem. We show that $2s$-sparse observable systems can be securely identified under Assumptions 1 and 2. We begin by stating how the proposed secure identification algorithm works. Our method looks for a subset of sensors $O \subseteq \{1,\dots,p\}$ with $|O| \geq p-s$ such that $\{(u(t), y(t)|_O)\}_{t=0}^{T-1}$ can be explained by an s-sparse observable system. Note that under Assumptions 1 and 2 such a subset always exists since the number of attacked sensors is upper-bounded by s, however this subset may not be unique.

Algorithm 1 summarizes the proposed method. Let us denote the output of this algorithm by S', which can be either $\emptyset$ (i.e., the model cannot be identified) or an identified model. In the case where the sensor measurements are exact, it can be shown that under Assumptions 1 and 2 this algorithm constructs a model S' that is similar modulo outputs to System (5) (as directly implied from Theorem 2 in [26]). In the noisy scenario, however, Assumption 1 can only be satisfied asymptotically, i.e., when the number of data points T, tends to infinity (see for example [29]). Moreover, there might be instances of the noise sequence which lead the algorithm to the wrong model. However, it can be shown that when the number of measurements tends to infinity, the probability of misidentifying the system approaches zero. Note that in this case, for any finite T, the identified model would not be "exactly" similar modulo outputs to the underlying LTI system, however, it would converge to a model which is similar modulo outputs to S.

As we elaborated before, two LTI systems that are similar to each other will have the exact same input-output characteristics, and in the context of system identification they are equivalent. We use the notation $\mathscr{L}$ to denote the set of equivalence classes of this relation.[1] In the rest of this section, with a slight abuse of notation we use S as its corresponding equivalence class of similarity transformation. We reserve $[S]$ for the equivalence class of similarity-modulo outputs. In order to formalize our convergence argument, we define the notion of ε-similarity modulo outputs.

Definition 3 (ε-similarity modulo outputs): Two LTI systems S_1 and S_2 are ε-similar modulo outputs if $d([S_1],[S_2]) < \varepsilon$, where d is the metric on the space of equivalence classes of similarity modulo outputs that makes the quotient map $(\mathscr{L}, d_0) \to (\mathscr{L}/\sim, d)$ continuous, and d_0 is the metric[2] on the space of equivalence classes of similarity.[3]

Algorithm 1: Pseudo-code of the proposed method.

```
Let $O_i$ for $i \in \{1,\dots,i_{\max}\}$ denote all possible subsets
of size $p-s$ of $\{1,\dots,p\}$;
input : $\{u(t)\}_{t=0}^{T-1}$ (Input), $\{y(t)\}_{t=0}^{T-1}$ (Output),
        Identify() (a system identification algorithm);
output: $S'$;
$S' \leftarrow \emptyset$;
$i \leftarrow 1$;
while $i \leq i_{max}$ do
    $S_{\text{temp}} \leftarrow \text{Identify}\left(\{u(t)\}_{t=0}^{T-1}, \{y(t)|_{O_i}\}_{t=0}^{T-1}\right)$;
    if $S_{temp}$ is s-sparse observable? then
        $S' \leftarrow S_{\text{temp}}$;
        break;
    end
    $i \leftarrow i+1$;
end
return $S'$
```

Theorem 1: Let us denote the output of Algorithm 1 by S'. For any $\varepsilon > 0$ the probability that S' is not ε-similar modulo outputs to the underlying LTI system S, approaches zero when the number of data points tends to infinity, provided

[1] Similar systems are similar modulo outputs, i.e., similarity modulo outputs induces an equivalence relation on $\mathscr{L}$, with a slight abuse of notation we also denote this relation by $\sim$.

[2] Note that d_0 should make the quotient map $(\mathbb{L}, d_S) \to (\mathscr{L}, d_0)$ continuous, where $\mathbb{L}$ is the set of LTI systems (matrices (A,B,C,D)) equipped with the metric d_S.

[3] These metrics exist. One can always endow the quotient space of a metric space with a (pseudo)metric. The only difference between pseudometrics and metrics is topological, and for T_0 topological spaces, a pseudometric is also a metric [31]. In this case, with a properly defined distance d_S on $\mathbb{L}$, quotient spaces ($\mathscr{L}$ and $\mathscr{L}/\sim$) are T_0, therefore d and d_0 can be constructed.

that S is $2s$-sparse observable and Assumptions 1 and 2 hold. Furthermore any feedback controller that stabilizes S' also stabilizes S with probability approaching one.

Proof: First we argue that the probability of $S' = \emptyset$ converges to zero as the number of measurements T, goes to infinity. Note that there always exists a set $O \subseteq \{1,\dots,p\}$ with $|O| = p - s$ that $\{(u(t), y(t)|_O)\}_{t=0}^{T-1}$ can be explained by an s-sparse observable system since at most s sensors are under attack and S is $2s$-sparse observable. We know that the noisy identification algorithm guarantees the exact model when the number of measurements tends to infinity. Therefore, with probability approaching one the identified model for this subset O is s-sparse observable. We conclude that Algorithm 1 does not return $\emptyset$ with high probability. However, such a subset and model are not unique.

Let us assume that the output of Algorithm 1 is not $\emptyset$, and we denote the corresponding subset by O'. We prove S' is ε-similar modulo outputs to S with high probability. Note that S' is not the "exact" model of the corresponding subset of sensors, we denote this exact model by $\hat{S}$. The noisy identification algorithm (identify(.) in Algorithm 1) guarantees the identification of the exact model when the number of measurements tends to infinity, i.e., for any $\varepsilon' > 0$ we have

$$\lim_{T\to\infty} \Pr(d_0(S', \hat{S}) < \varepsilon') = 1, \tag{6}$$

where d_0 is the metric on the space of equivalence classes of similarity transformation.

Our argument consists in 2 steps. First we argue that if S' is s-sparse observable then $\hat{S}$ is also s-sparse observable with probability approaching one. In the second step, we claim that if $\hat{S}$ is s-sparse observable then it should be similar modulo outputs to S, hence $d([S'],[\hat{S}]) = d([S'],[S])$, where d is the metric on the space of equivalence classes of similarity modulo outputs. The quotient map $S \mapsto [S]$ is continuous, and (6) holds for any $\varepsilon' > 0$. Therefore we conclude that

$$\lim_{T\to\infty} \Pr(d([S'],[S]) < \varepsilon) = 1, \tag{7}$$

or equivalently S' is ε-similar modulo outputs to S with probability approaching one.

Note that s-sparse observability is a robust property, i.e., an s-sparse observable system remains s-sparse observable by a small enough perturbation of the system dynamics, therefore (6) implies that if S' is s-sparse observable then $\hat{S}$ is also s-sparse observable with probability approaching one.

Now we prove that if $\hat{S}$ is s-sparse observable, then it should be similar modulo outputs to S. By Assumption 2 at most s sensors are under attack hence there exists a subset of O', denoted by O'_{clean} that corresponds to $p - 2s$ attack-free sensors. Note that $\hat{S}$ is the exact model and therefore $\hat{S}|_{O_{\text{clean}}}$ and $S|_{O_{\text{clean}}}$ both represent the same input-output behavior. More precisely, they are related by a similarity transformation.

Proposition 2 (see Corollary 1 in [26]): Assume that $S := (A,B,C,D)$ is an s- sparse observable system with p sensors. For any $\Gamma \subseteq \{1,\dots,p\}$ with $|\Gamma| \geq p - s$, $S|_\Gamma := (A,B,C|_\Gamma,D|_\Gamma)$ is similar modulo outputs to S.

This proposition implies that $\hat{S} \sim \hat{S}|_{O_{\text{clean}}}$ and $S|_{O_{\text{clean}}} \sim S$. Similarity modulo outputs is an equivalence relation [26], hence $S \sim \hat{S}$.

We showed that with high probability the identified model is ε-similar modulo outputs to S. Now we are ready to prove the stability result. First we show that any controller that stabilizes $\hat{S} := (\hat{A},\hat{B},\hat{C},\hat{D})$ also stabilizes S. Proposition 1 implies that there exists a linear change of coordinates, P, such that $\hat{A} = PAP^{-1}$ and $\hat{B} = PB$. Let us denote the state variable corresponding to these systems by $\hat{x}(t)$ and $x(t)$, respectively. According to the definition of similarity modulo outputs and given the fact that both systems have the same input sequences, we know that $\hat{x}(t) = Px(t)$. The controller makes $\hat{S}$ asymptotically stable, i.e., $\lim_{t\to\infty}\|\hat{x}(t)\| = 0$. Note that $\|x(t)\| \leq \|P^{-1}\|\|\hat{x}(t)\|$, so $\lim_{t\to\infty}\|x(t)\| = 0$. It is a standard result in control theory that a small enough perturbation on system dynamics does not affect the stabilization for feedback controllers. Therefore any feedback controller that stabilizes S' also stabilizes $\hat{S}$ with probability approaching one. We conclude that the same feedback controller makes S asymptotically stable with high probability.

■

We conclude this section by briefly discussing the requirement of Assumption 1. Note that our method is predicated on Assumption 1 being satisfied. One should take this consideration into account when implementing our proposed method. Subspace identification algorithms require a long training sequence in order to guarantee identification in the noisy scenario. Therefore, the secure identification method also requires a long training sequence.

IV. Conclusion

In this paper we considered the problem of the secure system identification under an adversarial attack on the sensor measurements. The attacker is omniscient and there is no restriction on how it alters the sensor measurements. We work in a noisy scenario in which sensor measurements are corrupted by additive noise in addition to the adversarial attack. It has been shown that given a bound on the number of attacked sensors and under suitable observability conditions one can still construct a meaningful model of the system that can be used for the stabilization.

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.

[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] 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.

[13] 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.

[14] 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.

[15] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 24–45, 2015.

[16] 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.

[17] Y. Shoukry, M. Chong, M. Wakiaki, P. de Nuzzo, A. D. Sangiovanni-Vincentelli, S. Seshia, J. P. Hespanha, and P. Tabuada, "Smt-based observer design for cyber physical systems under sensor attacks," in *American Control Conference (ACC)*, pp. 2439–2444, 2015.

[18] 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.

[19] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *Automatic Control, IEEE Transactions on*, vol. 60, no. 4, pp. 1145–1151, 2015.

[20] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, 2013.

[21] Y. Mo, J. P. Hespanha, and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 31–43, 2014.

[22] 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.

[23] S. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *IEEE International Conference on Decision and Control (CDC)*, 2015.

[24] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 2929–2933, 2015.

[25] 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.

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

[27] P. J. Antsaklis and A. N. Michel, *Linear systems*. Springer Science & Business Media, 2006.

[28] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and approximate modeling of linear systems: A behavioral approach*, vol. 11. SIAM, 2006.

[29] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory-Implementation-Applications*. Springer Science & Business Media, 2012.

[30] L. Ljung, *System identification: Theory for the User*. Englewood Cliffs, 1987.

[31] N. R. Howes, *Modern analysis and topology*. Springer Science & Business Media, 2012.