# The Watermark-Securable Subspace of a Linear System Containing a Single Malicious Actuator

Bharadwaj Satchidanandan and P. R. Kumar, *Fellow, IEEE*

***Abstract*—Consider a multiple-input, multiple-output, perfectly observed linear dynamical system containing an arbitrary set of malicious sensors, and at most one malicious actuator. The malicious sensors need not report their measurements truthfully and a malicious actuator may not apply inputs in accordance with the control law. The honest actuators in the system, if there are any, employ Dynamic Watermarking in order to detect the presence of malicious nodes. The state space of such a system can be decomposed into two orthogonal subspaces, called the watermark-securable and the watermark-unsecurable subspaces, such that the malicious sensors and actuators cannot degrade the state estimation performance of the honest sensors and actuators along the watermark-securable subspace if they wish to remain undetected. This paper presents a precise characterization of the watermark-securable subspace for any system containing at most one malicious actuator.**

***Index Terms*—Cyber-Physical Systems, Security, Securable Subspace, Unsecurable Subspace, Dynamic Watermarking.**

## I. Introduction

Cyber-Physical Systems (CPS) can potentially play an indispensable role in confronting some of the major challenges that societies face in areas such as energy, water, heath-care, manufacturing, and transportation [1]. Examples include increasing the penetration of renewable energy, sustaining water and other natural resources for a growing population, automating transportation systems to improve their safety and efficiency, etc. However, one of the foremost impediments to the rapid proliferation of CPSs is their vulnerability to cyber attacks. Many instances of attacks on real-world cyber-physical systems in the past, such as the Stuxnet attack [2] and the Maroochy-Shire incident [3], reaffirm this concern. Unless provable security guarantees are provided for cyber-physical systems, there will be resistance to their large-scale adoption, especially in sectors such as energy or transportation where a security breach could result in severe economic consequences or even loss of life.

The approach of Dynamic Watermarking [4]–[8] provides provable security guarantees to control systems against arbitrary attack strategies. Most prior works on this topic have addressed the scenario where an arbitrary combination of sensors could be malicious, but all actuators are honest. Another line of work [9]–[11] studies the setting where an arbitrary combination of both sensors and actuators could be malicious, but wherein the honest actuators do not employ Dynamic Watermarking for defense. The notions of securable and unsecurable subspaces of a linear system were introduced in this context [9], and their operational significance established [9]–[11]. In this paper, we go beyond these two settings and consider a system where in addition to an arbitrary combination of sensors being malicious, at most one arbitrary actuator could also be malicious, and the honest actuators, if there are any, employ Dynamic Watermarking to detect the presence of adversarial sensors and actuators. For any such system, we (i) show that its state space can be decomposed into two orthogonal subspaces, called the watemark-securable and the watermark-unsecurable subspaces, such that the malicious sensors and actuators cannot degrade the state estimation performance of the honest sensors and actuators along the watermark-securable subspace if they wish to remain undetected, and (ii) present a precise characterization of the watermark-securable and the watermark-unsecurable subspaces.

The rest of the paper is organized as follows. Section II describes some related work on the topic of CPS security. Section III formulates the problem that is addressed in the paper. Section IV contains the main result of the paper, viz., a characterization of the watermark-securable subspace of a linear system containing at most one malicious actuator. Section V concludes the paper.

**Notation:** The following notation is used throughout the paper. Given a vector $\mathbf{v}$, we denote by $v_i$ the $i^{th}$ component of $\mathbf{v}$, and by $\mathbf{v}^{TT}$ the matrix $\mathbf{v}\mathbf{v}^T$. Given a sequence $\{\mathbf{v}\}$, we denote by $\mathbf{v}^t$ the sequence $\{\mathbf{v}[0], \dots, \mathbf{v}[t]\}$. Given a vector $\mathbf{v}$ and a subspace $\mathcal{S}$, we denote by $\mathbf{v}_{\mathcal{S}}$ the orthogonal projection of $\mathbf{v}$ on the subspace $\mathcal{S}$. Given a subspace $\mathcal{S}$, $\mathcal{S}^{\perp}$ denotes the orthogonal complement of $\mathcal{S}$ and $P_{\mathcal{S}}$ denotes the matrix that projects a vector right-multiplying it onto the subspace $\mathcal{S}$. Given two subspaces $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{X} \oplus \mathcal{Y}$ denotes their span, and $\mathcal{X}_{\mathcal{Y}}$ denotes the projection of subspace $\mathcal{X}$ on subspace $\mathcal{Y}$. Given a matrix $A$, we denote by $\mathcal{R}(A)$ its range space.

## II. Related Work

Three classes of attack detectors for CPS, namely, static detectors, dynamic detectors and active detectors, have been defined in [12]. Limitations on what attacks can be detected by each of these classes of attack detectors are also derived. The problem of estimating the state of a system in the presence of adversarial sensors is addressed in [13], [14]. The problem of designing systems that are robust to stealthy attacks is addressed in [15], and bounds on the state estimation error that an adversarial actuator can introduce while remaining stealthy are derived in [16]. Reference [4] introduces the idea of Physical Watermarking to detect replay attacks in a control system, and [17] shows how it can aid in detecting a more sophisticated attack. However, as shown in [7], there exist attacks that cannot be detected by the attack

978-1-7281-3187-0/20/$31.00 ©2020 IEEE

detectors proposed in [4], [17] and related works. Consequently, it is shown in [6], [7] that by a careful design of tests for attack detection, Dynamic Watermarking can serve as a common defense strategy against not just specific attacks, but arbitrary attacks which introduce "significant" distortion. The security guarantee provided by Dynamic Watermarking for more general systems is presented in [18]. These tests are presented in an asymptotic form, and a finite-time statistical version of these tests is presented in [19] and its efficacy demonstrated. Necessary and sufficient conditions for designing security-guaranteeing dynamic watermarks for systems affected by arbitrarily distributed noise is addressed in [20]. A demonstration of the efficacy of Dynamic Watermarking in securing real-world cyberphysical systems are reported in [19], [21].

## III. Problem Setup

Consider a multiple-input, multiple-output, stochastic linear dynamical system described by

$$\begin{aligned}\mathbf{x}[t+1] &= A\mathbf{x}[t] + B\mathbf{u}[t] + \mathbf{w}[t+1],\\ \mathbf{y}[t+1] &= C\mathbf{x}[t+1],\end{aligned} \tag{1}$$

where $\mathbf{x}[t] \in \mathbb{R}^p$ is the state of the system at time $t$, $\mathbf{u}[t] \in \mathbb{R}^m$ is the input applied to the system at time $t$, $\{\mathbf{w}\}$ is the process noise sequence, independent and identically distributed (i.i.d.) across time with $\mathbf{w}[1] \sim \mathcal{N}(0, \sigma_W^2 I)$, and $A, B$ and $C$ are matrices of appropriate dimensions. We restrict attention to perfectly-observed systems in this paper, and therefore, the measurement matrix $C$ is simply the identity matrix of size $p$.

The setting that we consider in this paper is one where an arbitrary subset of sensors and at most one arbitrary actuator[1] could be malicious in system (1). Henceforth, by "node," we refer to a sensor or an actuator. We allow for the malicious nodes know the identity of all other malicious nodes in the system. On the other hand, the honest nodes do not know which other nodes are honest or malicious, or even if there are any malicious nodes present in the system. We also suppose that there is an underlying communication network connecting all the nodes using which the malicious nodes can exchange information among themselves if they wish to do so, and carry out coordinated attacks.

A malicious sensor may not report the measurements that it observes in a truthful manner. Instead, it could distort its measurements in an arbitrary manner and report the distorted measurements to the honest nodes. We denote by $\mathbf{z}[t]$ the measurements that the sensors report at time $t$ to the honest nodes in the system. If sensor $i$, $i \in \{1, \dots, p\}$, is honest, then $z_i[t] = x_i[t]$ for all $t$. While the malicious sensors can report different measurements to different nodes, the fact that nodes can exchange the reported measurements among themselves to check for consistency of the reported measurements constrains the malicious sensors to report the same values to all honest nodes in the system. The malicious nodes, however, can exchange the true measurements among themselves, so that they have the knowledge of both the true measurements as well as the measurements that the malicious nodes have reported to the honest nodes in the system.

The honest actuators employ Dynamic Watermarking in order to detect the presence of adversarial nodes in the system. The approach involves the honest actuators superimposing on their control policy-specified input a random signal known as the private excitation or "watermark," and conducting certain tests to detect the presence of malicious sensors. To elaborate, denote by $g_{i,t} : \mathbb{R}^{p\times(t+1)} \to \mathbb{R}$ the arbitrary and possibly history-dependent control policy that actuator $i$, $i \in \{1, \dots, m\}$, is supposed to employ at time $t$, so that $g_{i,t}(\mathbf{z}^t)$ is the control policy-specified input of actuator $i$ at time $t$ (recall the notation $\mathbf{z}^t$ from Section I). Each honest actuator $i$, $i \in \{1, \dots, m\}$, chooses a private excitation $e_i[t]$ at time $t$ according to the distribution $\mathcal{N}(0, \sigma_e^2)$, independent of all other random variables realized until that time, and superimposes it on its control policy-specified input. While the control policy-specified input of actuator $i$ as well as the statistics of its private excitation sequence are known to all nodes in the system, including the malicious nodes, actuator $i$ does not reveal the actual realization of its private excitation sequence to any other node in the system. As shown in prior works [6]–[8], it is this concealment of private excitation that bestows upon an honest actuator the ability to diagnose the system and detect the presence of malicious nodes. The net input applied by actuator $i$ at time $t$ is therefore

$$u_i[t] = g_{i,t}(\mathbf{z}^t) + e_i[t].$$

A malicious actuator can apply its inputs in an arbitrary fashion. Recall that in this paper, we restrict attention to the case where there could be at most one malicious actuator. We suppose without any loss of generality that it is actuator $m$ that could be malicious, since the actuators can always be relabelled so that the malicious actuator, if there is one, is indexed $m$. Define

$$e_m[t] := u_m[t] - g_{m,t}(\mathbf{z}^t).$$

Since actuator $m$ could be malicious, the statistics of the sequence $\{e_m\}$ need not be equal to the statistics that is prescribed for the private excitation sequence, viz., i.i.d. normal with variance $\sigma_e^2$ and at each time, independent of all other random variables realized until that time.

The system (1) evolves in closed loop as

$$\begin{aligned}\mathbf{x}[t+1] = A\mathbf{x}[t] + B\mathbf{g}_t(\mathbf{z}^t) + B^H\mathbf{e}^H[t] + B^M e_m[t]\\ + \mathbf{w}[t+1],\end{aligned} \tag{2}$$

where the vector $\mathbf{e}^H[t] := [e_1[t] \dots e_{m-1}[t]]^T$, the vector $\mathbf{g}_t(\mathbf{z}^t) := [g_{1,t}(\mathbf{z}^t) \dots g_{m,t}(\mathbf{z}^t)]^T$, $B^H$ is the sub-matrix of $B$ formed by its columns corresponding to the honest actuators, and $B^M$ is the sub-matrix of $B$ formed by its column corresponding to the malicious actuator. We always regard

[1] More generally, the results of this paper hold as long as the columns of the matrix $B$ corresponding to malicious actuators span a subspace of dimension no more than one.

$B^M$ as a column vector, and so the case when there are no malicious actuators is thought of as there being an artificial malicious actuator with $B^M = 0$. We will also assume throughout this paper that $\mathcal{R}(B^M) \cap \mathcal{U} = \{0\}$, where $\mathcal{U}$ is the unsecurable subspace of system (2) and $\mathcal{R}(B^M)$ denotes the range space of the matrix $B^M$. The unsecurable subspace is defined in [9]–[11] and also recapitulated in Section IV below.

In order to check whether or not there are any malicious nodes in (2), the following two tests are conducted by the honest nodes.

**Test 1:** Each honest actuator $i$ checks if

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} e_i[k](\mathbf{z}[k+1] - A\mathbf{z}[k] - B\mathbf{g}_k(\mathbf{z}^k)) = B_{\cdot i}\sigma_e^2, \tag{3}$$

where $B_{\cdot i}$ denotes the $i^{th}$ column of the matrix $B$.

**Test 2:** Each honest node checks if

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{z}[k+1] - A\mathbf{z}[k] - B\mathbf{g}_k(\mathbf{z}^k))^{TT} = BB^T\sigma_e^2 + \sigma_W^2 I. \tag{4}$$

The notation $(\cdot)^{TT}$ in the above equality can be recalled from Section I.

If and only if either of these tests fail does an honest node declare the presence of adversarial nodes in the system. Note that if all nodes in the system are honest, then both the tests are passed by the reported measurements $\{\mathbf{z}\}$ almost surely. Even though these tests are presented as asymptotic tests, as shown in [19], they can be converted into finite-time statistical tests with desired detection delays and false alarm rates using standard approaches.

Suppose that the reported measurements pass tests (3) and (4) so that the malicious nodes, if there are any, remain undetected. In that case, the honest nodes do not have any reason to suspect that there is any malicious node in the system, and consequently, their estimate of the system's state at time $t$ is simply $\mathbf{z}[t]$. The actual state of the system at time $t$, however, is $\mathbf{x}[t]$, which may or may not be equal to $\mathbf{z}[t]$. Define the state estimation error incurred by the honest nodes at time $t$ as

$$\mathbf{m}[t] := \mathbf{z}[t] - \mathbf{x}[t]. \tag{5}$$

The question that we address in the paper is this: What is the largest subspace $\mathcal{WS}$ of the state space such that if the reported measurements $\{\mathbf{z}\}$ pass Tests (3) and (4), then, the projection of the state estimation error $\{\mathbf{m}\}$ onto the subspace $\mathcal{WS}$ is of zero power? I.e., what is the largest subspace $\mathcal{WS}$ such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{WS}}[k]||^2 = 0. \tag{6}$$

This subspace is termed the "watermark-securable subspace" of system (2). Whatever attack strategy the malicious sensors and actuators employ, they cannot degrade the state estimation performance of the honest nodes along the watermark-securable subspace if they wish to remain undetected. The next section presents a precise characterization of this subspace.

## IV. The Watermark-Securable Subspace of a Linear System Containing a Single Malicious Actuator

In this section, we characterize the watermark-securable subspace of system (2). A few definitions are in order before presenting the result. We first recall from [9] the notions of securable and unsecurable subspaces of a linear system. Let $C^H$ be the sub-matrix of $C$ formed by its rows corresponding to the honest sensors.

**Definition 1** ( [9])**.** The *Unsecurable subspace* of system (2) is the maximal subspace $\mathcal{U} \subseteq \mathbb{R}^p$ such that for all $\mathbf{u} \in \mathcal{U}$,

1) $C^H\mathbf{u} = 0$, and
2) there exists $\mathbf{d}$ such that $A\mathbf{u} + B^M\mathbf{d} \in \mathcal{U}$.

**Definition 2** ( [9])**.** The *Securable subspace* $\mathcal{S}$ of system (2) is the orthogonal compliment of its unsecurable subspace. I.e.,

$$\mathcal{S} := \mathcal{U}^\perp.$$

Recall from Section III the assumption that $\mathcal{R}(B^M) \cap \mathcal{U} = \{0\}$. From this assumption and the fact that $B^M$ is a column vector, it follows that there exists a matrix $M_{\mathcal{S}\to\mathcal{R}(B^M)}$ such that $M_{\mathcal{S}\to\mathcal{R}(B^M)}[B^M x]_{\mathcal{S}} = B^M x$ for all $x \in \mathbb{R}$. We define the noise-securable subspace in terms of this matrix.

**Definition 3.** The *Noise-securable subspace* $\mathcal{NS}$ of system (2) is the largest subspace of the state space such that

$$\mathcal{NS} \subseteq \mathcal{U}, \tag{7}$$

and for every $\mathbf{r} \in \mathcal{R}^\perp(B^H)$,

$$P_{\mathcal{NS}} M_{\mathcal{S}\to\mathcal{R}(B^M)} P_{\mathcal{S}} \mathbf{r} = P_{\mathcal{NS}}\mathbf{r}. \tag{8}$$

It is a simple exercise to verify that the set of subspaces that satisfy (7) and (8) is closed under the operation of span, so that if $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are two subspaces that satisfy (7) and (8), then $\mathcal{Y}_1 \oplus \mathcal{Y}_2$ also satisfies (7) and (8). Consequently, the term "largest" is well-defined in the above definition; the noise-securable subspace is essentially the span of all subspaces that satisfy (7) and (8).

**Definition 4.** The *Noise-unsecurable subspace* $\mathcal{NU}$ of system (2) is the orthogonal complement of the noise-securable subspace that is contained in the unsecurable subspace. I.e.,

$$\mathcal{NU} := \mathcal{NS}^\perp \cap \mathcal{U}. \tag{9}$$

That $\mathcal{R}(B^M) \cap \ \mathcal{U} = \{0\}$ also implies that for every vector $\mathbf{m} \in \mathcal{U}$, there exists a unique vector $B^M d^{\mathcal{U}}$ such that $-A\mathbf{m} + B^M d^{\mathcal{U}} \in \mathcal{U}$. To see this, suppose that $B^M d^1$ and $B^M d^2$ are two vectors such that $-A\mathbf{m} + B^M d^1 \in \mathcal{U}$ and $-A\mathbf{m} + B^M d^2 \in \mathcal{U}$. Then, the difference of these two vectors must also belong to $\mathcal{U}$, i.e., $B^M(d^1 - d^2) \in \mathcal{U}$. Hence, $B^M(d^1 - d^2) \in \mathcal{R}(B^M) \cap \mathcal{U} = \{0\}$, which in turn implies that $B^M d^1 = B^M d^2$. If $B^M \neq 0$, this in turn implies that there exists a unique input $d^{\mathcal{U}}$ that the malicious actuator can apply such that $-A\mathbf{m} + B^M d^{\mathcal{U}} \in \mathcal{U}$. It can also be shown that there exists a matrix $F \in \mathbb{R}^{1 \times p}$ such that for every $\mathbf{m} \in \mathcal{U}$, the unique vector $B^M d^{\mathcal{U}}$ such that $-A\mathbf{m} + B^M d^{\mathcal{U}} \in \mathcal{U}$ is $B^M F\mathbf{m}$ [22]. We define the subspace $\mathcal{WS}_1$ in terms of this matrix $F$.

**Definition 5.** The subspace $\mathcal{WS}_1$ is the largest subspace such that

$$\mathcal{WS}_1 \subseteq \mathcal{NS}, \tag{10}$$

for every $\mathbf{r} \in \mathcal{WS}_1^{\perp} \cap \mathcal{U}$,

$$P_{\mathcal{WS}_1}[(A - B^M F)\mathbf{r}] = 0, \tag{11}$$

and

$$\lim_{k\to\infty} [P_{\mathcal{WS}_1}(A - B^M F)]^k \mathcal{WS}_1 = \{0\}. \tag{12}$$

**Definition 6.** The *watermark-securable subspace* $\mathcal{WS}$ of system (2) is

$$\mathcal{WS} := \mathcal{WS}_1 \oplus \mathcal{S}, \tag{13}$$

and its *watermark-unsecurable subspace* $\mathcal{WU}$ is

$$\mathcal{WU} := \mathcal{WS}^{\perp}. \tag{14}$$

We are now ready to present the main result of the paper.

**Theorem 1.** *Consider system (2) and suppose that the reported sequence of measurements* $\{\mathbf{z}\}$ *pass Tests (3) and (4). Then,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{WS}}[k+1]||^2 = 0. \tag{15}$$

*Proof.* That $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{S}}[k+1]||^2 = 0$ has been established in our prior work [23, Theorem 1]. We only need to show that $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{WS}_1}[k+1]||^2 = 0$.

Since the sequence $\{\mathbf{z}\}$ satisfies (4), it follows that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{m}[k+1] - A\mathbf{m}[k] + B^H \mathbf{e}^H[k] + B^M e_m[k] + \mathbf{w}[k+1])_{\mathcal{S}}^{TT} = P_{\mathcal{S}}(BB^T\sigma_e^2 + \sigma_w^2 I)P_{\mathcal{S}}^T.$$

Resolving the vectors $\mathbf{m}[k]$ and $\mathbf{m}[k+1]$ into their constituent components along subspaces $\mathcal{S}$ and $\mathcal{U}$ and simplifying, we get

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{m}_{\mathcal{S}}[k+1] - A\mathbf{m}_{\mathcal{S}}[k] - A\mathbf{m}_{\mathcal{U}}[k] + B^H \mathbf{e}^H[k] + B^M e_m[k] + \mathbf{w}[k+1])_{\mathcal{S}}^{TT} = P_{\mathcal{S}}(BB^T\sigma_e^2 + \sigma_w^2 I)P_{\mathcal{S}}^T. \tag{16}$$

Define

$$\bar{e}_m[k] := e_m[k] - d^{\mathcal{U}}[k] = e_m[k] - F\mathbf{m}_{\mathcal{U}}[k]. \tag{17}$$

Substituting (17) in (16) and using the fact that $-A\mathbf{m}_{\mathcal{U}}[k] + B^M d^{\mathcal{U}}[k] \in \mathcal{U}$, we get

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{m}_{\mathcal{S}}[k+1] - A\mathbf{m}_{\mathcal{S}}[k] + B^H \mathbf{e}^H[k] + B^M \bar{e}_m[k] + \mathbf{w}[k+1])_{\mathcal{S}}^{TT} = P_{\mathcal{S}}(BB^T\sigma_e^2 + \sigma_w^2 I)P_{\mathcal{S}}^T.$$

It follows from [23, Theorem 1] that the sequence $\{\mathbf{m}_{\mathcal{S}}\}$ is of zero power. Using this, the above equality simplifies to

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (B^H \mathbf{e}^H[k] + B^M \bar{e}_m[k] + \mathbf{w}[k+1])_{\mathcal{S}}^{TT} = P_{\mathcal{S}}(BB^T\sigma_e^2 + \sigma_w^2 I)P_{\mathcal{S}}^T. \tag{18}$$

Define $\sigma$-algebra $\mathcal{F}_k := \sigma(\mathbf{x}^k, \mathbf{e}^{H^{k-1}}, e_m^k)$. Then, $(\mathbf{e}^H[k-1], \mathcal{F}_k)$ is a martingale difference sequence, $(\mathbf{w}[k], \mathcal{F}_k)$ is a martingale difference sequence, and $\bar{e}_m[k] \in \mathcal{F}_k$. Using the Martingale Stability Theorem (MST) [24], we have

$$\sum_{k=0}^{T-1} \mathbf{w}[k+1]\bar{e}_m[k] = \begin{bmatrix} o(\sum_{k=0}^{T-1} \bar{e}_m^2[k]) \\ \vdots \\ o(\sum_{k=0}^{T-1} \bar{e}_m^2[k]) \end{bmatrix}, \tag{19}$$

and

$$\sum_{k=0}^{T-1} \mathbf{e}^H[k]\bar{e}_m[k] = \begin{bmatrix} o(\sum_{k=0}^{T-1} \bar{e}_m^2[k]) \\ \vdots \\ o(\sum_{k=0}^{T-1} \bar{e}_m^2[k]) \end{bmatrix}. \tag{20}$$

Expanding the Left-Hand Side (LHS) of (18), using (19) and (20) to simplify the result, yields

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (B^M \bar{e}_m[k])^{TT} = B^M (B^M)^T \sigma_e^2. \tag{21}$$

Consider first the case that $B^M \neq 0$. Then, the above equality implies that $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \bar{e}_m^2[k] = \sigma_e^2$. Multiplying and dividing the Right Hand Side (RHS) of (19) and (20) by the term $\sum_{k=0}^{T-1} \bar{e}_m^2$, dividing the two equalities by $T$, letting $T \to \infty$, and using (21) implies

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbf{w}[k+1]\bar{e}_m[k](B^M)^T = 0, \tag{22}$$

and

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbf{e}^H[k]\bar{e}_m[k](B^M)^T = 0. \tag{23}$$

Consider next the case when $B^M = 0$. Equalities (22) and (23) hold trivially in that case.

Now, since the sequence $\{\mathbf{z}\}$ satisfies (3), we have

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k](\mathbf{m}[k+1]-A\mathbf{m}[k]+B^H\mathbf{e}^H[k]\\ +B^M e_m[k]+\mathbf{w}[k+1])^T=\sigma_e^2(B^H)^T\end{aligned}$$

whence

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k](\mathbf{m}[k+1]-A\mathbf{m}[k]+B^M e_m[k])^T=0.$$

Resolving the vectors $\mathbf{m}[k]$ and $\mathbf{m}[k{+}1]$ into their constituent components along subspaces $\mathcal{S}$ and $\mathcal{U}$, and using the fact that the sequence $\{\mathbf{m}_{\mathcal{S}}\}$ is of zero power [23, Theorem 1], the above equality reduces to

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k](\mathbf{m}_{\mathcal{U}}[k+1]-A\mathbf{m}_{\mathcal{U}}[k]\\ +B^M e_m[k])^T=0.\end{aligned}$$

Substituting (17) in the above equality and right-multiplying the resulting equality by the matrix $P_{\mathcal{U}}^T$ gives

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k](\mathbf{m}_{\mathcal{U}}[k+1]-A\mathbf{m}_{\mathcal{U}}[k]\\ +B^M d^{\mathcal{U}}[k]+B^M\bar{e}_m[k])_{\mathcal{U}}^T=0.\end{aligned}\tag{24}$$

Define

$$\bar{\mathbf{m}}_{\mathcal{U}}[k+1]:=\mathbf{m}_{\mathcal{U}}[k+1]-A\mathbf{m}_{\mathcal{U}}[k]+B^M d^{\mathcal{U}}[k].\tag{25}$$

Substituting (25) in (24) and simplifying gives

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k](\bar{\mathbf{m}}_{\mathcal{U}}[k+1]+B^M\bar{e}_m[k])_{\mathcal{U}}^T=0,$$

and using (23) in the above equality yields

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbf{e}^H[k]\bar{\mathbf{m}}_{\mathcal{U}}^T[k+1]=0.\tag{26}$$

It also follows from (4) that

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{m}[k+1]-A\mathbf{m}[k]+B^H\mathbf{e}^H[k]\\ +B^M e_m[k]+\mathbf{w}[k+1])_{\mathcal{NS}}^{TT}\\ =P_{\mathcal{NS}}(BB^T\sigma_e^2+\sigma_w^2 I)P_{\mathcal{NS}}^T.\end{aligned}$$

Resolving the vectors $\mathbf{m}[k+1]$ and $\mathbf{m}[k]$ along subspaces $\mathcal{S}$ and $\mathcal{U}$ and using the fact that $\{\mathbf{m}_{\mathcal{S}}\}$ is of zero power gives

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{m}_{\mathcal{U}}[k+1]-A\mathbf{m}_{\mathcal{U}}[k]+B^M d^{\mathcal{U}}[k]\\ +B^H\mathbf{e}^H[k]+B^M\bar{e}_m[k]+\mathbf{w}[k+1])_{\mathcal{NS}}^{TT}\\ =P_{\mathcal{NS}}(BB^T\sigma_e^2+\sigma_w^2 I)P_{\mathcal{NS}}^T.\end{aligned}$$

Substituting (25) in the above equality and simplifying gives

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\bar{\mathbf{m}}_{\mathcal{U}}[k+1]+B^H\mathbf{e}^H[k]\\ +B^M\bar{e}_m[k]+\mathbf{w}[k+1])_{\mathcal{NS}}^{TT}\\ =P_{\mathcal{NS}}(BB^T\sigma_e^2+\sigma_w^2 I)P_{\mathcal{NS}}^T.\end{aligned}$$

Expanding the LHS of the above equality and using (26), (22), (23) and (21) to simplify yields

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}^{TT}[k{+}1]+\bar{\mathbf{m}}_{\mathcal{NS}}[k+1]\bar{e}_m[k](B_{\mathcal{NS}}^M)^T\\ +\bar{\mathbf{m}}_{\mathcal{NS}}[k+1]\mathbf{w}_{\mathcal{NS}}^T[k+1]\\ +B_{\mathcal{NS}}^M\bar{e}_m[k]\bar{\mathbf{m}}_{\mathcal{NS}}^T[k+1]\\ +\mathbf{w}_{\mathcal{NS}}[k+1]\bar{\mathbf{m}}_{\mathcal{NS}}^T[k+1]=0.\end{aligned}\tag{27}$$

We have from (4) that

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\mathbf{m}[k+1]-A\mathbf{m}[k]+B^H\mathbf{e}^H[k]\\ +B^M e_m[k]+\mathbf{w}[k+1])_{\mathcal{NS}}\\ \times(\mathbf{m}[k+1]-A\mathbf{m}[k]+B^H\mathbf{e}^H[k]\\ +B^M e_m[k]+\mathbf{w}[k+1])_{\mathcal{S}}^T\\ =P_{\mathcal{NS}}(BB^T\sigma_e^2+\sigma_w^2 I)P_{\mathcal{S}}^T.\end{aligned}$$

Substituting (25) in the above equality and simplifying yields

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}(\bar{\mathbf{m}}_{\mathcal{U}}[k+1]+B^H\mathbf{e}^H[k]\\ +B^M\bar{e}_m[k]+\mathbf{w}[k+1])_{\mathcal{NS}}\\ \times(B^H\mathbf{e}^H[k]+B^M\bar{e}_m[k]+\mathbf{w}[k+1])_{\mathcal{S}}^T\\ =P_{\mathcal{NS}}(BB^T\sigma_e^2+\sigma_w^2 I)P_{\mathcal{S}}^T.\end{aligned}$$

Using (26), (23), (21), and (22) to further simplify the above equality yields

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1](B^M\bar{e}_m[k]+\mathbf{w}[k+1])_{\mathcal{S}}^T=0.\tag{28}$$

Note that the malicious sensors, in addition to adapting the measurements that they report at any time $k$ to all random variables whose realization they know by time $k$, can also introduce additional randomization. Let $\boldsymbol{\theta}[k]$ denote a random vector that is independent of all random variables that have been realized until time $k$ and to which the malicious sensors could adapt the measurements that they report at time $k$. Define $\sigma-$algebra $\mathcal{G}_k := \sigma(\mathbf{x}^k, \mathbf{e}^{H^{k-2}}, e_m^{k-1}, \boldsymbol{\theta}^k)$. Also define $\widehat{\mathbf{w}}[k+1] := \mathbb{E}(\mathbf{w}[k+1]|\mathcal{G}_{k+1})$, $\widehat{\mathbf{w}}_{\mathcal{R}(B^H)}[k+1] := \mathbb{E}(\mathbf{w}_{\mathcal{R}(B^H)}[k+1]|\mathcal{G}_{k+1})$, and $\widehat{\mathbf{w}}_{\mathcal{R}^\perp(B^H)}[k+1] := \mathbb{E}(\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1]|\mathcal{G}_{k+1})$. Let $B^H = QR$ be the QR-decomposition of the matrix $B^H$ so that the columns of $Q$

form an orthonormal basis for $\mathcal{R}(B^H)$ and $R$ is an upper-triangular matrix with full row rank. Then, $\mathbf{w}_{\mathcal{R}(B^H)}[k+1] = Q\mathbf{w}^c[k+1]$ for some vector $\mathbf{w}^c[k+1]$. Let $\widehat{\mathbf{w}}^c[k+1] := \mathbb{E}(\mathbf{w}^c[k+1]|\mathcal{G}_{k+1})$. It can be shown after some algebra that

$$\widehat{\mathbf{w}}^c[k+1] = K_W(R\mathbf{e}^H[k] + \mathbf{w}^c[k+1]) \tag{29}$$

where $K_W := \sigma_w^2(RR^T\sigma_e^2 + \sigma_w^2 I)^{-1}$. Define $\widetilde{\mathbf{w}}^c[k+1] := \mathbf{w}^c[k+1] - \widehat{\mathbf{w}}^c[k+1]$. Substituting for $\widehat{\mathbf{w}}^c[k+1]$ from (29) and rearranging the terms gives $\mathbf{w}^c[k+1] = (I - K_W)^{-1}K_W R\mathbf{e}^H[k] + (I-K_W)^{-1}\widetilde{\mathbf{w}}^c[k+1]$, where the existence of the matrix $(I-K_W)^{-1}$ follows from the fact that the matrix $R$ has full row rank. Consequently,

$$\begin{aligned}\mathbf{w}_{\mathcal{R}(B^H)}[k+1] &= Q\mathbf{w}^c[k+1] \\ &= Q(I-K_W)^{-1}K_W R\mathbf{e}^H[k] \\ &\quad + Q(I-K_W)^{-1}\widetilde{\mathbf{w}}^c[k+1].\end{aligned} \tag{30}$$

We have $\mathbf{w}[k+1] = \mathbf{w}_{\mathcal{R}(B^H)}[k+1] + \mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1]$. Substituting for $\mathbf{w}_{\mathcal{R}(B^H)}[k+1]$ from (30) yields

$$\begin{aligned}\mathbf{w}[k+1] = Q(I-K_W)^{-1}K_W R\mathbf{e}^H[k] \\ +Q(I-K_W)^{-1}\widetilde{\mathbf{w}}^c[k+1] \\ +\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1].\end{aligned} \tag{31}$$

Substituting (31) in (28) and using (26) to simplify yields

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1](B^M\bar{e}_m[k] \\ +Q(I-K_W)^{-1}\widetilde{\mathbf{w}}^c[k+1] \\ +\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1])_{\mathcal{S}}^T = 0.\end{aligned} \tag{32}$$

It is straightforward to verify that $(\widetilde{\mathbf{w}}^c[k], \mathcal{G}_{k+1})$ is a martingale difference sequence, and that $\bar{\mathbf{m}}_{\mathcal{NS}}[k+1] \in \mathcal{G}_{k+1}$. Using MST, we have

$$\begin{aligned}&\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1](\widetilde{\mathbf{w}}^c[k+1])^T = \\ &\begin{bmatrix} o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) \\ \vdots & \ddots & \vdots \\ o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) \end{bmatrix},\end{aligned} \tag{33}$$

and substituting this back in (32) yields

$$\begin{aligned}&\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1](B^M\bar{e}_m[k] + \mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1])_{\mathcal{S}}^T \\ &= \begin{bmatrix} o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) \\ \vdots & \ddots & \vdots \\ o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) \end{bmatrix} \\ &\quad + [o(T)],\end{aligned} \tag{34}$$

where $[o(T)]$ denotes a matrix all of whose entries are $o(T)$. Right multiplying (34) by $(P_{\mathcal{NS}}M_{\mathcal{S}\to\mathcal{R}(B^M)})^T$ and rearranging the terms gives

$$\begin{aligned}&\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1]\bar{e}_m[k](B^M)^T P_{\mathcal{NS}}^T \\ &= -\sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1]\mathbf{w}^T_{\mathcal{R}^\perp(B^H)}[k+1]P_S^T M^T_{\mathcal{S}\to\mathcal{R}(B^M)}P^T_{\mathcal{NS}} \\ &+ \begin{bmatrix} o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) \\ \vdots & \ddots & \vdots \\ o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) \end{bmatrix} \\ &+ [o(T)].\end{aligned} \tag{35}$$

Next, substituting (31) in (27) and simplifying the resulting equality using (26) and (33) yields

$$\begin{aligned}\sum_{k=0}^{T-1}\bar{\mathbf{m}}^{TT}_{\mathcal{NS}}[k+1] + \bar{\mathbf{m}}_{\mathcal{NS}}[k+1]\bar{e}_m[k](B^M_{\mathcal{NS}})^T \\ +\bar{\mathbf{m}}_{\mathcal{NS}}[k+1](\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1])^T_{\mathcal{NS}} \\ +B^M_{\mathcal{NS}}\bar{e}_m[k]\bar{\mathbf{m}}^T_{\mathcal{NS}}[k+1] \\ +(\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1])_{\mathcal{NS}}\bar{\mathbf{m}}^T_{\mathcal{NS}}[k+1] \\ = \begin{bmatrix} o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) \\ \vdots & \ddots & \vdots \\ o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) \end{bmatrix} \\ +[o(T)].\end{aligned} \tag{36}$$

For brevity of notation, define

$$\boldsymbol{\omega}[t+1] := M_{\mathcal{S}\to\mathcal{R}(B^M)}P_S\mathbf{w}_{\mathcal{R}^\perp(B^H)}[t+1].$$

Substituting (35) in (36) yields

$$\begin{aligned}&\sum_{k=0}^{T-1}\bar{\mathbf{m}}^{TT}_{\mathcal{NS}}[k+1] \\ &+ \sum_{k=0}^{T-1}\bar{\mathbf{m}}_{\mathcal{NS}}[t+1](P_{\mathcal{NS}}\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1] - P_{\mathcal{NS}}\boldsymbol{\omega}[t+1])^T \\ &+ \sum_{k=0}^{T-1}(P_{\mathcal{NS}}\mathbf{w}_{\mathcal{R}^\perp(B^H)}[k+1] - P_{\mathcal{NS}}\boldsymbol{\omega}[t+1])\bar{\mathbf{m}}^T_{\mathcal{NS}}[t+1] \\ &= \begin{bmatrix} o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_1}[k+1]) \\ \vdots & \ddots & \vdots \\ o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) & \dots & o(\sum_{k=0}^{T-1}\bar{m}^2_{\mathcal{NS}_p}[k+1]) \end{bmatrix} \\ &+ [o(T)].\end{aligned}$$

Since the subspace $\mathcal{NS}$ satisfies (8), only the first term is nonzero in the LHS of the above equality. Dividing the equality by $T$, taking the limit as $T \to \infty$, and equating the trace implies

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}||\bar{\mathbf{m}}_{\mathcal{NS}}[k+1]||^2 = 0. \tag{37}$$

We have from (25) that

$$\mathbf{m}_{\mathcal{U}}[k+1] = (A - B^M F)\mathbf{m}_{\mathcal{U}}[k] + \bar{\mathbf{m}}_{\mathcal{NS}}[k+1] \\ + \bar{\mathbf{m}}_{\mathcal{NU}}[k+1].$$

Let $\mathcal{V} := \mathcal{WU} \cap \mathcal{U}$. The above equality implies that

$$\mathbf{m}_{\mathcal{WS}_1}[k+1] = P_{\mathcal{WS}_1}(A - B^M F)\mathbf{m}_{\mathcal{WS}_1}[k] \\ + P_{\mathcal{WS}_1}(A - B^M F)\mathbf{m}_{\mathcal{V}}[k] \\ + P_{\mathcal{WS}_1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1] \\ + P_{\mathcal{WS}_1}\bar{\mathbf{m}}_{\mathcal{NU}}[k+1],$$

Invoking (11) implies that the second term in the RHS of the above equality is zero. Since we have $\mathcal{WS}_1 \subseteq \mathcal{NS}$ from (10), and $\mathcal{NU} \perp \mathcal{NS}$, the last term in the RHS of the above equality also vanishes. We therefore obtain

$$\mathbf{m}_{\mathcal{WS}_1}[k+1] = P_{\mathcal{WS}_1}(A - B^M F)\mathbf{m}_{\mathcal{WS}_1}[k] \\ + P_{\mathcal{WS}_1}\bar{\mathbf{m}}_{\mathcal{NS}}[k+1].$$

Invoking (37) and (12) yields

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{WS}_1}[k+1]||^2 = 0. \tag{38}$$

□

While the above result establishes that the state estimation error sequence $\{\mathbf{m}\}$, when projected on the watermark-securable subspace $\mathcal{WS}$, is of zero power, it can also be shown that for any subspace $\mathcal{N}$ of the watermark-unsecurable subspace $\mathcal{WU}$, there exists an attack such that $\lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} ||\mathbf{m}_{\mathcal{N}}[k+1]||^2 \neq 0$. The proof of this result will be reported in a subsequent paper in a more general context. The subspace $\mathcal{WS}$ is therefore the maximal subspace of the state space along which the projection of the state estimation error of the honest nodes is of zero power.

## V. Conclusion

In this paper, we have considered a perfectly-observed linear stochastic system containing an arbitrary set of malicious sensors and at most one malicious actuator. The honest actuators, if there are any, employ Dynamic Watermarking to detect whether or not there are malicious nodes present in the system. For such a system, we have characterized its watermark-securable subspace and have established its operational significance. Extensions of this work include characterizing the watermark-securable subspace for a system containing an arbitrary number of malicious actuators, and for partially-observed systems affected by both process and measurement noises.

## References

[1] K.-D. Kim and P. R. Kumar, "Cyber-Physical Systems: A Perspective at the Centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[2] D. Kushner, "The Real Story of Stuxnet," *IEEE Spectrum*, vol. 50, no. 3, pp. 48–53, March 2013.

[3] M. Abrams, "Malicious Control System Cyber Security Attack Case Study-Maroochy Water Services, Australia," 2008.

[4] Y. Mo and B. Sinopoli, "Secure Control Against Replay Attacks," in *47th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2009.

[5] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical Authentication of Control Systems: Designing Watermarked Control Inputs to Detect Counterfeit Sensor Outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, Feb 2015.

[6] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyberphysical systems," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 219–240, Feb 2017.

[7] ——, "On minimal tests of sensor veracity for dynamic watermarking-based defense of cyber-physical systems," in *Communication Systems and Networks (COMSNETS), 2017 9th International Conference on*. IEEE, 2017, pp. 23–30.

[8] ——, "Secure control of networked cyber-physical systems," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 283–289.

[9] ——, "Control systems under attack: The securable and unsecurable subspaces of a linear stochastic system," in *Emerging Applications of Control and Systems Theory*. Springer, 2018, pp. 217–228.

[10] ——, "The securable subspace of a linear stochastic system with malicious sensors and actuators," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 911–917.

[11] ——, "On the operational significance of the securable subspace for partially observed linear stochastic systems," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018, pp. 2068–2073.

[12] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack Detection and Identification in Cyber-Physical Systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[13] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure State-Estimation for Dynamical Systems under Active Adversaries," in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2011.

[14] 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 *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 2929–2933.

[15] A. Teixeira, I. Shames, H. Sandberg, and K. Johansson, "Revealing Stealthy Attacks in Control Systems," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2012.

[16] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Security in Stochastic Control Systems: Fundamental Limitations and Performance Bounds," *American Control Conference*, pp. 195–200, 2015.

[17] S. Weerakkody, Y. Mo, and B. Sinopoli, "Detecting Integrity Attacks on Control Systems using Robust Physical Watermarking," in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 3757–3764.

[18] P. Hespanhol, M. Porter, R. Vasudevan, and A. Aswani, "Dynamic watermarking for general LTI systems," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 1834–1839.

[19] M. Porter, A. Joshi, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, "Simulation and real-world evaluation of attack detection schemes," *arXiv preprint arXiv:1810.07773*, 2018.

[20] B. Satchidanandan and P. R. Kumar, "On the design of security-guaranteeing dynamic watermarks," *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 307–312, April 2020.

[21] W.-H. Ko, B. Satchidanandan, and P. R. Kumar, "Theory and implementation of dynamic watermarking for cybersecurity of advanced transportation systems," in *2016 IEEE Conference on Communications and Network Security (CNS)*, Oct 2016, pp. 416–420.

[22] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*. Springer-Verlag, 1985.

[23] B. Satchidanandan and P. R. Kumar, "On the watermark-securable subspace of a linear stochastic system," in *Proceedings of the sixth Indian Control Conference (ICC)*, to appear.

[24] T. L. Lai and C. Z. Wei, "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," *The Annals of Statistics*, pp. 154–166, 1982.