

Dynamic Watermarking: Active Defense of Networked Cyber–Physical Systems

This paper investigates the problem of secure control of networked cyber–physical systems.

By BHARADWAJ SATCHIDANANDAN AND P. R. KUMAR, Fellow IEEE

ABSTRACT | The coming decades may see the large scale deployment of networked cyber-physical systems to address global needs in areas such as energy, water, health care, and transportation. However, as recent events have shown, such systems are vulnerable to cyber attacks. Being safety critical, their disruption or misbehavior can cause economic losses or injuries and loss of life. It is therefore important to secure such networked cyber-physical systems against attacks. In the absence of credible security guarantees, there will be resistance to the proliferation of cyber-physical systems, which are much needed to meet global needs in critical infrastructures and services. This paper addresses the problem of secure control of networked cyber-physical systems. This problem is different from the problem of securing the communication network, since cyber-physical systems at their very essence need sensors and actuators that interface with the physical plant, and malicious agents may tamper with sensors or actuators, as recent attacks have shown. We consider physical plants that are being controlled by multiple actuators and sensors communicating over a network, where some sensors could be “malicious,” meaning that they may not report the measurements that they observe. We address a general technique by which the actuators can detect the actions of malicious sensors in the system and disable closed-loop control based on their information. This

technique, called “watermarking,” employs the technique of actuators injecting private excitation into the system, which will reveal malicious tampering with signals. We show how such an active defense can be used to secure networked systems of sensors and actuators.

KEYWORDS | Cyber-Physical Systems (CPS); dynamic watermarking; networked control systems; networked cyber-physical systems; secure control

I. INTRODUCTION

The 21st century could well be the era of large-scale system building. Such large-scale systems are envisioned to be formed by the interconnection of many embedded devices communicating with each other and interacting with the physical world. Their operation requires tight integration of communication, control, and computation, and they have been termed broadly as Cyber–Physical Systems (CPS). The smart energy grid, intelligent transportation systems, Internet of Things, telesurgical systems, and robotics are examples of such cyber–physical systems.

Although the importance and benefits of cyber–physical systems require no emphasis, their sustained proliferation is contingent on some key challenges being addressed of which security is a primary one. Since CPSs have many applications in safety-critical scenarios, security breaches of these systems can have adverse consequences, including economic loss, injury, and death.

There have been many instances of demonstrated attacks on cyber–physical systems in the recent past [1], [2]. In 2003, in Maroochy-Shire, Australia, a disgruntled ex-employee of a sewage treatment corporation hacked into the computers controlling the sewage system and issued commands that led to a series of faults in the system [1], [3]. This is an insider attack whereby the

Manuscript received January 10, 2016; accepted May 23, 2016. Date of publication November 8, 2016; date of current version January 18, 2017. This paper is partially based on work supported by the National Science Foundation (NSF) Science and Technology Center under Grant CCF-0939370, NSF Contract No. CNS-1302182, by the US Army Research Office under Contract No. W911NF-15-1-0279, and by the Air Force Office of Scientific Research under Contract No. FA-9550-13-1-0008.

The authors are with Texas A&M University, Electrical and Computer Engineering, College Station, TX 77843 USA (e-mail: bharadwaj.s1990@tamu.edu).

Digital Object Identifier: 10.1109/JPROC.2016.2575064

0018-9219 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

adversary has the necessary credentials to access and issue control commands to the system. We will return to this point shortly. Another example is the attack on computers controlling the Davis-Besse nuclear power plant in Ohio, USA. In the year 2003, the Slammer worm, which infected about 75000 hosts in the Internet in under ten minutes, infected the computers controlling the nuclear power plant, disabling the safety monitoring systems [1]. While the Slammer worm was not designed to target the nuclear power plant, the use of commodity IT software in control systems made them vulnerable to such attacks [1]. Another pertinent example is the Stuxnet worm, which in 2010 exploited a vulnerability in Microsoft Windows to subvert critical computers controlling centrifuges in Iran's uranium enrichment facility [4]. Having subverted the computers, it issued control commands that caused the centrifuges to operate at abnormally high speeds, causing them to tear themselves apart. To keep the attacks undetected by software-implemented alarm routines and officials in the control room, Stuxnet recorded the sensor values in the facility for twenty-one seconds before carrying out each attack, and replayed those twenty-one seconds in a constant loop during the attack. Stuxnet has been claimed to be the first known digital weapon [4], and since then, cyberwarfare has emerged as a serious concern for cyber-physical systems due to the many advantages it offers to the attacker, such as allowing it to remain anonymous, attack without geographical constraints, etc. Today, the resources required to carry out such attacks on critical infrastructures are generally available [5], underlining the urgent need for the research community to pay attention to this problem.

In this paper, we examine the problem of detecting attacks on networked cyber-physical systems. These systems can be thought of as having two layers: a physical layer, which consists of the plant, actuators, controllers, and sensors, that interact with physical signals and a cyber layer, which networks the components of the physical layer. While securing the cyber layer is certainly of importance, by itself, it does not constitute the security of the cyber-physical system as a whole. The Maroochy-Shire incident is a classic illustration of this point in which the malfunctioning of the plant was not the result of an attack on the network layer but of authorized individuals attacking the physical layer by issuing improper control commands.

At first look, it appears as though securing the physical layer is more difficult than securing the cyber layer. In the problem of network security, there is a clear distinction between an honest party and an adversary. Attributes, such as credentials and cryptographic keys, distinguish honest parties from adversaries. However, when it comes to securing the physical layer, no such demarcation exists. Any authorized party is also a potential adversary.

However, we show in this paper that what works in favor of securing the physical layer, and what we exploit, is the fact that the actions of every node interfacing with the physical layer get transformed into physical signals, and these signals can be subject to scrutiny for semantic consistency. To elaborate, consider a physical system consisting of a plant, some actuators, and some sensors. If an actuator injects into the system a probing signal that is not disclosed to other nodes in the system, then, combined with the knowledge of the plant dynamics, the actuator expects the signals to appear in transformed ways at various points in the system. Based on the information that the actuator receives from the sensors about the signals at various points, it can potentially infer whether there is malicious activity in the system.

We develop these ideas to secure the physical layer of a noisy dynamical system. We examine a protocol whereby honest actuator nodes deliberately superimpose certain stochastically independent probing signals on top of the control law they are intended to apply. We then propose specific "tests" that these actuator nodes perform to infer malicious activity, and establish their effectiveness. As a particular illustration of the results, for example, we can show that even if all the sensors are malicious, and the actuators have absolutely no measurements that they can directly make and have to completely rely on the malicious sensors for *all* their purported measurements, then even in such an adverse environment, the actuators can, under appropriate conditions, ensure that the additional distortion on performance that the malicious sensors can cause is of mean-square zero, if they are to remain undetected. Using this approach of active defense, we establish that under appropriate conditions, no matter in what way the adversarial sensors collude, the amount of distortion that they can add without exposing their presence can have an average power of only zero.

The method we examine is a dynamic version of "watermarking," [6] whereby certain indelible patterns are imprinted into a medium that can detect tampering [6]–[8]. It shows how one can watermark dynamic signals so that one can detect malicious misbehavior on the part of sensors or actuators. On top of a secure communication system, it provides overall security to a cyber-physical system against malicious sensors and actuators.

This paper is organized as follows. Section II describes prior work in this area. Section III provides a system-theoretic formulation of the problem. Section IV describes our approach of active defense for networked cyber-physical systems. Section V opens by describing the method in the relatively simple context of a scalar linear Gaussian system and rigorously establishes the associated theoretical guarantees. Section VI treats the more general class of scalar auto-regressive systems with exogenous noise (ARX systems) that is Gaussian. Section VII extends these ideas to the more general

ARMAX systems with arbitrary delay, a model that is frequently encountered in process control. Section VIII deals with partially observed SISO systems with Gaussian process and measurement noise. Section IX considers multi-input, multi-output linear Gaussian systems in state-space form. Section X describes how our results can potentially be extended to non-Gaussian systems. Section XI shows how the theoretical results lead to statistical tests that can be used to detect malicious behavior within a delay bound with a controlled false alarm rate. Section XII provides our concluding remarks.

II. PRIOR WORK

The vulnerability and the need to secure critical infrastructures from cyberattacks has been recognized at least as early as in 1997 [9]. Subsequent reports [10]–[15] cited demonstrated attacks, identifying potential threats, and analyzing the effects of successful attacks on specific systems. The large-scale replacement of proprietary control software and protocols by commodity IT software and protocols by the industry to allow for interoperability and rapid scalability has increased the vulnerability of Industrial Control Systems (ICS) to cyberattacks, and roadmaps were prepared to address security of control systems in various sectors, such as the energy sector [16], water sector [17], chemical sector [18], and the transportation sector [19].

Some of the initial work on secure control [20] has addressed the definition of what constitutes a secure control system. Certain key operational goals such as closed-loop stability and performance metric of interest are noted in [20], and it is proposed that a secure control system must achieve these operational goals even when under attack, or at least cause only a gradual degradation. It also identifies how the problem of secure control of networked systems departs from the traditional problems of network security and information security. In the former, authorized users or insiders can launch attacks on the system causing physical damage, as in the Maroochy-Shire incident. Hence, network and information security measures, such as intrusion prevention and detection, authentication, access control, etc., fundamentally cannot address these attacks. Therefore, securing the network does not amount to securing the NCS. In this paper, we build a framework on top of a secure communication network to secure the NCS. There has been recent work showing how one indeed can build a communication network that provides provable guarantees on security, throughput, as well as delays [21].

A theoretical study of secure control benefits from having a model for the adversary, and [22] defines certain adversary and attack models. A popular attack on communication networks is the Denial-of-Service (DoS) attack in which the adversary floods the communication network with useless packets, rendering it incapable of

transporting useful information. Another attack is the deception attack whereby an adversary impersonates as another node and transmits false information on its behalf. In the context of a networked control system, the adversary could employ the DoS attack to prevent the controller and actuator from receiving the data required for their operation, and could employ deception to cause an actuator node to issue incorrect actuation signals. A framework to study the evolution of the physical process under Denial-Of-Service (DoS) and deception attacks is presented in [22].

Classical estimation algorithms, such as the Kalman filter, assume perfect communication between the sensor and the controller. However, in the presence of an unreliable network, which may stem from either the characteristics of the network or from adversarial presence, certain packets may be dropped. This affects the performance of the estimation and control algorithm. Motivated by this scenario, a large body of research has been devoted to developing estimation and control algorithms for systems with intermittent observations, c.f. [23]–[27] and references therein. Though this line of work does not explicitly model adversarial behavior, some of these ideas have been employed in the literature of secure control. For instance, [28] studies, using some of the machinery developed in [23] in the context of control over lossy communication networks, the effect of DoS attack on the control performance. Also addressed in the literature is the effect of deception attacks on estimation and control performance. In [29], the effect of false data fed by the compromised sensors to the state estimator is studied. The goal is to characterize the set of all estimation biases that an adversary can inject without being identified. A related work is [30] in which fundamental trade-offs between the detection probability (for a fixed false alarm rate) and the estimation error that the adversary can cause, are studied for single-input-single-output systems.

Several techniques have also been developed to counter attacks on cyber-physical systems. A technique for correct recovery of the state estimate in the presence of malicious sensors is presented in [31], and also a characterization of the number of malicious sensors that can be tolerated by the algorithm. A well-known attack on control systems is the replay attack whereby the adversary records the sensor measurements for a fixed period of time and replays them during the attack so as to maintain the illusion of a normal operating condition. It was shown in [32] that only systems for which the matrix $(A + BL)(I - KC)$ is stable are susceptible to replay attacks, where L is the feedback gain and K is the steady-state Kalman gain of the system's Kalman filter. Consequently, for such systems, a method to secure the system from replay attack is presented in [32]–[35]. The fundamental idea of these is to inject into the actuation signal a component that is not known in advance. Specifically, [32], [34] consider the replay attack employed in Stuxnet and introduce a technique, termed Physical

Watermarking, wherein the controller commands the actuators to inject into the system a component that is random and not known in advance to secure the system against such an attack. To the best of our knowledge, this is the first use of the idea of watermarking. It is shown that by employing Physical Watermarking, the covariance of the innovations process when the system is “healthy” and that when it is under attack are significantly different, enabling the estimator to detect the attack using a χ^2 detector. This technique is extended in [35] to detect an adversary employing more intelligent attack strategies. Specifically, the adversary is assumed to possess a set of capabilities, based on which a specific attack strategy, consisting of the adversary generating false measurement values that are reported to the estimator, is identified. It is shown that Physical Watermarking can counter such an adversary. Including a random component in the actuation signal would clearly affect the running cost, and [36] develops an optimal policy to switch between cost-centric and security-centric controllers by formulating the problem as a stochastic game between the system and the adversary. A method to detect false-data injection is presented in [37] wherein the focus is on zero-dynamics attacks, attacks which cannot be detected based on input and output measurements. A method to verify the measurements received from multiple sensors is presented in [38], which exploits correlations between sensor measurements and other features to weed out the measurements from malicious sensors. Though not presented in the context of a dynamical system, the ideas presented in [38] could in principle be extended to incorporate system dynamics.

At a high level, the techniques generally proposed in the literature for secure control, such as estimation with intermittent or incorrect observations, exploiting known correlations between sensor measurements to weed out suspicious measurements or techniques inspired from fault-tolerant control, can be classified as passive techniques, meaning that they anticipate that the adversary will inject malicious signals into the system and therefore employ a design that minimizes the damage that can be so caused. On the contrary, in this paper, we pursue active defense along the lines of [32], [34], [35], an alternative approach for secure control in which, over and above the control-policy specified excitation, the system is excited in ways unknown to the adversary, thereby preventing the adversary from injecting malicious signals into the system. This constitutes a form of watermarking of a dynamic system [6]–[8]. Since it does not distinguish between an adversarial node and a faulty node, this approach also falls under the purview of fault-tolerant control. Such a singular departure from passive approach is [32], [34], [35], where it has been proposed to inject a signal to guard against adversarial attacks. This paper develops a comprehensive treatment of such dynamic watermarking. To the best of our

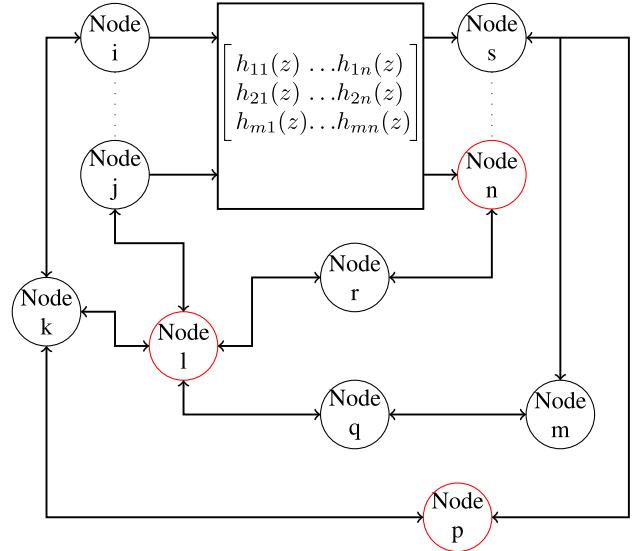


Fig. 1. A networked cyber-physical system.

knowledge, the general theory and techniques of active defense for *arbitrary* attacks does not appear to have been studied in the literature thus far. That is what is developed in this paper.

III. PROBLEM FORMULATION

Fig. 1 illustrates the basic architecture of a networked cyber-physical system. At the heart of the system is a physical plant with m inputs and n outputs. Each input is controlled by an independent actuator, and each output is measured by an independent sensor. A known transformation capturing the dynamics of the physical plant maps the actuation signals applied by the actuators to outputs that are measured by the sensors. These measurements are communicated to entities called controllers through an underlying communication network. Each controller’s job is to compute, in accordance with a control policy, the particular actuation signals that must be applied. The result of this computation is then communicated to the actuators through the communication network, which then apply the actuation signals. We assume that the network is complete so that every node in the network can communicate with every other node. We use the term “nodes” generically to refer to any entity in the network. Therefore, in Fig. 1, some nodes are actuators, some are sensors, some are controllers, and some could just be relays whose only job is to forward the information from one node to another. In this work, we assume that each node has both communication and computational capabilities, thereby allowing the controllers to be collocated with the actuators.

However, certain sensors in the system could be “malicious” (suggestive of this, some nodes in Fig. 1

are marked in red), and the other nodes are said to be “honest.” We further assume that the malicious nodes know the identity of all other malicious nodes in the system, allowing them to collude to achieve their objective, whereas the honest nodes don’t know which of the other nodes are malicious or honest. A malicious sensor is a sensor node that does not report accurately the measurements that it observes. Rather, it reports a distorted version of the measurements. A malicious router node may not forward the packets that it receives, may forward packets that it does not receive (while claiming otherwise), alter packets before forwarding, introduce intentional delays, impersonate some other node in the system, etc. Therefore, the challenges of securing a networked cyber–physical system are two-fold.

- 1) Secure the cyber layer that comprises the communication network, ensuring confidentiality, integrity, and availability of network packets.
- 2) Secure the sensors and actuators interfacing with the physical layer.

The former is achieved by a combination of traditional approaches, such as cryptography, and a more recent line of work reported in [21], [39], and [40], whereas the latter is the subject of this paper.

Based on our assumption that the communication network is complete, and that the cyber layer has been secured, we suppose that every node in the network knows the identity of the node from which a packet that it receives originated, and knows whether a packet that it receives was tampered with by any node along the route. Therefore, going forward, we abstract the cyber layer as consisting of secure, reliable, delay-guaranteed bit pipes between any pair of nodes in the network. In particular, we imagine that there exists a secure, reliable, delay-guaranteed bit pipe between any particular sensor and actuator node.

Finally, the plant that is being controlled is abstracted as a stochastic linear dynamical system described by time-invariant parameters. Although any system that is of practical interest is most certainly nonlinear, we focus on linear systems for two reasons. The first is that a linear system lends itself to tractable analysis, and enables one to separate the complexity arising out of the problem at hand from the complexity arising as a consequence of the system’s nonlinearity. Secondly, a theory developed for linear systems provides valuable insights and design principles that often transcend the particulars of the model and apply to a much broader class of systems. The wide applicability of Kalman’s pioneering work on linear control systems [41] stands testimony to this fact.

We are now in a position to state the problem in precise terms (with notation as indicated in the Appendix). Consider an $m \times n$ stochastic linear dynamical system of order p , described by

$$\mathbf{x}[t+1] = \mathbf{A}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \mathbf{w}[t+1] \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{B} \in \mathbb{R}^{p \times m}$, $\mathbf{u}[t]$ is the input applied to the plant at time t , $\{\mathbf{w}\}$ is a sequence of independent and identically distributed (i.i.d.) Gaussian random vectors with zero mean and covariance matrix $\sigma_w^2 \mathbf{I}$, independent of the initial state \mathbf{x}_0 of the system.

While a malicious sensor can report different measurements to different actuators, the consistency of the reported measurements can be checked by allowing the actuators to exchange the reported measurements among themselves. This constrains the malicious sensors to report the same value to all honest nodes in the system. The sensor node j reports to the honest nodes in the system the value $\mathbf{z}_j[t]$ as the measurement that it observes at time t . We define $\mathbf{z}[t] := [\mathbf{z}_1[t] \ \mathbf{z}_2[t] \ \mathbf{z}_3[t] \cdots \mathbf{z}_n[t]]^T$. We will call $\mathbf{z}[t]$ for $t \geq 0$ the measurements reported by the sensors. Note that the sensor j is honest if $\mathbf{z}_j[t] = \mathbf{x}_j[t] \ \forall t$.

We assume that a control policy is in place, known to all nodes in the system, and allow for it to be history dependent, so that the i th input at time t , $u_i^g[t]$, dictated by the policy is

$$u_i^g[t] = g_t^i(\mathbf{z}^t) \quad (2)$$

where $\mathbf{z}^t := \{\mathbf{z}[0], \mathbf{z}[1], \dots, \mathbf{z}[t]\}$. We suppose that the controller that computes this control law is collocated with the actuator node.

Our goal in this paper is to secure the control system by developing techniques that prevent the malicious nodes from causing excessive distortion if they are to remain undetected. We will suppose that the purpose of control law (2) is to improve the regulation performance of the system with respect to the disturbances affecting it. For some of the performance results, we will assume that the system is open-loop stable and show that the adversarial nodes cannot affect the performance of the system without remaining undetected, and, if detected, then they can be disconnected, returning the system to stable behavior. The fundamental results characterizing what is the most that the adversarial nodes can do while remaining undetected, are applicable to all systems, stable or unstable.

IV. DYNAMIC WATERMARKING: AN ACTIVE DEFENSE FOR NETWORKED CYBER–PHYSICAL SYSTEMS

The key idea that allows the honest nodes to detect the presence of malicious nodes in the system is the following. Let g denote the control policy in place that specifies inputs to be applied in response to observed outputs. At each time instant t , an actuator node superimposes on its control policy-specified input $u_i^g[t]$, a random variable $e_i[t]$ that it draws independently from a

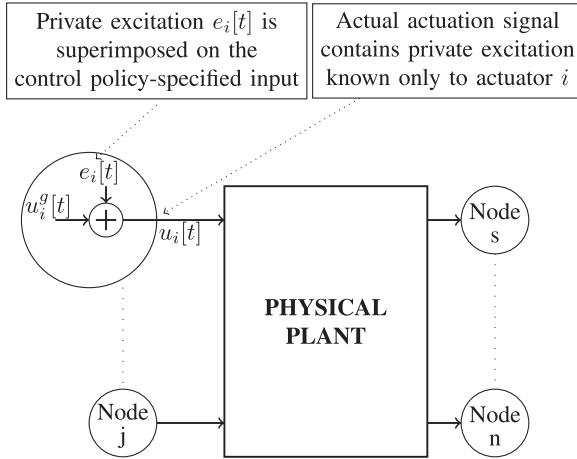


Fig. 2. The actuator node i superimposes a private excitation, whose realization is unknown to other nodes, onto its control inputs.

specified distribution. Therefore, the input that actuator i applies at time t is

$$u_i[t] = u_i^g[t] + e_i[t]. \quad (3)$$

This is illustrated in Fig. 2. The random variables $e_0[t], e_1[t], e_2[t], \dots$ are independent and identically distributed (i.i.d.), independent of the control policy-specified input. The distribution that they are chosen from is made public (i.e., made known to every node in the system), but the actual values of the excitation are not disclosed. In fact, this is how the honest actuators can check whether signals are being tampered with as they travel around the control loop. We refer to these random variables as an actuator node's *privately imposed excitation*, since only that actuator node knows the actual realization of the sequence.

To see why private excitation helps, and sometimes even necessary, consider the example of a single-input-single-output (SISO) system whereby the sensor is malicious and the actuator is honest. Suppose that the control policy in place is $g = (g_1, g_2, \dots, g_t, \dots)$, where g_t specifies the input to be applied at time t in response to outputs up till that time. The actual outputs of the plant up to time t are $x^t := (x[0], x[1], \dots, x[t])$. However, the outputs reported to it by the malicious sensor are $z^t := (z[0], z[1], \dots, z[t])$, which may differ from x^t . The actuator therefore applies the control input $u[t] = g_t(z^t)$ at time t , without using any private excitation. Then, we have the system

$$x[t+1] = ax[t] + bg_t(z^t) + w[t+1] \quad (4)$$

where $w[t]$ is the zero-mean process noise with variance σ_w^2 .

If the actuator does not superimpose a private excitation, the sensor knows, for each time t , the input $u[t]$ applied by the actuator because it knows both the measurement sequence z^t that it reported to the actuator as well as the control policy $\{g_1, g_2, \dots\}$ that the honest actuator has implemented. Hence, the malicious sensor can report a sequence of measurements $\{z[t]\}$ to the actuator without even "looking" at the output but by simply "simulating a linear stochastic system" after generating its own i.i.d. process noise $\{w'\}$ from the same distribution as that of $\{w\}$, as follows:

$$z(t+1) = az(t) + bg_t(z^t) + w'(t+1). \quad (5)$$

The actuator cannot detect that the sensor is malicious since the sequence $\{w'\}$, having been chosen from the same distribution as $\{w\}$, could have been the actual process noise.

However, by superimposing a private excitation that is *unknown* to the sensor, the actuator forces the sensor to report measurements that are correlated with $\{e_i\}$, lest it be exposed, as we will show in the sequel. In the following sections, we prove that by constraining the sensor to report measurements that are correlated with the private excitation essentially limits the amount of distortion that the sensor can get away with while remaining undetected to be essentially zero in a mean-square sense.

This active defense technique is similar in spirit to the technique of digital watermarking [6] in electronic documents. Electronic documents can be easily transmitted and reproduced in large numbers. In doing so, the source of the document may be deleted and can result in copyright violations. To protect the identity of its author, or any other information about the document, the electronic document is "watermarked" before being made available electronically. A digital watermark is a digital code that is robustly embedded in the original document [7]. By robust, it is meant that the code cannot be destroyed without destroying the contents of the document. This code typically contains information about the document that needs to be preserved. Though preferable, it is not a requirement that the watermark be imperceptible. The only requirement is that it does not distort the actual contents beyond certain acceptable limits [7]. Applications of digital watermarking also include data authentication whereby fragile watermarks are used that get destroyed when the data is tampered with [7], and data monitoring and tracking.

This approach for secure control is analogous to digital watermarking. As we show in the subsequent sections, with regard to, for example, the above case in

which we considered a compromised sensor, injecting private excitation that is unknown to the sensor effectively “watermarks” the process noise, in the sense that the sensor cannot separate the private excitation from the process noise. Hence, any attempt on the part of the sensor to distort the process noise (which is the only component of the output unknown to the actuators) will also distort the “watermark,” allowing the honest nodes to detect malicious activity.

Specifically, given a general MIMO system of the form (1), we define

$$\mathbf{v}[k] := \mathbf{z}[k] - A\mathbf{z}[k-1] - B\mathbf{u}[k-1] - \mathbf{w}[k]$$

so that if the sensors truthfully report $\mathbf{z}[t] \equiv \mathbf{y}[t]$, then $\mathbf{v} \equiv 0$. As will be shown, the sequence $\{\mathbf{v}\}$ has the interpretation of the additive distortion introduced by the malicious sensors to the process noise. A similar definition can also be provided for a partially observed system, as we will show in Section VIII. In this case, the sequence $\{\mathbf{v}\}$ has the interpretation of the additive distortion introduced by the malicious sensors to the innovations process. Now, based on $\{\mathbf{v}\}$, we define the following quantity.

Definition: (Additive Distortion Power of Malicious Sensors): We call

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \|\mathbf{v}[k]\|^2$$

as the *additive distortion power* of the malicious sensors.

The fundamental security guarantee provided by Dynamic Watermarking is that the additive distortion power is restricted to be zero if the malicious sensors are to remain undetected. We establish this in several linear control system contexts in this paper.

V. ACTIVE DEFENSE FOR NETWORKED CYBER-PHYSICAL SYSTEMS: THE SISO CASE WITH GAUSSIAN NOISE

In this section, to illustrate the results in a simple context, we focus on single-input, single-output linear stochastic dynamical systems with Gaussian noise. The system is described by

$$\mathbf{x}[t+1] = a\mathbf{x}[t] + b\mathbf{u}[t] + \mathbf{w}[t+1] \quad (6)$$

where $a, b, \mathbf{x}[t], \mathbf{u}[t], \mathbf{w}[t] \in \mathbb{R}$, with $\{\mathbf{w}[t]\}$ being zero-mean i.i.d. Gaussian process noise of variance σ_w^2 . The

actuator wishes to implement a control law $\{g_t\}$, i.e., it wishes to implement $\mathbf{u}[t] = g_t(\mathbf{x}^t)$, where $\mathbf{x}^t := (\mathbf{x}[0], \mathbf{x}[1], \dots, \mathbf{x}[t])$. However, the actuator does not have access to \mathbf{x}^t . It relies on a sensor that measures $\mathbf{x}[t]$. However, since the sensor could be malicious, it reports measurements $\mathbf{z}[t]$ to the actuator, where $\mathbf{z}[t]$ could differ from $\mathbf{x}[t]$. We suppose that the actuator is *honest*. It is meant to implement the control policy g and add a private excitation $\{e\}$ as a defense, which it does. Specifically, the actuator applies to the system the input

$$\mathbf{u}[t] = g_t(\mathbf{z}^t) + e[t]. \quad (7)$$

Note that even though it implements the control policy $\{g\}$, the policy is applied to the measurements $\mathbf{z}[t]$ reported by the sensor, which could differ from the true output $\mathbf{x}[t]$. The private excitation $e[\cdot]$ added is independent and identically distributed (i.i.d.) and Gaussian of mean 0 and variance σ_e^2 . Therefore, the system evolves in closed-loop as

$$\mathbf{x}[t+1] = a\mathbf{x}[t] + b g_t(\mathbf{z}^t) + b e[t] + \mathbf{w}[t+1]. \quad (8)$$

We propose that the honest actuator performs certain “tests” to check whether the sensor is malicious. Toward developing these tests, note that the actual sequence of states $\{\mathbf{x}[t]\}$ of the system satisfies

$$\mathbf{x}[t+1] - a\mathbf{x}[t] - b g_t(\mathbf{z}^t) = b e[t] + \mathbf{w}[t+1]. \quad (9)$$

Therefore, we have

$$\begin{aligned} \{\mathbf{x}[t+1] - a\mathbf{x}[t] - b g_t(\mathbf{z}^t)\}_t &\sim \\ \text{i.i.d. } \mathcal{N}(0, b^2 \sigma_e^2 + \sigma_w^2) \end{aligned} \quad (10)$$

$$\begin{aligned} \{\mathbf{x}[t+1] - a\mathbf{x}[t] - b g_t(\mathbf{z}^t) - b e[t]\}_t &\sim \\ \text{i.i.d. } \mathcal{N}(0, \sigma_w^2). \end{aligned} \quad (11)$$

Based on the above observations, we propose that the actuator perform the following natural tests for variance to check whether the sensor is honestly reporting $\mathbf{x}[t]$. The actuator checks whether the reported sequence $\{\mathbf{z}[t]\}$ satisfies conditions (10) and (11), which the true output $\{\mathbf{x}[t]\}$ would satisfy if the sensor were truthfully reporting $\mathbf{z}[t] \equiv \mathbf{x}[t]$. We write the tests in an asymptotic form below, as a test conducted over an infinite time interval. They can be reduced to statistical tests over a

finite time interval in standard ways, which we elaborate on in Section XI.

- 1) **Actuator Test 1:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left(z[k+1] - az[k] - bg_k(z^k) - be[k] \right)^2 = \sigma_w^2. \quad (12)$$

- 2) **Actuator Test 2:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} & \left(z[k+1] - az[k] - bg_k(z^k) \right)^2 \\ & = \left(b^2 \sigma_e^2 + \sigma_w^2 \right). \quad (13) \end{aligned}$$

Define

$$v[t+1] := z[t+1] - az[t] - bg_t(z^t) - be[t] - w[t+1]$$

so that for an honest sensor that reports $z[t] \equiv x[t]$, $v[t] = 0 \forall t$. We term the quantity

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k]$$

the *additive distortion power* of a malicious sensor for reasons explained later. The ensuing theorem proves that a malicious sensor with only zero additive distortion power can pass the above two tests, thereby remaining undetected.

Theorem 1: If $\{z[t]\}$ passes tests (12) and (13), thereby remaining undetected, then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] = 0. \quad (14)$$

Proof: Since $\{z\}$ satisfies (12), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T (v[k] + w[k])^2 = \sigma_w^2. \quad (15)$$

Hence,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] + 2v[k]w[k] + w^2[k] = \sigma_w^2.$$

Since $\lim_{T \rightarrow \infty} (1/T) \sum_{k=1}^T w^2[k] = E\{w^2[k]\} = \sigma_w^2$, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T 2v[k]w[k] = 0. \quad (16)$$

Since $\{z\}$ also satisfies (13), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T (v[k] + be[k-1] + w[k])^2 = b^2 \sigma_e^2 + \sigma_w^2.$$

So,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T (v[k] + w[k])^2 & + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T b^2 e^2[k-1] \\ & + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T 2be[k-1](v[k] + w[k]) = b^2 \sigma_e^2 + \sigma_w^2. \end{aligned}$$

Using (15), and the fact that $\{e\}$ has variance σ_e^2 , the above reduces to

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T e[k-1](v[k] + w[k]) = 0.$$

Invoking the fact that $e[k-1]$ and $w[k]$ are independent $\forall k$ and zero mean, it further reduces to

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T e[k-1]v[k] = 0. \quad (17)$$

The above equation implies that the sequence $\{v\}$, added by the sensor, must be empirically uncorrelated with the actuator's private noise sequence $\{e\}$.

Let $\mathcal{S}_k := \sigma(x^k, z^k, e^{k-2})$, $\hat{w}[k] := E[w[k]|\mathcal{S}_k]$. Since

$$w[k] = x[k] - ax[k-1] - bg_{k-1}(z^{k-1}) - be[k-1]$$

we have

$$(x^{k-2}, e^{k-2}) \rightarrow (x[k-1], x[k], z^k) \rightarrow w[k]$$

forming a Markov chain. Consequently, $\hat{w}[k] := E[w[k]|\sigma(e^{k-2}, x^{k-2}, x[k-1], x[k], z^k)] = E[w[k]|\sigma(x[k-1], x[k], z^k)]$. Since $x[k] - ax[k-1] - bg_{k-1}(z^{k-1})$ (which is equal to $be[k-1] + w[k]$) is i.i.d. Gaussian for different k , we have [42]

$$\begin{aligned} \hat{w}[k] &= \frac{\sigma_w^2}{b^2\sigma_e^2 + \sigma_w^2} (be[k-1] + w[k]) \\ &= \beta(be[k-1] + w[k]) \end{aligned} \quad (18)$$

where $\beta := (\sigma_w^2/(b^2\sigma_e^2 + \sigma_w^2)) < 1$.

Let $\tilde{w}[k] := w[k] - \hat{w}[k]$. Then, $(\tilde{w}[k-1], \mathcal{S}_k)$ is a Martingale difference sequence. This is because $\tilde{w}[k-1] \in \mathcal{S}_k$, and

$$E[\tilde{w}[k] \mid \mathcal{S}_k] = 0. \quad (19)$$

We also have $v[k] \in \mathcal{S}_k$ (in fact, $v[k] \in \sigma(x^k, z^k)$). Hence, Martingale Stability Theorem (MST) [43] applies, and we have

$$\sum_{k=1}^T v[k]\tilde{w}[k] = o\left(\sum_{k=1}^T v^2[k]\right) + O(1). \quad (20)$$

Now

$$\begin{aligned} \sum_{k=1}^T v[k]w[k] &= \sum_{k=1}^T v[k](\hat{w}[k] + \tilde{w}[k]) \\ &= \sum_{k=1}^T v[k]\hat{w}[k] + o\left(\sum_{k=1}^T v^2[k]\right) + O(1). \end{aligned}$$

Employing the specific form of the estimate (18), we have from the above

$$\begin{aligned} \sum_{k=1}^T v[k]w[k] &= \beta b \sum_{k=1}^T v[k]e[k-1] + \beta \sum_{k=1}^T v[k]w[k] \\ &\quad + o\left(\sum_{k=1}^T v^2[k]\right) + O(1). \end{aligned}$$

Hence,

$$\sum_{k=1}^T v[k]w[k] = \frac{\beta b}{1-\beta} \sum_{k=1}^T v[k]e[k-1] + o\left(\sum_{k=1}^T v^2[k]\right) + O(1).$$

From (17), we have $\sum_{k=1}^T v[k]e[k-1] = o(T)$. It follows that

$$\sum_{k=1}^T v[k]w[k] = o\left(\sum_{k=1}^T v^2[k]\right) + o(T) + O(1). \quad (21)$$

So

$$\begin{aligned} \sum_{k=1}^T v^2[k] + \sum_{k=1}^T 2v[k]w[k] &= (1 + o(1)) \left(\sum_{k=1}^T v^2[k] \right) \\ &\quad + o(T) + O(1). \end{aligned}$$

Dividing the above equation by T , taking the limit as $T \rightarrow \infty$, and invoking (16) completes the proof. \square

Remark: Note that the only sources of uncertainty in the system are the initial state of the system $x[0]$ and the sequence of noise realizations $\{w[1], w[2], w[3], \dots\}$. The sensor reporting a sequence of measurements is equivalent to it reporting a sequence of process noise realizations, since the actuator expects $z[t+1] - az[t] - bg_t(z^t) - be[t]$, which it can compute, to be equal to the process noise $w[t+1]$. From the definition of $v[t]$, we have

$$z[t+1] - az[t] - bg_t(z^t) - be[t] = w[t+1] + v[t+1].$$

The left-hand side of the above equation can be computed by the actuator, and therefore it can also compute the sequence $\{w + v\}$. What the theorem states is that a malicious sensor cannot distort the noise realization $\{w[1], w[2], w[3], \dots\}$ beyond adding a zero-power sequence to it. Suppose the control law g has been designed to provide good noise regulation performance of a stable system, then the performance of the system is good, subject only to the slightly increased cost of the private excitation of variance σ_e^2 . As we discuss in Section XI, this can be reduced to a low enough value that permits detection of a malicious sensor within a specified delay with an acceptable level of false alarm probability.

Theorem 2: Suppose $|a| < 1$, i.e., the system is stable.

- (i) Define the distortion $d[t] := z[t] - x[t]$. Then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} d^2[k] = 0.$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} d^2[k] = 0. \quad (22)$$

- (ii) If the malicious sensor is to remain undetected, the mean-square performance of $x[t]$ is the same as the reported mean-square performance $z[t]$ that the actuator believes it is

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} z^2[k].$$

Note that $x[k] = z[k] - d[k]$. Hence, $x^2[k] = z^2[k] + d^2[k] - 2(\gamma z[k])(\gamma^{-1}d[k])$. Now $|2(\gamma z[k])(\gamma^{-1}d[k])| \leq (\gamma^2 z^2[k]) + (\gamma^{-2}d^2[k])$. Therefore,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] &\leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (1 + \gamma^2)z^2[k] \\ &\quad + (1 + \gamma^{-2})d^2[k]. \end{aligned}$$

- (iii) Suppose the control law is $u^g(t) = fx(t)$ with $|a + bf| < 1$. The malicious sensor cannot compromise the performance of the system if it is to remain undetected, i.e., the mean-square performance of the system is

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] = \frac{\sigma_w^2 + b^2 \sigma_e^2}{1 - |a + bf|^2}.$$

Since the result is true for any $\gamma > 0$, taking the limit $\gamma \rightarrow 0$ and noting that the mean-square of d is 0 gives

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] \leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} z^2[k]. \quad (23)$$

Similarly, since $z[k] = x[k] + d[k]$, we have $z^2[k] = x^2[k] + d^2[k] + 2(\gamma x[k])(\gamma^{-1}d[k])$. Repeating the same argument as before, we have

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} z^2[k] &\leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (1 + \gamma^2)x^2[k] \\ &\quad + (1 + \gamma^{-2})d^2[k]. \end{aligned}$$

Taking the limit as $\gamma \rightarrow 0$, and noting that mean-square of d is 0, the above reduces to

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} z^2[k] \leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k]. \quad (24)$$

The distortion experienced by the actuator can therefore be thought of as the output of a linear dynamical system driven by an input sequence $\{v[t]\}$ satisfying (14). Therefore

$$d[k] = \sum_{n=0}^{k-1} a^n v[k-n]$$

The second result follows from (23) and (24). The third result is immediate noting that the mean-square of $\{z\}$ converges to $(\sigma_w^2 + b^2 \sigma_e^2)/(1 - |a + bf|^2)$. \square

VI. ACTIVE DEFENSE FOR NETWORKED CYBER-PHYSICAL SYSTEMS: THE SISO ARX CASE

The results developed in the previous section can be extended to a more general ARX system model. Specifically, we consider a unit delay, strictly minimum phase, single-input, single-output system described by

$$y[t+1] = -\sum_{m=0}^p a_m y[t-m] + \sum_{r=0}^h b_r u[t-r] + w[t+1] \quad (25)$$

where $a_m, b_r, y[t], u[t], w[t] \in \mathbb{R}$, $b_0 \neq 0$, and with $\{w[t]\}$ being zero-mean i.i.d. Gaussian process noise of variance σ_w^2 . Let q^{-1} denote the backward shift operator. The above system can equivalently be expressed as

$$A(q^{-1})y[t] = q^{-1}B(q^{-1})u[t] + w[t] \quad (26)$$

where $A(q^{-1}) := 1 + a_0q^{-1} + a_1q^{-2} + \cdots + a_pq^{-(p+1)}$, and $B(q^{-1}) := b_0 + b_1q^{-1} + \cdots + b_hq^{-h}$, with $B(q^{-1})$ being strictly minimum phase, i.e., all its roots lie strictly outside the unit circle.

We consider an honest actuator that is meant to, and implements the control policy $\{g\}$, and adds a private excitation $\{e\}$ as a defense. Unlike in the system considered in Section V, the output of the ARX system at any particular time depends on the past inputs. Specifically, the output at any time instant contains contributions of private excitation injected in the past. Hence, simply injecting an i.i.d. sequence of Gaussian random variables will not result in an output distribution that is i.i.d. across time.

However, since the actuator knows the past values of the private excitation, and also the transfer function of the system, it can perform “pre-equalization” by filtering the private excitation sequence before injecting it into the system. The filter, which we refer to as the pre-equalizer, has to be chosen in such a way that the component of the private excitation that appears in the output of the plant is an i.i.d. sequence of mean 0 and variance $b_0^2\sigma_e^2$. We let

$$e'[t] := -\frac{1}{b_0} (b_1 e'[t-1] + b_2 e'[t-2] + \cdots + b_h e'[t-h]) + e[t]$$

where $e[t]$ is i.i.d. and Gaussian of mean 0 and variance σ_e^2 . Since the system is strictly minimum phase, this is a stable generation of an excitation sequence [42].

With a pre-equalizer in place, the effective input applied by the actuator is

$$u[t] = g_t(z^t) + e'[t] \quad (27)$$

where $\{e'\}_t$ is the output of the pre-equalizer, and z^t is the sequence of measurements reported by the sensor up to time t . Therefore, the system evolves in closed-loop as

$$\begin{aligned} y[t+1] &= -\sum_{m=0}^p a_m y[t-m] + \sum_{r=0}^h b_r g_{t-r}(z^{t-r}) \\ &\quad + b_0 e[t] + w[t+1] \end{aligned} \quad (28)$$

where $\{e[t]\}$ is a sequence of i.i.d. Gaussian random variables with mean 0 and variance σ_e^2 . Hence, from the point of view of actuator tests and associated results, the above problem is similar to that described in Section V.

We propose that the actuator perform the following tests to check whether the sensor is reporting the measurements honestly.

- 1) **Actuator Test 1:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left(z[k+1] + \sum_{m=0}^p a_m z[k-m] - \sum_{r=0}^h b_r g_{k-r}(z^{k-r}) - b_0 e[k] \right)^2 = \sigma_w^2. \quad (29)$$

- 2) **Actuator Test 2:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left(z[k+1] + \sum_{m=0}^p a_m z[k-m] - \sum_{r=0}^h b_r g_{k-r}(z^{k-r}) \right)^2 = (b_0^2 \sigma_e^2 + \sigma_w^2). \quad (30)$$

Consequently, Theorem 1 and Theorem 2 can be extended as follows.

Theorem 3:

- 1) Let $v[t+1] := z[t+1] + \sum_{m=0}^p a_m z[k-m] - \sum_{r=0}^h b_r g_{k-r}(z^{k-r}) - b_0 e[t] - w[t+1]$, so that for an honest sensor which reports $z[t] \equiv y[t]$,

$v[t] = 0 \forall t$. If $\{z[t]\}$ passes tests (29) and (30), then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] = 0. \quad (31)$$

- 2) If the malicious sensor is to remain undetected, the mean-square performance of $\{y\}$ is what the actuator believes it is, which is the mean-square performance of $\{z\}$. Hence, if the control law was designed to provide a certain mean-square performance, then that is the value that is indeed attained.

Proof: Omitted, since the proof follows the same sequence of arguments as the proof of Theorem 1. \square

In the following section, we extend the technique to systems with arbitrary delay and colored noise, i.e., the more general ARMAX model, see [44].

VII. ACTIVE DEFENSE FOR NETWORKED CYBER-PHYSICAL SYSTEMS: THE SISO ARMAX CASE

A general ARMAX system with arbitrary but finite delay is a model that is encountered often in process control. In this section, we develop a Dynamic Watermarking scheme to secure such systems. We show that by commanding the actuator to inject private excitation whose spectrum is matched to that of the colored process noise entering the system, it can be ensured that a malicious sensor is constrained to distorting the process noise, the only quantity unknown to the actuator, by at most a zero-power signal. This amounts to a form of Internal Model Principle [45] for Dynamic Watermarking, and is a phenomenon that doesn't emerge in the analysis of a simple SISO or an ARX system treated in the previous sections.

A general ARMAX system with finite delay l is described by

$$y[t] = - \sum_{k=1}^p a_k y[t-k] + \sum_{k=0}^h b_k u[t-l-k] + \sum_{k=0}^r c_k w[t-k]. \quad (32)$$

Without loss of generality, we assume that $c_0 = 1$. Let q^{-1} be the backward shift operator so that $q^{-1}y[t] = y[t-1]$. Then, the above system can be expressed as

$$A(q^{-1})y[t] = q^{-l}B(q^{-1})u[t] + C(q^{-1})w[t] \quad (33)$$

where $a_0 := 1$, $A(q^{-1}) := a_0 + a_1q^{-1} + \dots + a_pq^{-p}$, $B(q^{-1}) := b_0 + b_1q^{-1} + \dots + b_hq^{-h}$, $C(q^{-1}) := c_0 + c_1q^{-1} + \dots + c_rq^{-r}$, and $B(q^{-1})$ and $C(q^{-1})$ are assumed to be strictly minimum phase. To secure the system, the actuator applies the control

$$u[t] = u^g[t] + B^{-1}(q^{-1})C(q^{-1})e[t] \quad (34)$$

where $u^g[t]$ is the control policy-specified input sequence and $\{e\}$ is a sequence of i.i.d Gaussian random variables of zero mean and variance σ_e^2 , denoting the actuator node's private excitation. Since $B(q^{-1})$ is assumed to be minimum phase, (34) is a stable generation of the control input $u[t]$. Consequently, the output of the system obeys

$$A(q^{-1})y[t] = q^{-l}B(q^{-1})u^g[t] + q^{-l}C(q^{-1})e[t] + C(q^{-1})w[t] \quad (35)$$

implying that

$$y[t] = - \sum_{k=1}^p a_k y[t-k] + \sum_{k=0}^h b_k u^g[t-l-k] + \sum_{k=0}^r c_k e[t-l-k] + \sum_{k=0}^r c_k w[t-k]. \quad (36)$$

While the q -domain derivation is just formal, the above difference equation can be obtained rigorously.

Define $\lambda[t] := e[t-l] + w[t]$. Then, the output of the system can be expressed as

$$y[t] = - \sum_{k=1}^p a_k y[t-k] + \sum_{k=0}^h b_k u^g[t-l-k] + \sum_{k=0}^r c_k \lambda[t-k]. \quad (37)$$

We now develop tests that the actuator should perform to detect maliciousness of the sensor. The fundamental idea behind the tests is to check whether the prediction-error of the reported sequence $\{z\}$ has the appropriate statistics. Specifically, the actuator processes the

reported measurements with a prediction-error filter defined through the following recursion for all $t \geq 0$

$$\begin{aligned} z_{t|t-1} &= -\sum_{k=1}^p a_k z[t-k] + \sum_{k=0}^h b_k u^g[t-l-k] \\ &\quad + \sum_{k=1}^r c_k \tilde{z}[t-k] \end{aligned} \quad (38)$$

$$\tilde{z}[t] = z[t] - z_{t|t-1}. \quad (39)$$

The filter is assumed to be initialized with $\tilde{z}[-k] = \lambda[-k]$, $k \in \{-1, -2, \dots, -r\}$. It is easy to verify that the above filter produces $\tilde{z}[t] \equiv \lambda[t]$ if $z[t] \equiv y[t]$. Based on this observation, we propose that the actuator perform the following tests.

- 1) **Actuator Test 1:** The actuator checks whether

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (\tilde{z}[k] - e[k-l])^2 = \sigma_e^2. \quad (40)$$

- 2) **Actuator Test 2:** The actuator checks whether

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \tilde{z}^2[k] = \sigma_w^2 + \sigma_e^2. \quad (41)$$

The following theorem shows that the above tests suffice to ensure that a malicious sensor cannot introduce any distortion beyond the addition of a zero-power signal to the process noise. i.e., a malicious sensor of only zero-additive distortion power can pass the above tests to remain undetected.

Theorem 4: Define $v[t] := \sum_{k=0}^p a_k z[t-k] - \sum_{k=0}^h b_k u^g[t-l-k] - \sum_{k=0}^r c_k \lambda[t-k]$ so that for an honest sensor reporting $z \equiv y$, $v \equiv 0$. If the sensor passes tests (40) and (41), then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} v^2[k] = 0.$$

Proof: From (38) and (39), one gets

$$\begin{aligned} \sum_{k=0}^r c_k \tilde{z}[t-k] &= z[t] + \sum_{k=1}^p a_k z[t-k] - \sum_{k=0}^h b_k u^g[t-l-k] \\ &= \sum_{k=0}^r c_k \lambda[t-k] + v[t]. \end{aligned} \quad (42)$$

Equivalently, assuming appropriate initial conditions, one has

$$C(q^{-1})\tilde{z}[t] = C(q^{-1})\lambda[t] + v[t].$$

This gives

$$\tilde{z}[t] = \lambda[t] + v_f[t]$$

where $v_f[t] := C^{-1}(q^{-1})v[t]$. Now

$$\begin{aligned} \tilde{z}[t] &= \lambda[t] + v_f[t] \\ &= w[t] + e[t-l] + v_f[t]. \end{aligned} \quad (43)$$

Since $\{\tilde{z}\}$ passes (40), we have from the above

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (v_f[k] + w[k])^2 = \sigma_w^2. \quad (44)$$

This gives

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} v_f^2[k] + 2v_f[k]w[k] = 0. \quad (45)$$

Since $\{\tilde{z}\}$ also passes (41), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (v_f[k] + w[k] + e[k-l])^2 = \sigma_w^2 + \sigma_e^2. \quad (46)$$

This gives

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} v_f^2[k] + 2v_f[k]w[k] + 2v_f[k]e[k-l] = 0.$$

Combining the above with (45), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} v_f[k]e[k-l] = 0. \quad (47)$$

Comparing (45) and (47) with (16) and (17), we see that the filtered distortion measure $\{v_f\}$ behaves the

same way $\{v\}$ does in the white noise case. Proceeding the same way as in the proof of Theorem 1, one arrives at

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} v_f^2[k] = 0. \quad (48)$$

Since $v[k] = C(q^{-1})v_f[k]$, and $C(q^{-1})$ is minimum phase, it follows that $\lim_{T \rightarrow \infty} (1/T) \sum_{k=0}^{T-1} v^2[k] = 0$. \square

VIII. ACTIVE DEFENSE FOR NETWORKED CYBER-PHYSICAL SYSTEMS: SISO SYSTEMS WITH PARTIAL OBSERVATIONS

In this section, we address SISO systems with noisy, partial observations. We consider a p th order single-input–single-output system described by

$$\mathbf{x}[t+1] = A\mathbf{x}[t] + Bu[t] + \mathbf{w}[t+1] \quad (49)$$

$$y[t+1] = C\mathbf{x}[t+1] + n[t+1]. \quad (50)$$

where $\mathbf{x}[t] \in \mathbb{R}^p$, and A , B , and C are known matrices of appropriate dimensions. The actuator, being honest, applies the input

$$u[t] = g_t(z^t) + e[t] \quad (51)$$

where $g_t(z^t)$ is the control policy-specified input, and $e[t] \sim \mathcal{N}(0, \sigma_e^2)$ is a sequence of i.i.d random variables denoting the actuator node's private excitation. Consequently, the system evolves as

$$\mathbf{x}[t+1] = A\mathbf{x}[t] + Bg_t(z^t) + be[t] + \mathbf{w}[t+1] \quad (52)$$

$$y[t+1] = C\mathbf{x}[t+1] + n[t+1]. \quad (53)$$

Let $z[t]$ be the measurement reported by the sensor at time t . Since the sensor can be malicious, $z[t]$ need not equal $y[t]$ for every t .

The actuator performs Kalman filtering on the reported measurements $\{z\}$ as follows:

$$\widehat{\mathbf{x}}_F(k+1|k) = A\widehat{\mathbf{x}}_F(k|k) + Bg_k(z^k) + Be[k] \quad (54)$$

$$\begin{aligned} \widehat{\mathbf{x}}_F(k+1|k+1) &= A\widehat{\mathbf{x}}_F(k|k) + Bg_k(z^k) + Be[k] \\ &\quad + K_k\nu_F[k+1] \end{aligned} \quad (55)$$

where $\nu_F[k+1] := z[k+1] - C\widehat{\mathbf{x}}_F(k+1|k)$, denotes the

(possibly) faulty innovations computed by the actuator, and K_k is the Kalman gain at time k .

We also define a Kalman filter that operates on the true measurements $y[t]$ as follows:

$$\widehat{\mathbf{x}}_R(k+1|k) = A\widehat{\mathbf{x}}_R(k|k) + Bg_k(z^k) + Be[k] \quad (56)$$

$$\begin{aligned} \widehat{\mathbf{x}}_R(k+1|k+1) &= A\widehat{\mathbf{x}}_R(k|k) + Bg_k(z^k) + Be[k] \\ &\quad + K_k\nu_R[k+1] \end{aligned} \quad (57)$$

where $\nu_R[k+1] := y[k+1] - C\widehat{\mathbf{x}}_R(k+1|k)$ is the “real innovations” or “true innovations” of the system. Of course, the actuator cannot implement the latter Kalman filter since it may not receive $y[k]$ from the sensor.

We now define

$$\begin{aligned} \mathbf{v}[k+1] &:= \widehat{\mathbf{x}}_F(k+1|k+1) - A\widehat{\mathbf{x}}_F(k|k) - Bg_k(z^k) \\ &\quad - Be[k] - K_k\nu_R[k+1] \end{aligned} \quad (58)$$

so that for an honest sensor that reports $z \equiv y$, we have $\nu \equiv \nu_R$ and consequently, $\mathbf{v} \equiv 0$. The actuator performs the following two tests to detect maliciousness of the sensor.

1) **Actuator Test 1:** Actuator checks whether

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} e[k] \left(\widehat{\mathbf{x}}_F(k+1|k+1) - A\widehat{\mathbf{x}}_F(k|k) - Bg_k(z^k) - Be[k] \right) = 0. \quad (59)$$

2) **Actuator Test 2:** Actuator checks whether

$$\begin{aligned} &\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \left(\widehat{\mathbf{x}}_F(k+1|k+1) - A\widehat{\mathbf{x}}_F(k|k) - Bg_k(z^k) - Be[k] \right)^T \\ &\quad \left(\widehat{\mathbf{x}}_F(k+1|k+1) - A\widehat{\mathbf{x}}_F(k|k) - Bg_k(z^k) - Be[k] \right) \\ &= \sigma_R^2 KK^T \end{aligned} \quad (60)$$

where σ_R^2 denotes the variance of the true innovations process, and K is the steady-state Kalman gain. We assume that (A, C) is observable so that the algebraic Riccati equation associated with the Kalman gain has a unique non-negative definite solution [42]. The following theorem shows that the above tests suffice to ensure that a

malicious sensor of only zero-additive distortion power can pass the tests to remain undetected.

Theorem 5: Suppose that the reported sequence of measurements passes the tests (59) and (60). Then

1)

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \|\mathbf{v}[k+1]\|^2 = 0. \quad (61)$$

Proof: Since the reported sequence of measurements passes (59), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} e[k](K_k \nu_R[k+1] + \mathbf{v}[k+1]) = 0.$$

Since the true innovations are independent of the zero-mean private excitation, it follows that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} e[k] \mathbf{v}[k+1] = 0. \quad (62)$$

Since the reported sequence of measurements also passes (60), we have

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (K_k \nu_R[k+1] + \mathbf{v}[k+1]) \\ (K_k \nu_R[k+1] + \mathbf{v}[k+1])^T = \sigma_R^2 K K^T. \end{aligned}$$

Simplifying, the above gives

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (K_k \nu_R[k+1] \mathbf{v}^T[k+1]) \\ + (K_k \nu_R[k+1] \mathbf{v}^T[k+1])^T \\ + (\mathbf{v}[k+1] \mathbf{v}^T[k+1]) = 0. \quad (63) \end{aligned}$$

We define $\mathcal{S}_k := \sigma(\hat{\mathbf{x}}_R^{k|k}, \hat{\mathbf{x}}_F^{k|k}, z^k, e^{k-2}, y^{k-1})$, where $\hat{\mathbf{x}}_R^{k|k} := \{\hat{\mathbf{x}}_R(k|k), \hat{\mathbf{x}}_R(k-1|k-1), \dots, \hat{\mathbf{x}}_R(-1|-1)\}$, and $\hat{\mathbf{x}}_F^{k|k}$ is defined likewise. Define $\widehat{\nu}_R[k] := E[\nu_R[k] | \mathcal{S}_k]$. From the Kalman filtering equations, we have

$$\begin{aligned} K_{k-1} \nu_R[k] &= \hat{\mathbf{x}}_R(k|k) - A \hat{\mathbf{x}}_R(k-1|k-1) \\ &\quad - B g_{k-1}(z^{k-1}) - Be[k-1] \end{aligned}$$

implying that

$$\begin{aligned} (\hat{\mathbf{x}}_R^{k-2|k-2}, \hat{\mathbf{x}}_F^{k|k}, e^{k-2}, y^{k-1}) &\rightarrow \\ (\hat{\mathbf{x}}_R(k-1|k-1), \hat{\mathbf{x}}_R(k|k), z^k) &\rightarrow \nu_R[k] \end{aligned}$$

forms a Markov chain. Therefore $\widehat{\nu}_R[k] := E[\nu_R[k] | \sigma(\hat{\mathbf{x}}_R^{k|k}, \hat{\mathbf{x}}_F^{k|k}, z^k, e^{k-2}, y^{k-1})] = E[\nu_R[k] | \sigma(\hat{\mathbf{x}}_R(k-1|k-1), \hat{\mathbf{x}}_R(k|k), z^k)]$. Therefore, we have

$$K_{k-1} \widehat{\nu}_R[k] = K_\nu(K_{k-1} \nu_R[k] + Be[k-1]) \quad (64)$$

where $K_\nu := \sigma_R^2 K_{k-1} K_{k-1}^T (\sigma_R^2 K_{k-1} K_{k-1}^T + \sigma_e^2 B B^T)^{-1}$. Define $K_{k-1} \widetilde{\nu}_R[k] := K_{k-1} \nu_R[k] - K_{k-1} \widehat{\nu}_R[k]$. Then, $(K_{k-1} \widetilde{\nu}_R[k-1], \mathcal{S}_k)$ is a martingale difference sequence. This is because $\widetilde{\nu}_R[k-1] \in \mathcal{S}_k$, and

$$E[\widetilde{\nu}_R[k] | \mathcal{S}_k] = 0.$$

Also, $\mathbf{v}[k] \in \mathcal{S}_k$. Hence, MST applies, and we have

$$\begin{aligned} \sum_{k=1}^T K_{k-1} \widetilde{\nu}_R[k] \mathbf{v}^T[k] \\ = \begin{bmatrix} o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \\ o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \\ \vdots & \dots & \vdots \\ o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \end{bmatrix} \\ + O(1) \quad (65) \end{aligned}$$

where $v_i[k]$ denotes the i th component of $\mathbf{v}[k]$. Now, since $K_k \nu_R[k] = K_k \widehat{\nu}_R[k] + K_k \widetilde{\nu}_R[k]$, from (64), we have

$$\begin{aligned} K_{k-1} \nu_R[k] &= (I - K_\nu)^{-1} Be[k-1] \\ &\quad + (I - K_\nu)^{-1} K_{k-1} \widetilde{\nu}_R[k]. \quad (66) \end{aligned}$$

Substituting this in (63) and equating the diagonals using (62) and (65) completes the proof. \square

IX. ACTIVE DEFENSE FOR NETWORKED CYBER-PHYSICAL SYSTEMS: MIMO SYSTEMS WITH GAUSSIAN NOISE

In this section, we investigate multiple-input–multiple-output (MIMO) linear stochastic dynamical systems. They pose additional challenges as compared to SISO systems since malicious sensors in a MIMO system can attempt to collude to prevent the other nodes from detecting their presence. In this section, we show how the actuators can prevent the malicious nodes from introducing “excessive” distortion, lest they expose their presence.

An m input MIMO Linear Dynamical System of order n is described by

$$\mathbf{x}[t+1] = \mathbf{A}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \mathbf{w}[t+1] \quad (67)$$

where $\mathbf{x}[t] \in \mathbb{R}^n$, $\mathbf{u}[t] \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, and $\{\mathbf{w}\}$ is a zero-mean i.i.d. sequence of Gaussian random vectors with covariance matrix $\sigma_w^2 I_n$. We consider the case where \mathbf{x} is perfectly observed.

Let $\mathbf{z}[t]$ be the measurements reported by the sensors to the other nodes in the system. Note that since the nodes in the system are allowed to exchange reported measurements among them, all malicious sensors have to be consistent and report the same (but possibly erroneous) measurements to all honest nodes in the system. A history-dependent control policy is assumed to be in place, which dictates that the i th input to the system at time t be

$$u_i^g[t] = g_t^i(\mathbf{z}^t). \quad (68)$$

As reasoned before, to secure the system, each actuator superimposes on the input specified by the control policy, an additional zero-mean private excitation that it draws from a distribution, here Gaussian, that is made public. It should be noted that while the *distribution* is made public, the actual values of the excitation are not revealed by the actuator to any other node. The private excitation value drawn at each time t is chosen to be independent of its private excitation values at all the other time instants, of the private excitation values of other actuator nodes, and of the control policy-specified input. Therefore, actuator i applies at time t the input

$$u_i[t] = u_i^g[t] + e_i[t] = g_t^i(\mathbf{z}^t) + e_i[t] \quad (69)$$

where $e_i[t] \sim \mathcal{N}(0, \sigma_e^2)$ is independent of $e_j[k]$ for $(j, k) \neq (i, t)$, $\mathbf{x}[m], \mathbf{z}[m]$ for $m \leq t$, and $\mathbf{w}[n]$ for all n . We assume that all actuators are honest, meaning that they apply control inputs in accordance to (69).

We propose the following tests to be performed by each actuator i .

- 1) **Actuator Test 1:** Actuator i checks whether the reported sequence of measurements $\{\mathbf{z}\}$ satisfies

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} & \left(\mathbf{z}[k+1] - \mathbf{A}\mathbf{z}[k] - \mathbf{B}g_k(\mathbf{z}^k) \right) \\ & \left(\mathbf{z}[k+1] - \mathbf{A}\mathbf{z}[k] - \mathbf{B}g_k(\mathbf{z}^k) \right)^T = \sigma_e^2 \mathbf{B} \mathbf{B}^T + \sigma_w^2 I_n. \end{aligned} \quad (70)$$

- 2) **Actuator Test 2:** Actuator i checks whether the reported sequence of measurements $\{\mathbf{z}\}$ satisfies

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} e_i[k] & \left(\mathbf{z}[k+1] - \mathbf{A}\mathbf{z}[k] - \mathbf{B}g_k(\mathbf{z}^k) \right) \\ & = B_{:,i} \sigma_e^2. \end{aligned} \quad (71)$$

In Actuator Test 2, we have simplified the test to directly check the quantity that is important, which in this case is the analog of (12). We note that for any set of conditions to be an admissible test, they have to be: 1) checkable by the sensors and actuators based on their observations and what is reported to them; and 2) satisfied if all parties are honest. The above two conditions satisfy these two conditions.

Define

$$\mathbf{v}[t+1] := \mathbf{z}[t+1] - \mathbf{A}\mathbf{z}[t] - \mathbf{B}g_t(\mathbf{z}^t) - \mathbf{B}\mathbf{e}[t] - \mathbf{w}[t+1]$$

so that if $\mathbf{z}[t] \equiv \mathbf{x}[t]$, $\mathbf{v}[t] \equiv 0$. It is easy to see that as in the case of SISO systems, the sequence $\{\mathbf{v}\}$ has the interpretation as the additive distortion of the process noise deliberately introduced by the malicious sensors. We term the quantity

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \|\mathbf{v}[k]\|^2$$

the *additive distortion power* of the malicious sensors. The following theorem, akin to Theorem 1, proves that malicious sensors of only zero-additive distortion power can pass the above tests, thereby remaining undetected.

Theorem 6: Suppose that the reported sequence of measurements passes the tests (70), and (71). If B is of rank n , then we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \|\mathbf{v}[k]\|^2 = 0. \quad (72)$$

Proof: Since $\{\mathbf{z}\}$ satisfies (71), we have $\forall i \in \{1, 2, \dots, m\}$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T e_i[k](B\mathbf{e}[k] + \mathbf{w}[k+1] + \mathbf{v}[k+1]) = \sigma_e^2 B_{(i)}. \quad (73)$$

It follows that for all $i \in \{1, 2, \dots, m\}$

forming a Markov chain. Consequently $\widehat{\mathbf{w}}[k] := E[\mathbf{w}[k] | \sigma(\mathbf{x}^{k-2}, \mathbf{e}^{k-2}, \mathbf{x}[k-1], \mathbf{x}[k], \mathbf{z}^k)] = E[\mathbf{w}[k] | \sigma(\mathbf{x}[k-1], \mathbf{x}[k], \mathbf{z}^k)]$. Therefore,

$$\widehat{\mathbf{w}}[k] = K_W(B\mathbf{e}[k-1] + \mathbf{w}[k]) \quad (77)$$

where $K_W := \sigma_w^2 (\sigma_e^2 BB^T + \sigma_w^2 I)^{-1}$. Now define $\widetilde{\mathbf{w}}[k] := \mathbf{w}[k] - \widehat{\mathbf{w}}[k]$. Then, $(\widetilde{\mathbf{w}}[k-1], \mathcal{S}_k)$ is a martingale difference sequence. This is because $\widetilde{\mathbf{w}}[k-1] \in \mathcal{S}_k$, and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T e_i[k]\mathbf{v}[k+1] = 0. \quad (78) \quad E[\widetilde{\mathbf{w}}[k] | \mathcal{S}_k] = 0.$$

Therefore,

Also, $\mathbf{v}[k] \in \mathcal{S}_k$. Hence, MST applies, and we have

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \mathbf{e}[k]\mathbf{v}^T[k+1] &= 0. \quad (74) \quad \sum_{k=0}^{T-1} \widetilde{\mathbf{w}}[k+1]\mathbf{v}^T[k+1] \\ &= \begin{bmatrix} o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \\ o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \\ \vdots & \dots & \vdots \\ o\left(\sum_{k=0}^{T-1} v_1^2[k+1]\right) & \dots & o\left(\sum_{k=0}^{T-1} v_p^2[k+1]\right) \end{bmatrix} + O(1) \end{aligned} \quad (79)$$

Since $\{\mathbf{z}\}$ also satisfies (70), we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (B\mathbf{e}[k] + \mathbf{w}[k+1] + \mathbf{v}[k+1])^T = \sigma_e^2 BB^T + \sigma_w^2 I_n. \quad (75)$$

Using (74) and the fact that the process noise $\mathbf{w}[k+1]$, and the private excitation of the actuators at time k are independent, the above simplifies to

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{w}[k+1]\mathbf{v}^T[k+1]) + (\mathbf{w}[k+1]\mathbf{v}^T[k+1])^T \\ (\mathbf{v}[k+1]\mathbf{v}^T[k+1]) = 0. \quad (76) \end{aligned}$$

Define $S_k := \sigma(\mathbf{x}^k, \mathbf{z}^k, \mathbf{e}^{k-2})$, and $\widehat{\mathbf{w}}[k] := E[\mathbf{w}[k] | S_k]$. Since

where $v_i[k+1]$ denotes the i th element of $\mathbf{v}[k+1]$.

Using the above, we have

$$\begin{aligned} \mathbf{w}[k+1] &= \widehat{\mathbf{w}}[k+1] + \widetilde{\mathbf{w}}[k+1] \\ &= K_W(B\mathbf{e}[k] + \mathbf{w}[k+1]) + \widetilde{\mathbf{w}}[k+1]. \end{aligned} \quad (80)$$

From the assumption on the rank of B , it follows that K_W has all eigenvalues strictly lesser than unity. Therefore, simplifying the above

$$\mathbf{w}[k] = \mathbf{x}[k] - A\mathbf{x}[k-1] - Bg_{k-1}(\mathbf{z}^{k-1}) - B\mathbf{e}[k-1]$$

$$\mathbf{w}[k+1] = (I - K_W)^{-1} K_W B \mathbf{e}[k] + (I - K_W)^{-1} \widetilde{\mathbf{w}}[k+1]. \quad (81)$$

Substituting this into (76), using (74) and (79), and equating the q th entry along the diagonal, we have

$$\sum_{k=0}^{T-1} v_q^2[k+1] + o\left(\sum_{k=0}^{T-1} v_q^2[k+1]\right) = o(T). \quad (82)$$

Since this is true for all $q \in \{1, 2, \dots, n\}$, dividing the above by T and taking the limit as $T \rightarrow \infty$ completes the proof. \square

X. EXTENSION TO NON-GAUSSIAN SYSTEMS

The results developed in the previous sections assumed that the process noise follows a Gaussian distribution. This assumption can be relaxed to a certain extent. In this section, we illustrate how this may be relaxed for a single-input-single-output system.

Consider a SISO system described by

$$x[t+1] = ax[t] + bu[t] + w[t+1] \quad (83)$$

where $w[t] \sim P_w$ is an i.i.d. process with mean 0 and variance σ_w^2 . In such a case, the actuator can choose the distribution of its private excitation so that the output of the private excitation has the same distribution as the process noise. Hence, the actuator applies to the system the input

$$u[t] = g_t(z^t) + e[t] \quad (84)$$

where $e[t] \sim P_e$ is an i.i.d. sequence. As before, even though the actuator implements the control policy $\{g\}$, the policy is applied to the measurements $z[t]$ reported by the sensor, which could differ from the true output $x[t]$. Therefore, the system evolves in closed-loop as

$$x[t+1] = ax[t] + bg_t(z^t) + be[t] + w[t+1]. \quad (85)$$

The actuator then performs the following tests to whether the sensor is malicious.

- 1) **Actuator Test 1:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (z[k+1] - az[k] - bg_k(z^k) - be[k])^2 \\ = \sigma_w^2. \end{aligned} \quad (86)$$

- 2) **Actuator Test 2:** Check whether the reported sequence of measurements $\{z[t]\}$ satisfies

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} (z[k+1] - az[k] - bg_k(z^k))^2 \\ = 2\sigma_w^2. \end{aligned} \quad (87)$$

As before, let

$$v[t+1] := z[t+1] - az[t] - bg_t(z^t) - be[t] - w[t+1]$$

so that for an honest sensor that reports $z[t] \equiv x[t]$, $v[t] = 0 \forall t$. The *additive distortion power* of a malicious sensor, too, is defined as before, i.e., as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k].$$

The following result, a generalization of Theorem 1, shows that the above tests suffice to ensure that a malicious sensor of only zero effective power can remain undetected.

Theorem 7: If $\{z[t]\}$ satisfies tests (86) and (87), then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] = 0. \quad (88)$$

Proof: Following the same sequence of arguments as in the proof of Theorem 1, we arrive at the following equalities:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T v^2[k] + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T 2v[k]w[k] = 0 \quad (89)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T e[k-1]v[k] = 0. \quad (90)$$

Let $\mathcal{S}_k := \sigma(x^k, z^k, e^{k-2})$, and $\widehat{w}[k] := E[w[k]|\mathcal{S}_k]$. The general complication is that due to non-Gaussianity, the conditional mean estimate may be nonlinear in $e[k-1]$ and $w[k]$, unlike in the Gaussian case. However, since the sequence of observations are i.i.d., and the

distribution of the actuator's private excitation is chosen to be the same as the distribution of the process noise, we have

$$\hat{w}[k] = \frac{1}{2}(be[k-1] + w[k]) = \beta(be[k-1] + w[k]) \quad (91)$$

where $\beta := 1/2$. This can be written as

$$\hat{w}[k] = \alpha e[k-1] + \beta w[k]$$

where $\alpha := b\beta$. Let $\tilde{w}[k] := w[k] - \hat{w}[k]$.

Now, the result below also holds following the same sequence of arguments employed in the proof of Theorem 1:

$$\sum_{k=1}^T v[k]\tilde{w}[k] = o\left(\sum_{k=1}^T v^2[k]\right) + O(1). \quad (92)$$

Hence, following similar arguments as in the proof of Theorem 1, we obtain

$$\sum_{k=1}^T v^2[k] + \sum_{k=1}^T 2v[k]w[k] = (1 + o(1))\left(\sum_{k=1}^T v^2[k]\right) + O(1).$$

Dividing the above equation by T , taking the limit as $T \rightarrow \infty$, and invoking (89) completes the proof. \square

Remark: The net import of this result is that to safeguard against malicious actions of the sensor, we need to add a level of noise exactly equal to the process noise already present in the system. Hence, there is an amplification of the process noise's standard deviation by $\sqrt{2}$ that appears in the price of guarding against malicious attacks by the sensor. In contrast, in the Gaussian case, we have seen that this extra cost can be made as small as desired by choosing σ_e^2 as small as desired, though, of course, at the cost of delaying detection with an acceptable false alarm probability, as we discuss in the next section.

XI. STATISTICAL TESTS FOR ACTIVE DEFENSE

The results developed in the previous sections are couched in an asymptotic fashion. They characterize what can be detected and how that may be done. These asymptotic characterizations can be used to develop

statistical tests that reveal malicious activity in a finite period of time with acceptable false alarm rates.

The task at hand is to develop statistical tests equivalent to asymptotic tests, such as (12), (13), (70), (71), which include tests for covariance matrices of certain random vectors. This problem has been addressed in [46] in the context of fault detection in control systems in which the innovation sequence is tested for the properties of whiteness, mean, and covariance. Since these are also the properties that have to be satisfied by the test sequences that we construct in (12), (13), and (71), the same test described in [46] can be used.

Yet another test that is particularly well-suited for the problem at hand is the sequential probability ratio test, introduced in [47]. In the standard setting of sequential hypothesis testing, for each time t , the set of all possible observations S_m (or the space of sufficient statistics) is partitioned into three sets R_0 , R_1 , R . At each time t , a decision is made whether to select the null hypothesis (and stop), select the alternative hypothesis (and stop), or continue observing. The null (alternative) hypothesis is accepted, and hypothesis testing is stopped if at some time t , the observations available till time t , denoted by $\{z_s^t\}$, fall in the set R_0 (R_1). Hypothesis testing is continued at time t if $\{z_s^t\}$ falls in R . One such sequential test is the sequential probability ratio test presented in [47] whereby the likelihood ratio of observations up to time t , denoted by Λ_t , is compared against two thresholds τ_1 and τ_2 . The null (alternative) hypothesis is accepted at time t if $\Lambda_t < \tau_1$ ($\Lambda_t > \tau_2$). If $\tau_1 \leq \Lambda_t \leq \tau_2$, another observation is drawn, and the process repeats.

There are a few aspects wherein the problem at hand departs from the above. These are enumerated below. In what follows, the null hypothesis is that the control system is not under attack.

- 1) In the problem of secure control, it is not possible to attribute a distribution (or even a parameterized, finite-dimensional family of distributions) from which the observations would occur under the alternative hypothesis (that the control system is under attack). Consequently, a likelihood ratio cannot be defined, and the sequential probability ratio test cannot be used. Hence, the problem at hand is not to test one hypothesis against another, but is only that of rejecting the null hypothesis or otherwise.
- 2) In the classical setup as described in [47], in a finite time and with probability 1, the test stops, and one hypothesis or the other is accepted. However, the adversary could be a "sleeper agent." It could behave as an honest party for a long time and then operate in a malicious manner. Hence, in our situation, at no time can one "accept" the null hypothesis forever for all subsequent time. Therefore, in the systems of

interest here, there are only two possible decisions at any time t , reject the null hypothesis or continue observations.

- 3) In the classical setting, optimal performance is obtained by considering all observations that are available up to the time of making a decision. However, in our problem, an adversary that behaves maliciously in a bursty fashion is not likely to be detected since long-term averaging would nullify these effects. To account for this, a moving window approach could be employed over which a test statistic can be calculated. That is, the statistical test is conducted over a window of l observations, $nl \leq t < (n+1)l$. In each window, it assesses whether to reject the null hypotheses. In this manner, malicious activity can be detected within l samples, with acceptable false alarm rate. The choice of the excitation variance σ_e^2 can be based on the acceptable detection delay l , and the false alarm rate.

Based on the above, the basis of a statistical hypothesis test for the problem can be chosen as follows. Under the null hypothesis, the sample covariance matrix follows the Wishart distribution [48] (the multidimensional counterpart of the Chi-squared distribution). Hence, for a given false alarm rate α , the threshold $\tau(\alpha)$ can be determined for the likelihood function of the observations, whereby the likelihood function considered for the test at time t is the probability density evaluated at the l most recent observations. If this likelihood function exceeds the threshold at any time t , an alarm is raised. If not, the test is repeated for the next time instant. The following simulation example is illustrative.

Example: We consider the scenario of an ARX system addressed in Section VI. Specifically, we consider a second order plant of the type common in process control. A single-input, single-output stable plant obeying (25), with parameters $a_0 = 0.7$, $a_1 = 0.2$, $b_0 = 1$, $b_1 = 0.5$ and $\sigma_w^2 = 1$ is considered. Note that the chosen system is minimum phase so that the pre-equalizer required to filter the sequence of private excitation is stable.

The sensor is assumed to be initially well-behaved but suddenly become malicious after a certain period (denoting the time of initiation of attack), while the actuator is assumed to remain uncompromised. The actuator applies control inputs

$$u[k] = -\frac{1}{b_0}(a_0 z[t] + a_1 z[t-1] + b_1 u[k-1]) + e[k]$$

where $z[k]$ is the measurement reported by the sensor at time k , and $e[k] \sim \text{i.i.d. } \mathcal{N}(0, 1)$ is the actuator node's

private excitation. Therefore, the system evolves as

$$\begin{aligned} y[k+1] &= 0.7(y[k] - z[k]) + 0.3(y[k-1] - z[k-1]) \\ &\quad + e[k] + w[k+1]. \end{aligned}$$

The adversary attacks the system at some random time and begins to report false measurements. To do so, the adversary estimates optimally the process noise at each instant from the measurements that it observes. Since the process noise and the private excitation have the same variance, the optimal estimate is simply $(1/2)(y[k+1] - 0.7(y[k] - z[k]) - 0.3(y[k-1] - z[k-1]))$. Once the adversary forms its estimate, it adds a noise correlated with the estimate as follows. The adversary forms $v[t] = n[t] - \hat{w}[t]$, where $n[t] \sim \mathcal{N}(0, 1)$, the same distribution as $w[t]$, and reports $z[t] = x[t] + v[t]$. Note that in the absence of dynamic watermarking, employing this strategy would enable the adversary to form perfect estimates of the process noise, and, consequently, as reasoned in Section IV, would render every detection algorithm useless.

Fig. 3 plots the evolution of the negative log likelihood function over time when dynamic watermarking is employed. In our simulation, the adversary initiates attack at time epoch 4500. As indicated in Fig. 3, the (windowed) negative log likelihood function corresponding to (29) stays within limits until the attack. Around time epoch 4500, the likelihood function steadily increases, indicating the onset of an attack. Based on tolerable false alarm rates, an appropriate detection threshold can be set.

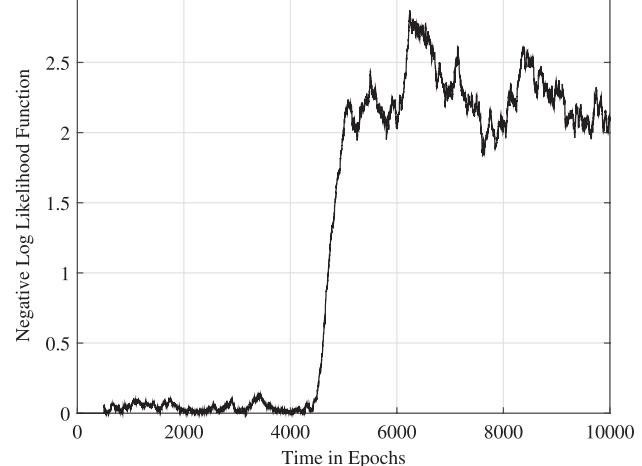


Fig. 3. Sequential hypothesis testing for attack detection.

XII. CONCLUSION

In this paper, we have considered the fundamental problem of security of networked cyber–physical systems. Securing such systems is not the same as securing a communication network due to the fundamental role played by sensors and actuators in interfacing with and controlling the physical plant. We have provided a general procedure for dynamic watermarking that imposes private excitation signals on the actuation signals whose presence can then be traced around the loop to detect malicious behaviors of sensors.

We have only explored the tip of the iceberg of this procedure. It can be used in more general contexts for a variety of plants of interest in control systems, such as process control systems, in which more nonlinear models may be of interest. Further study into performance of finite-time tests for Dynamic Watermarking would also be of practical interest.

There is much to be done in this area if cyber–physical system applications are to proliferate to address critical infrastructural and societal needs. ■

REFERENCES

- [1] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop Future Directions in Cyber-Physical Syst. Security* Newark, NJ, USA, 2009.
- [2] Y. Mo et al., "Cyber-physical security of a smart grid infrastructure," *Proc. IEEE*, vol. 100, no. 1, pp. 195–209, 2012.
- [3] M. Abrams and J. Weiss, "Malicious Control System Cyber Security Attack Case Study-Maroochy Water Services, Australia." [Online]. Available: https://www.mitre.org/sites/default/files/pdf/08_1145.pdf
- [4] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security Privacy*, vol. 9, no. 3, pp. 49–51, 2011.
- [5] S. Stanford, V. Paxson, and N. Weaver, "How to own the internet in your spare time," in *Proc. 11th USENIX Security Symp.* Berkeley, CA, USA: 2002, pp. 149–167. [Online]. Available: <http://dl.acm.org/citation.cfm?id=647253.720288>
- [6] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O. Gorman, "Electronic marking and identification techniques to discourage document copying," *IEEE J. Sel. Areas Comm.*, vol. 13, no. 8, pp. 1495–1504, 1995.
- [7] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.
- [8] A. Z. Tirkel, G. Rankin, R. Van Schyndel, W. Ho, N. Mee, and C. F. Osborne, "Electronic watermark," *Digital Image Comput., Technol. Appl.*, pp. 666–673, 1993.
- [9] "President's Commission on Critical Infrastructure Protection. Critical Foundations: Protecting America's Infrastructures," 1997. [Online]. Available: <https://www.fas.org/sgp/library/pccip.pdf>
- [10] Y. Y. Haimes, N. C. Matalas, J. H. Lambert, B. A. Jackson, and J. F. R. Fellows, "Reducing vulnerability of water supply systems to attack," *J. Infrastructure Syst.*, vol. 4, no. 4, pp. 164–177, 1998. [Online]. Available: [http://dx.doi.org/10.1061/\(ASCE\)1076-0342\(1998\)4:4\(164\)](http://dx.doi.org/10.1061/(ASCE)1076-0342(1998)4:4(164))
- [11] M. Bishop and E. Goldman, "The strategy and tactics of information warfare," *Contemporary Security Policy*, vol. 24, no. 1, pp. 113–139, 2003. [Online]. Available: <http://dx.doi.org/10.1080/13523260312331271839>
- [12] K. Barnes and B. Johnson, "Introduction to SCADA: Protection and Vulnerabilities," Tech. Rep. INEEL/EXT-04-01710, Idaho National Engineering and Environmental Laboratory, 2004.
- [13] Y. Y. Haimes, "Risk of terrorism to cyber-physical and organizational-societal infrastructures," *Public Works Management and Policy*, vol. 6, no. 4, pp. 231–240, 2002. [Online]. Available: <http://pwm.sagepub.com/content/6/4/231.abstract>
- [14] D. Watts, "Security and vulnerability in electric power systems," *35th North Amer. Power Symp.*, pp. 559–566, 2003.
- [15] A. S. Brown, "SCADA vs. the hackers," *Mech. Eng.*, vol. 124, no. 12, p. 37, 2002.
- [16] J. Eisenhauer, P. Donnelly, M. Ellis, and M. O'Brien, "Roadmap to secure control systems in the energy sector," Energetics Inc., Sponsored by U.S. Dept. Energy U.S. Dept. Homeland Security, Jan. 2006.
- [17] "Roadmap to Secure Control Systems in the Water Sector," *Water Sector Coordinating Council Cyber Security Work. Group*' Mar. 2006. [Online]. Available: <http://www.awwa.org/Portals/0/files/legreg/Security/SecurityRoadmap.pdf>
- [18] "Roadmap to Secure Control Systems in the Chemical Sector," *Chemical Sector Roadmap Work. Group*, Sep. 2009. [Online]. Available: <https://scadahacker.com/library/Documents/Roadmaps/Roadmap%20to%20Secure%20Control%20Systems%20in%20the%20Chemical%20Sector.pdf>
- [19] "Roadmap to Secure Control Systems in the Transportation Sector," *Ind. Control Syst. Joint Work. Group*, Aug. 2012. [Online]. Available: <https://ics-cert.us-cert.gov/sites/default/files/ICSJWG-Archive/TransportationRoadmap20120831.pdf>
- [20] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *The IEEE 28th Int. Conf. Distributed Comput. Syst. Workshops*, 2008, pp. 495–500.
- [21] J. Ponniah, Y.-C. Hu, and P. R. Kumar, "A clean slate approach to secure wireless networking," *Foundations and Trends Netw.*, vol. 9, no. 1, pp. 1–105, 2014. [Online]. Available: <http://dx.doi.org/10.1561/1300000037>
- [22] A. A. Cardenas, S. Amin, and S. Sastry, "Research Challenges for the Security of Control Systems." [Online]. Available: https://www.usenix.org/legacy/events/hotsec08/tech/full_papers/cardenas/cardenas_html/hotsecHTML.html
- [23] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, no. 1, pp. 163–187, 2007.
- [24] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S. S. Sastry, "Optimal linear LQG control over lossy networks without packet acknowledgment," *Asian J. Control*, vol. 10, no. 1, pp. 3–13, 2008.
- [25] V. Gupta, B. Hassibi, and R. M. Murray, "Optimal LQG control across packet-dropping links," *Syst. Control Lett.*, vol. 56, no. 6, pp. 439–446, 2007.
- [26] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Automat. Control*, vol. 46, no. 9, pp. 1384–1400, 2001.
- [27] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses," in *Proc. IEEE 43rd Conf. Decision Control*, vol. 4, Dec. 2004, pp. 4180–4186.
- [28] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control*. Berlin, Germany:Springer-Verlag, 2009.

APPENDIX

Notation

This paper uses the following notation.

- Scalars are denoted using lowercase: a
- Vectors are denoted using lowercase boldface: \mathbf{x}
- The i th component of vector \mathbf{x} : x_i
- Matrices are denoted using uppercase: A
- The i th column of matrix A : $A_{\cdot i}$
- The i th row of matrix A : $A_{i \cdot}$
- The submatrix of A formed by the intersection of rows i through j and columns p through q : $A_{i:j,p:q}$
- The matrix with i th column of A removed: $A_{\cdot(i)}$
- The matrix with i th row of A removed: $A_{(-i)}$
- The vector with i th component of \mathbf{x} removed: \mathbf{x}_{-i}

Acknowledgment

The authors would like to thank Bruno Sinopoli (Carnegie Mellon University (CMU), Pittsburgh, PA, USA), Yilin Mo (Nanyang Technological University, Singapore) and Sean Weerakkody (CMU) for their valuable comments.

- [29] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conf. Decision and Control*, 2010, pp. 5967–5972.
- [30] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," *Amer. Control Conf.*, pp. 195–200, 2015.
- [31] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *49th IEEE Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 337–344.
- [32] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th IEEE Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2009, pp. 911–918.
- [33] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, 2014.
- [34] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.
- [35] S. Weerakkody, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on control systems using robust physical watermarking," in *53rd IEEE Conf. Decision Control*, Dec. 2014, pp. 3757–3764.
- [36] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *2013 IEEE 52nd Annu. Conf. Decision Control*, 2013, pp. 1854–1859.
- [37] A. Teixeira, I. Shames, H. Sandberg, and K. Johansson, "Revealing stealthy attacks in control systems," in *50th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2012, pp. 1806–1813.
- [38] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Shield: A data verification framework for participatory sensing systems," in *Proc. 8th ACM Conf. Security Privacy in Wireless Mobile Networks*. New York, NY, USA, 2015. [Online]. Available: <http://doi.acm.org/10.1145/2766498.2766503>
- [39] J. Ponniah, Y.-C. Hu, and P. Kumar, "A system-theoretic clean slate approach to provably secure ad hoc wireless networking," *IEEE Trans. Control Network Syst.*, vol. 3, no. 2, pp. 206–217, May 2015.
- [40] I.-H. Hou, V. Borkar, and P. Kumar, "A theory of QoS for wireless," in *IEEE INFOCOM*, 2009, pp. 486–494.
- [41] R. Kalman, "On the general theory of control systems," *IRE Trans. Automat. Control*, vol. 4, no. 3, pp. 110–110, Dec. 1959.
- [42] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.
- [43] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Statist.*, pp. 154–166, 1982.
- [44] K. J. Åström, *Introduction to Stochastic Control*. New York, NY, USA: Academic Press, 1970.
- [45] B. A. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, 1976.
- [46] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.
- [47] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.
- [48] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 1958, vol. 2.

ABOUT THE AUTHORS

Bharadwaj Satchidanandan received the master's degree from Indian Institute of Technology Madras, Tamil Nadu, India, in 2015, where he worked on wireless communications. He is a doctoral student in the Electrical and Computer Engineering Department at Texas A&M University, College Station, TX, USA.



From May 2015 and August 2015, he interned at Intel Labs, Santa Clara, CA, USA, where he worked on interference cancellation algorithms for next-generation wireless networks. His research interests include cyberphysical systems, power systems, security, database privacy, communications, control, and signal processing.

P. R. Kumar (Fellow, IEEE) received the B. Tech. from Indian Institute of Technology (IIT) Madras, Chennai, India, in 1973, the D.Sc. from Washington University in St. Louis, St. Louis, MO, USA, in 1977.



From 1977 to 1984, he was a Faculty Member at University of Maryland, Baltimore County, Baltimore, MD, USA and at University of Illinois, Urbana-Champaign, Champaign, IL, USA, from 1985 to 2011. He is currently at Texas A&M University, College Station, TX, USA. He is currently a D. J. Gandhi Distinguished Visiting Professor at IIT Bombay, Maharashtra, India, and an Honorary Professor at IIT Hyderabad. His current research is focused on stochastic systems, energy systems, wireless networks, security, automated transportation, and cyberphysical systems.

Dr. Kumar was Leader of the Guest Chair Professor Group on Wireless Communication and Networking at Tsinghua University, Beijing, China. He is a member of the US National Academy of Engineering and The World Academy of Sciences. He was awarded a Doctor Honoris Causa by ETH, Zurich. He has received the IEEE Field Award for Control Systems, the Donald P. Eckman Award of the AACC, Fred W. Ellersick Prize of the IEEE Communications Society, the Outstanding Contribution Award of ACM SIGMOBILE, the Infocom Achievement Award, and the SIGMOBILE Test-of-Time Paper Award. He is an ACM Fellow. He was awarded the Distinguished Alumnus Award from IIT Madras, the Alumni Achievement Award from Washington University, and the Daniel Drucker Eminent Faculty Award from the College of Engineering at the University of Illinois.