Attack Detection and Identification in Cyber-Physical Systems

Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo

Abstract

Cyber-physical systems are ubiquitous in power systems, transportation networks, industrial process control and critical infrastructures. These systems need to operate reliably in the face of unforeseen failures and external malicious attacks. In this paper (i) we propose a mathematical framework for cyber-physical systems, attacks, and monitors; (ii) we characterize fundamental monitoring limitations from system-theoretic and graph-theoretic perspectives; and (iii) we design centralized and distributed attack detection and identification monitors. Finally, we validate our findings through compelling examples.

I. Introduction

Cyber-physical systems integrate physical processes, computational resources, and communication capabilities. Examples of cyber-physical systems include transportation networks, power generation and distribution networks, water and gas distribution networks, and advanced communication systems. As recently highlighted by the Maroochy water breach [1] in March 2000, multiple recent power blackouts in Brazil [2], the SQL Slammer worm attack on the Davis-Besse nuclear plant in January 2003 [3], the StuxNet computer worm [4] in June 2010, and by various industrial security incidents [5], cyber-physical systems are prone to failures and attacks on their physical infrastructure, and cyber attacks on their data management and communication layer.

Concerns about security of control systems are not new, as the numerous manuscripts on systems fault detection, isolation, and recovery testify [6], [7]. Cyber-physical systems, however, suffer from specific vulnerabilities which do not affect classical control systems, and for which

This material is based upon work supported in part by NSF grant CNS-1135819 and by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo are with the Center for Control, Dynamical Systems and Computation, University of California at Santa Barbara, {fabiopas,dorfler,bullo}@engineering.ucsb.edu.

appropriate detection and identification techniques need to be developed. For instance, the reliance on communication networks and standard communication protocols to transmit measurements and control packets increases the possibility of intentional and worst-case attacks against physical plants. On the other hand, information security methods, such as authentication, access control, and message integrity, appear inadequate for a satisfactory protection of cyber-physical systems. Indeed, these security methods do not exploit the compatibility of the measurements with the underlying physical process or the control mechanism, and they are therefore ineffective against insider attacks targeting the physical dynamics [1].

Related work. The analysis of vulnerabilities of cyber-physical systems to external attacks has received increasing attention in the last years. The general approach has been to study the effect of specific attacks against particular systems. For instance, in [8] *deception* and *denial of service* attacks against a networked control system are defined, and, for the latter ones, a countermeasure based on semi-definite programming is proposed. Deception attacks refer to the possibility of compromising the integrity of control packets or measurements, and they are cast by altering the behavior of sensors and actuators. Denial of service attacks, instead, compromise the availability of resources by, for instance, jamming the communication channel. In [9] *false data* injection attacks against static state estimators are introduced. False data injection attacks are specific deception attacks in the context of static estimators. It is shown that undetectable false data injection attacks can be designed even when the attacker has limited resources. In a similar fashion, *stealthy deception attacks* against the Supervisory Control and Data Acquisition system are studied, among others, in [10]. In [11] the effect of *replay attacks* on a control system is discussed. Replay attacks are cast by hijacking the sensors, recording the readings for a certain amount of time, and repeating such readings while injecting an exogenous signal into the system. It is shown that these attacks can be detected by injecting a signal unknown to the attacker into the system. In [12] the effect of *covert attacks* against control systems is investigated. Specifically, a parameterized decoupling structure allows a covert agent to alter the behavior of the physical plant while remaining undetected from the original controller. In [13] a resilient control problem is studied, in which control packets transmitted over a network are corrupted by a human adversary. A receding-horizon Stackelberg control law is proposed to stabilize the control system despite the attack. Recently the problem of estimating the state of a linear system with corrupted measurements has been studied [14]. More precisely, the maximum number of tolerable faulty sensors is characterized, and a decoding algorithm is proposed to detect corrupted measurements. Finally, security issues of specific cyber-physical systems have received considerable attention, such as power networks [15]–[19], linear networks with misbehaving components [20], [21], and water networks [22], [23].

Contributions. The contributions of this paper are as follows. First, we describe a unified modeling framework for cyber-physical systems and attacks (Section II). Motivated by existing cyber-physical systems and existing attack scenarios, we model a cyber-physical system under attack as a descriptor system subject to unknown inputs affecting the state and the measurements. For our model, we define the notions of *detectability* and *identifiability* of an attack by its effect on output measurements. Informed by the classic work on geometric control theory [24], [25], our framework includes the *deterministic static detection problem* considered in [9], [10], and the prototypical deception and denial of service [8], stealth [16], (dynamic) false-data injection [26], replay attacks [11], and covert attacks [12] as special cases.

Second, we show the fundamental limitations of a class of monitors (Section III-A). This class includes the widely-studied static, dynamic, and active monitors. We prove that (i) a cyber-physical attack is undetectable by our monitors if and only if the attackers' signal excites uniquely the zero dynamics of the input/output system, and (ii) that undetectable and unidentifiable attacks can be cast without knowing monitoring signals or the system noise.

Third, we provide a graph-theoretic characterization of undetectable attacks (Section III-B). We borrow some tools from the theory of structured systems, and we identify conditions on the system interconnection structure for the existence of undetectable attacks. These conditions are *generic*, in the sense that they hold for almost all numerical systems with the same structure, and they can be efficiently verified. As a complementary result, we extend a result of [27] on structural left-invertibility to regular descriptor systems. Finally, with respect to our earlier work [20], [21], we consider continuous-time descriptor systems, and we include parameters constraints.

Fourth, we design centralized and distributed monitors (Section IV). Our centralized monitors and our distributed detection monitor are complete, in the sense that they detect and identify every (detectable and identifiable) attack. Our centralized monitors are designed by leveraging on tools from geometric control theory, while our distributed detection monitor relies upon techniques from distributed control and parallel computation. Additionally, we characterize the computational complexity of the attack identification problem.

Fifth and finally, we illustrate the potential impact of our theoretical findings through compelling examples. In particular, (i) we design an undetectable state attack to destabilize the WSSC 3-machine 6-bus power system, (ii) we characterize the resilience to output attacks of the IEEE 14 bus system, (iii) we show the detection performance of our distributed monitor on the IEEE 118 bus system, and (iv) we use the RTS 96 network model to illustrate that our methods are effective also in the presence of system noise, nonlinearities, and modeling uncertainties. Through these examples we show the advantages of dynamic monitors against static ones, and we provide insight on the design of attacks.

II. Problem Setup and Preliminary Results

In this paper we model cyber-physical systems under attack as linear time-invariant descriptor systems subject to unknown inputs. This simplified model neglects system nonlinearities and the presence of noise in the dynamics and the measurements. Nevertheless, such a simplified model has long proven useful in studying stability, faults, and attacks in, for instance, power networks, sensor networks, and water networks. It is our premise that more detailed models are unlikely to change the basic conclusions of this work.

Model of cyber-physical systems under attack. We consider the descriptor system[1]

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, $E \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, and $D \in \mathbb{R}^{p \times m}$. Here the matrix E is possibly singular, and the inputs Bu and Du are unknown signals describing disturbances affecting the plant. Besides reflecting the genuine failure of systems components, these disturbances model the effect of attacks against the cyber-physical system. Without loss of generality, we assume that each state and output variable can be independently compromised by an attacker, and we let $B = [I_{n \times n} \; 0_{n \times p}]$ and $D = [0_{p \times n} \; I_{p \times p}]$.

The attack signal $t \mapsto u(t) \in \mathbb{R}^{n+p}$ depends upon the specific attack strategy. In particular, if $K \subseteq \{1, \dots, n+p\}$ is the *attack set*, with $|K| = k$, then all (and only) the entries of u indexed by K are nonzero over time, that is, for each $i \in K$, there exists a time t such that $u_i(t) \neq 0$,

[1] The results stated in this paper for continuous-time descriptor systems hold also for discrete-time descriptor systems and nonsingular systems. Moreover, due to linearity of (1), known inputs do not affect our results.

and $u_j(t) = 0$ for all $j \notin K$ and at all times. To underline this sparsity relation, we sometimes use u_K to denote the attack signal, that is the subvector of u indexed by K. Accordingly, the pair (B_K, D_K), where B_K and D_K are the submatrices of B and D with columns in K, denote the *attack signature*. Hence, $Bu(t) = B_K u_K(t)$, and $Du(t) = D_K u_K(t)$. Since the matrix E may be singular, we make the following assumptions on system (1):

(A1) the pair (E, A) is regular, that is, the determinant $|sE - A|$ does not vanish identically;

(A2) the initial condition $x(0) \in \mathbb{R}^n$ is consistent, that is, $(Ax(0) + Bu(0)) \in \operatorname{Im}(E)$; and

(A3) the input signal $u(t)$ is smooth.

Assumption (A1) assures the existence of a unique solution $x(t)$ to (1). Assumptions (A2) and (A3) guarantee smoothness of the state trajectory $x(t)$ and the measurements $y(t)$, [28, Lemma 2.5]. If assumptions (A2) and (A3) are dropped, then there are inconsistent initial conditions and impulsive inputs by which a powerful attacker can avoid detection; see Remark 4. Throughout the paper, the cardinality k of the attack set, or an upper bound, is assumed to be known.

Remark 1: ***(Examples of cyber-physical systems requiring advanced security mechanisms)*** Future power grids will combine physical dynamics with a sophisticated coordination infrastructure. The cyber-physical security of the grid has been identified as an issue of primary concern, see [19], [29] and [10], [16]–[18], [30], [31].

Mass transport networks are cyber-physical systems, such as gas transmission and distribution networks [32], large-scale process engineering plants [33], and water networks. Examples of water networks include open channel flows [34] for irrigation purposes and municipal water networks [35], [36]. The vulnerability of open channel networks to cyber-physical attacks has been studied in [12], [22], and municipal water networks are also known to be susceptible to attacks on the hydraulics [1] and biochemical contamination threats [23].

Power networks and mass transport network under attack can be modeled by descriptor systems with unknown inputs. For instance, the small-signal version of the classical structure-preserving power network model reads as [30], [31]

$$\begin{bmatrix} I & 0 & 0 \\ 0 & M_{\mathrm{g}} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\delta}(t) \\ \dot{\omega}(t) \\ \dot{\theta}(t) \end{bmatrix} = - \begin{bmatrix} 0 & -I & 0 \\ \mathcal{L}_{\mathrm{gg}} & D_{\mathrm{g}} & \mathcal{L}_{\mathrm{gl}} \\ \mathcal{L}_{\mathrm{lg}} & 0 & \mathcal{L}_{\mathrm{ll}} \end{bmatrix} \begin{bmatrix} \delta(t) \\ \omega(t) \\ \theta(t) \end{bmatrix} + \begin{bmatrix} 0 \\ P_{\omega}(t) \\ P_{\theta}(t) \end{bmatrix}, \qquad (2)$$

where δ and ω denote the generator rotor angles and frequencies, θ are the voltage angles at the buses, $\mathcal{L} = \left[\begin{smallmatrix} \mathcal{L}_{\mathrm{gg}} & \mathcal{L}_{\mathrm{gl}} \\ \mathcal{L}_{\mathrm{lg}} & \mathcal{L}_{\mathrm{ll}} \end{smallmatrix}\right]$ is the network susceptance matrix, M_{g} and D_{g} are the diagonal matrices of the generator inertial and damping coefficients, and P_ω and P_δ are power injections at the generators and buses. We refer to [35], [36] for the modeling of water networks. □

Model of monitors. A *monitor* is a deterministic algorithm $\Phi : \Lambda \rightarrow \Psi$ with access to continuous-time measurements and knowledge of the system dynamics, that is, $\Lambda = \{E, A, C, y(t) \ \forall t \in \mathbb{R}_{\geq 0}\}$. The output of a monitor is $\Psi = \{\psi_1, \psi_2\}$, with $\psi_1 \in \{\text{True}, \text{False}\}$, and $\psi_2 \subseteq \{1, \dots, n+p\}$. In particular, the output ψ_1 reveals the presence of attacks, while ψ_2 corresponds to the attack set.

Let $y(x, u, t)$ be the output signal of (1) generated from the initial state x by the attack input u. Then, the monitoring input $y(t)$ equals $y(x_0, u_K, t)$ at all times, where x_0 is the system initial state and u_K is the attack signal of the attack set K. Since we only consider deterministic cyber-physical systems, we assume monitors to be *consistent*, that is,

(i) $\psi_1 =$ True *only if* the attack set K is nonempty ($\psi_1 =$ False, otherwise),

(ii) $\psi_2 = \emptyset$ *if and only if* $\psi_1 =$ False, and

(iii) $\psi_2 = K$ *only if* K is the (unique) smallest subset $S \subseteq \{1, \dots, n+p\}$ satisfying $y(t) = y(x_1, u_S, t)$ for some initial state x_1 and at all times $t \in \mathbb{R}_{\geq 0}$ ($\psi_2 = \{1, \dots, n+p\}$, otherwise).

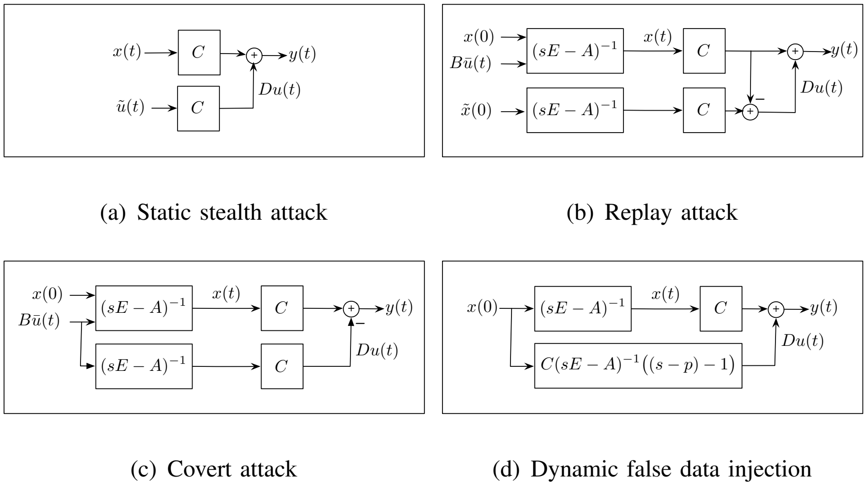
Observe that, if $S = \{1, \dots, n+p\}$, then there always exists an attack signal u_S satisfying $y(t) = y(x_0, u_K, t) = y(x_1, u_S, t)$. Our consistency assumption ensures that false-alarms are not triggered by our monitors. Examples of monitors can be found in [10], [11], [17].

The objective of a monitor is twofold:

Definition 1: ***(Attack detection and identification)*** Consider system (1) with nonzero attack $(B_K u_K, D_K u_K)$. The attack $(B_K u_K, D_K u_K)$ is *detected* by a monitor Φ if $\psi_1 =$ True. The attack $(B_K u_K, D_K u_K)$ is *identified* by the monitor Φ if $\psi_2 = K$.

An attack is *undetectable* (respectively *unidentifiable*) if no monitor detects (respectively identifies) the attack. Of course, an undetectable attack is also unidentifiable, since it cannot be distinguished from the zero attack. An attack set K is undetectable (respectively unidentifiable) if there exists an undetectable (respectively unidentifiable) attack $(B_K u_K, D_K u_K)$.

Model of attacks. In this work we consider colluding omniscient attackers with the ability of altering the cyber-physical dynamics through exogenous inputs. In particular, we let the attack $(Bu(t), Du(t))$ in (1) be designed based on knowledge of the system structure and parameters E, A, C, and the full state $x(t)$ at all times. Additionally, attackers have unlimited computation



(a) Static stealth attack

(b) Replay attack

(c) Covert attack

(d) Dynamic false data injection

Fig. 1. A block diagram illustration of prototypical attacks is here reported. In Fig. 1(a) the attacker corrupts the measurements y with the signal $D_K u_K \in \operatorname{Im}(C)$. Notice that in this attack the dynamics of the system are not considered. In Fig. 1(b) the attacker affects the output so that $y(t) = y(x(0), [\bar{u}^\mathsf{T} \; u^\mathsf{T}]^\mathsf{T}, t) = y(\tilde{x}(0), 0, t)$. The covert attack in Fig. 1(c) is a feedback version of the replay attack, and it can be explained analogously. In Fig. 1(d) the attack is such that the unstable pole p is made unobservable.

capabilities, and their objective is to disrupt the physical state or the measurements while avoiding detection or identification. Note that specific attacks may be cast by possibly-weaker attackers.

Remark 2: ***(Existing attack strategies as subcases)*** The following prototypical attacks can be modeled and analyzed through our theoretical framework:

(i) *stealth attacks* defined in [16] correspond to output attacks compatible with the measurements equation;

(ii) *replay attacks* defined in [11] are state and output attacks which affect the system dynamics and reset the measurements;

(iii) *covert attacks* defined in [12] are closed-loop replay attacks, where the output attack is chosen to cancel out the effect on measurements of the state attack; and

(iv) *(dynamic) false-data injection attacks* defined in [26] are output attacks rendering an unstable mode (if any) of the system unobservable.

A possible implementation of the above attacks in our model is illustrated in Fig. 1. □

To conclude this section we remark that our modeling framework captures failures and attacks against power networks and water supply networks. Possible genuine failures include variations in demand and supply of power or water, line outages or pipe leakages, and failures of sensors and actuators. Possible cyber-physical attacks include measurements corruption [9], [10], [22] and attacks on the control architecture or the physical state [1], [15], [18], [19].

III. Fundamental Monitoring Limitations

In this section we highlight fundamental monitoring limitations from system-theoretic and graph-theoretic perspectives.

A. System-theoretic monitoring limitations

Following the discussion in Section II, an attack is undetectable if the measurements due to the attack coincide with the measurements due to some nominal operating condition.

Lemma 3.1: ***(Undetectable attack)*** For the descriptor system (1), the nonzero attack $(B_K u_K, D_K u_K)$ is undetectable if and only if $y(x_1, u_K, t) = y(x_2, 0, t)$ for some initial states $x_1, x_2 \in \mathbb{R}^n$ and for all $t \in \mathbb{R}_{\geq 0}$.

Proof: (If) Let $y(x_1, u_K, t) = y(x_2, 0, t)$. Since monitors are deterministic, the monitor inputs $y_1(t) = y(x_1, u_K, t)$ and $y_2(t) = y(x_2, 0, t)$ yield the same output $\{\psi_1, \psi_2\}$. Since monitors are consistent, we have $\psi_1 =$ False for the input y_2. Hence, $\psi_1 =$ False also for the input y_1, and the attack is undetectable.

(Only if) Suppose that $y(x_1, u_K, t) \neq y(x_2, 0, t)$ for every initial states x_1 and x_2. Then the attack $(B_K u_K, D_K u_K)$ is distinguishable from nominal operating conditions via the system output. Hence, the attack is detectable. See Section IV for a complete detection monitor. ■

Analogous to detectability, the identifiability of an attack is the possibility to distinguish from measurements between the action of two distinct attacks. We measure the strength of an attack through the cardinality of the corresponding attack set. Since an attacker can independently compromise any state variable or measurement, every subset of the states and measurements of fixed cardinality is a potential attack set.

Lemma 3.2: ***(Unidentifiable attack)*** For the descriptor system (1), the nonzero attack $(B_K u_K, D_K u_K)$ is unidentifiable if and only if $y(x_1, u_K, t) = y(x_2, u_R, t)$ for some initial states $x_1, x_2 \in \mathbb{R}^n$, attack $(B_R u_R, D_R u_R)$ with $|R| \leq |K|$ and $R \neq K$, and for all $t \in \mathbb{R}_{\geq 0}$.

A proof of Lemma 3.2 follows the same reasoning as the proof of Lemma 3.1. We now elaborate on the above lemmas to derive fundamental detection and identification limitations. For a vector $x \in \mathbb{R}^n$, let $\text{supp}(x) = \{i \in \{1, \dots, n\} : x_i \neq 0\}$, and let $\|x\|_{\ell_0} = |\text{supp}(x)|$ denote the number of non-zero entries.

Theorem 3.3: ***(Detectability of cyber-physical attacks)*** For the descriptor system (1) and an attack set K, the following statements are equivalent:

(i) the attack set K is undetectable; and

(ii) there exist $s \in \mathbb{C}$, $g \in \mathbb{R}^{|K|}$, and $x \in \mathbb{R}^n$, with $x \neq 0$, such that $(sE - A)x - B_K g = 0$ and $Cx + D_K g = 0$.

Moreover, there exists an undetectable attack set K, with $|K| = k$, if and only if there exist $s \in \mathbb{C}$ and $x \in \mathbb{R}^n$ such that $\|(sE - A)x\|_0 + \|Cx\|_0 = k$.

Proof: By Lemma 3.1 and linearity of system (1), the attack u_K is undetectable if and only if there exists x_0 such that $y(x_0, u_K, t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$, that is, if and only if system (1) features zero dynamics. For a linear descriptor system with smooth input and consistent initial condition, the existence of zero dynamics is equivalent to the existence of invariant zeros as in (ii), see [28, Theorem 3.2 and Proposition 3.4]. The equivalence of statements (i) and (ii) follows. The last statement follows from (ii), and the fact that $B = [I, 0]$ and $D = [0, I]$. ■

Following Theorem 3.3, an attack $(B_K u_K, D_K u_K)$ is undetectable if it excites *only* zero dynamics for the dynamical system (1). Moreover, the existence of undetectable attacks for the attack set K is equivalent to the existence of *invariant zeros* for the system (E, A, B_K, C, D_K). For the notions of zero dynamics and invariant zeros we refer the reader to [25], [28]. The following theorem shows that analogous statements hold for the identifiability of attacks.

Theorem 3.4: ***(Identifiability of cyber-physical attacks)*** For the descriptor system (1) and an attack set K, the following statements are equivalent:

(i) the attack set K is unidentifiable; and

(ii) there exists an attack set R, with $|R| \leq |K|$ and $R \neq K$, $s \in \mathbb{C}$, $g_K \in \mathbb{R}^{|K|}$, $g_R \in \mathbb{R}^{|R|}$, and $x \in \mathbb{R}^n$, with $x \neq 0$, such that $(sE - A)x - B_K g_K - B_R g_R = 0$ and $Cx + D_K g_K + D_R g_R = 0$.

Moreover, there exists an unidentifiable attack set K, with $|K| = k \in \mathbb{N}_0$, if and only if there exists an undetectable attack set $\bar{K}$, with $|\bar{K}| \leq 2k$.

Proof: Due to linearity of the system (1), the unidentifiability condition in Lemma 3.2 is equivalent to the condition $y(x_K - x_R, u_K - u_R, t) = 0$, for some initial conditions x_K, x_R, and attack signals u_K, u_R. The equivalence between statements (i) and (ii) follows analogously to the proof of Theorem 3.3. Finally, the last statement follows from Theorem 3.3, and the fact that $B = [I, 0]$ and $D = [0, I]$. ■

Theorem 3.4 shows that the existence of an unidentifiable attack set K of cardinality k is equivalent to the existence of invariant zeros for the system $(E, A, B_{\bar{K}}, C, D_{\bar{K}})$, with $|\bar{K}| \leq 2k$.

Remark 3: ***(Static and active monitors, and noisy dynamics)*** A particular monitor is the so-called *static monitor* which verifies the consistency of the measurements without knowledge of the system dynamics and without exploiting relations among measurements taken at discrete time instants. For instance, the *bad data detector* in [9], [37] is a static monitor. Then, an attack $(B_K u_K, D_K u_K)$ is undetectable by a static monitor if and only if, for some state trajectory $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ and for all times $t \in \mathbb{N}$ it holds $Cx(t) + D_K u_K(t) = 0$. Note that state attacks are undetectable by static monitors [17].

An *active monitor* injects an auxiliary input (Bv, Dv) to reveals attacks [11]. Since auxiliary inputs do not alter the invariant zeros of system (1), active monitors share the same fundamental limitations of our monitors.

An analogous reasoning shows that the existence of undetectable attacks for a noise-free system implies the existence of undetectable attacks for the same system driven by noise. The converse does not hold, since attackers may remain undetected by injecting a signal compatible with the noise statistics. □

Remark 4: ***(Inconsistent initial conditions and impulsive attacks)*** If the consistency assumption (A2) is dropped, then discontinuities in the state $x(t \downarrow 0)$ may affect the measurements $y(t \downarrow 0)$. For instance, for index-one systems, inconsistent initial conditions lead to initial jumps for the algebraic equations to be satisfied. Consequently, the inconsistent initial value $[0^\mathsf{T} \; x_2(0)^\mathsf{T}]^\mathsf{T} \in \operatorname{Ker}(E)$ cannot be recovered through measurements.

Assumption (A3) requires the attack signal to be sufficiently smooth such that $x(t)$ and $y(t)$ are at least continuous. Suppose that assumption (A3) is dropped and the input $u(t)$ belongs to the class of impulsive smooth distributions $\mathcal{C}_{\text{imp}} = \mathcal{C}_{\text{smooth}} \cup \mathcal{C}_{\text{p-imp}}$, that is, loosely speaking, the class of functions given by the linear combination of a smooth function on $\mathbb{R}_{\geq 0}$ (denoted by $\mathcal{C}_{\text{smooth}}$) and Dirac impulses and their derivatives at $t = 0$ (denoted by $\mathcal{C}_{\text{p-imp}}$) [28]. In this case, an attacker commanding an impulsive input can reset the initial state and evade detection.

The discussion in the previous two paragraphs can be formalized as follows. Let $\mathcal{V}_c$ be the subspace of points $x_0 \in \mathbb{R}^n$ of consistent initial conditions for which there exists an input $u \in \mathcal{C}^m_{\text{smooth}}$ and a state trajectory $x \in \mathcal{C}^n_{\text{smooth}}$ to the descriptor system (1) such that $y(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$. Let $\mathcal{V}_d$ (respectively $\mathcal{W}$) be the subspace of points $x_0 \in \mathbb{R}^n$ for which there exists an input $u \in \mathcal{C}_{\text{imp}}^{n+p}$ (respectively $u \in \mathcal{C}_{\text{p-imp}}^{n+p}$) and a state trajectory $x \in \mathcal{C}_{\text{imp}}^{n}$ (respectively $x \in \mathcal{C}_{\text{p-imp}}^{n}$) to the descriptor system (1) such that $y(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$. From [28, Theorem 3.2 and Proposition 3.4] it is known that $\mathcal{V}_d = \mathcal{V}_c + \mathcal{W} + \text{Ker}(E)$.

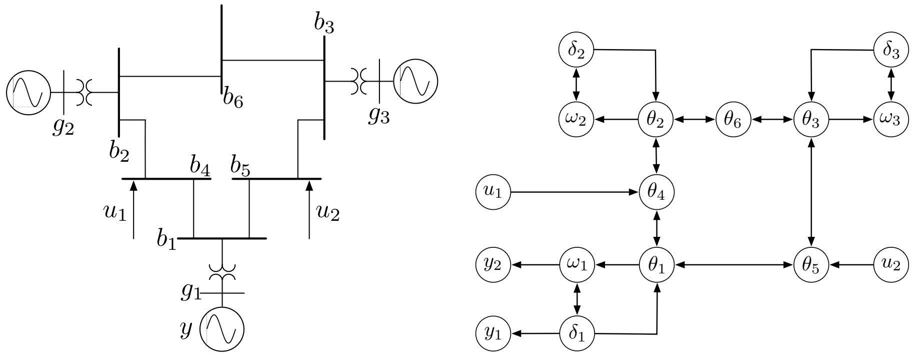
In this work we focus on the smooth output-nulling subspace $\mathcal{V}_c$, which is exactly space of zero dynamics identified in Theorems 3.3 and 3.4. Hence, for inconsistent initial conditions, the results presented in this section are valid only for strictly positive times. On the other hand, if an attacker injects impulsive signals, then it can avoid detection for initial conditions in $\mathcal{W}$. □

B. Graph-theoretic monitoring limitations

In this section we derive detectability conditions based upon a connectivity property of a graph associated with the dynamical system. For the ease of notation, in this subsection we drop the subscript K from B_K, D_K, and u_K. Let $([E], [A], [B], [C], [D])$ be the tuple of structure matrices associated with the system (1) [27]. We associate a directed input/state/output graph $\mathcal{G}_{\text{iso}} = (\mathcal{V}, \mathcal{E})$ with ($[E]$,$[A]$,$[B]$,$[C]$,$[D]$). The vertex set $\mathcal{V} = \mathcal{U} \cup \mathcal{X} \cup \mathcal{Y}$ consists of input, state, and output vertices given by $\mathcal{U} = \{u_1, \dots, u_m\}$, $\mathcal{X} = \{x_1, \dots, x_n\}$, and $\mathcal{Y} = \{y_1, \dots, y_p\}$, respectively. The set of directed edges $\mathcal{E}$ is $\mathcal{E}_{[E]} \cup \mathcal{E}_{[A]} \cup \mathcal{E}_{[B]} \cup \mathcal{E}_{[C]} \cup \mathcal{E}_{[D]}$, where $\mathcal{E}_{[E]} = \{(x_j, x_i) : [E]_{ij} \neq 0\}$, $\mathcal{E}_{[A]} = \{(x_j, x_i) : [A]_{ij} \neq 0\}$, $\mathcal{E}_{[B]} = \{(u_j, x_i) : [B]_{ij} \neq 0\}$, $\mathcal{E}_{[C]} = \{(x_j, y_i) : [C]_{ij} \neq 0\}$, and $\mathcal{E}_{[D]} = \{(u_j, y_i) : [D]_{ij} \neq 0\}$. In the latter, the expression $[E]_{ij} \neq 0$ means that the (i, j)-th entry of $[E]$ is a free parameter. For the graph $\mathcal{G}_{\text{iso}}$, a set of l mutually disjoint and simple paths between two sets of vertices S_1, S_2 is called *linking* of size l from S_1 to S_2. Finally, the matrix $s[E]-[A]$ is *structurally non-degenerate* if the determinant $|sE-A| \neq 0$ for a *generic* realization of E and A, that is, $|sE - A| \neq 0$ holds in the whole parameter space of elements of E and A with exception of a low dimensional variety [24], [38].

Example 1: ***(Power network structural analysis)*** Consider the power network illustrated in Fig. 2(a), where, being e_i the i-th canonical vector, we take $[E] = \text{blkdiag}(1, 1, 1, M_1, M_2, M_3, 0, 0, 0, 0, 0, 0)$, $[B] = [e_8\ e_9]$, $[C] = [e_1\ e_4]^{\mathsf{T}}$, $[D] = 0$, and $[A]$ equal to

$$\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
a_{4,1} & 0 & 0 & a_{4,4} & 0 & 0 & a_{4,7} & 0 & 0 & 0 & 0 & 0 \\
0 & a_{5,2} & 0 & 0 & a_{5,5} & 0 & 0 & a_{5,8} & 0 & 0 & 0 & 0 \\
0 & 0 & a_{6,3} & 0 & 0 & a_{6,6} & 0 & 0 & a_{6,9} & 0 & 0 & 0 \\
a_{7,1} & 0 & 0 & 0 & 0 & 0 & a_{7,7} & 0 & 0 & a_{7,10} & a_{7,11} & 0 \\
0 & a_{8,2} & 0 & 0 & 0 & 0 & 0 & a_{8,8} & 0 & a_{8,10} & 0 & a_{8,12} \\
0 & 0 & a_{9,3} & 0 & 0 & 0 & 0 & 0 & a_{9,9} & 0 & a_{9,11} & a_{9,12} \\
0 & 0 & 0 & 0 & 0 & 0 & a_{10,7} & a_{10,8} & 0 & a_{10,10} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_{11,7} & 0 & a_{11,9} & 0 & a_{11,11} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12,8} & a_{12,9} & 0 & 0 & a_{12,12}
\end{bmatrix}$$



(a) WSSC power system

(b) Input/output graph $\mathcal{G}_{\text{iso}}$

Fig. 2. Fig. 2(a) shows the WSSC power system with 3 generators and 6 buses. The numerical value of the network parameters can be found in [30]. The digraph associated with the network in Fig. 2(a). The self-loops of the vertices $\{\delta_1, \delta_2, \delta_3\}$, $\{\omega_1, \omega_2, \omega_3\}$, and $\{\theta_1, \dots, \theta_6\}$ are not drawn. The inputs u_1 and u_2 affect respectively the bus b_4 and the bus b_5. The measured variables are the rotor angle and frequency of the first generator.

The digraph associated with the structure matrices $([E], [A], [B], [C], [D])$ is in Fig. 2(b). □

Recall from Lemma 3.1 that an attack u is undetectable if $y(x_1, u, t) = y(x_2, 0, t)$ for some initial states x_1 and x_2. In the following result, we consider the particular case that the system initial state is known. Hence, an attack u is undetectable if $y(x_0, u, t) = y(x_0, 0, t)$ for some initial state x_0. Equivalently, the system fails to be left-invertible [25]. We say that the structured system $([E], [A], [B], [C], [D])$ is *structurally left-invertible* if every admissible realization (E, A, B, C, D) is left-invertible, with exception, possibly, of a low dimensional variety.

Theorem 3.5: ***(Structurally undetectable attack)*** Let the parameters space of the structured system $([E], [A], [B], [C], [D])$ define a polytope in $\mathbb{R}^d$ for some $d \in \mathbb{N}_0$. Assume that $s[E] - [A]$ is structurally non-degenerate. The system $([E], [A], [B], [C], [D])$ is structurally left-invertible if and only if there exists a linking of size $|\mathcal{U}|$ from $\mathcal{U}$ to $\mathcal{Y}$.

Theorem 3.5 extends the structural left-invertibility results known for nonsingular systems to regular descriptor systems, and its proof relies on classical concepts from structural analysis, algebraic geometry, and graph theory. Additionally, Theorem 3.5 gives a characterization of structurally undetectable attacks. The following result is useful to prove Theorem 3.5.

Lemma 3.6: ***(Polytopes and algebraic varieties)*** Let $S \subseteq \mathbb{R}^d$ be a polytope, and let $T \subseteq \mathbb{R}^d$ be an algebraic variety. Then, either the set $S \subseteq T$, or the set $S \setminus (S \cap T)$ is generic in S.

Proof: Let $T \subseteq \mathbb{R}^d$ be the algebraic variety described by the locus of common zeros of the polynomials $\{\phi_1(x), \dots, \phi_t(x)\}$, with $t \in \mathbb{N}$, $t < \infty$. Then $S \subseteq T$ if and only if every polynomial ϕ_i vanishes identically on S. Suppose that some polynomials, say ϕ_i, do not vanish identically on S. Then, $S \cap T \neq S$, and $S \cap T = \{x \in S : \phi_i(x) = 0\}$ is nowhere dense in S, since its closure has empty interior [39]. Hence, $S \cap T$ is a meagre subset of S, and its complement $S \setminus (S \cap T)$ is a generic subset of S [39]. ■

In Lemma 3.6 interpret the polytope S as the admissible parameters space of a structured cyber-physical system. Then we have shown that left-invertibility of a cyber-physical system is a generic property even when the admissible parameters space is a polytope of the whole parameters space. Consequently, for a structured cyber-physical system, if the initial state is known, either every admissible realization admits undetectable attacks, or there is no undetectable attack for every realization, except possibly for those lying on a low dimensional variety.

Proof of Theorem 3.5: Because of Lemma 3.6, we need to show that, if there are $|\mathcal{U}|$ disjoint paths from $\mathcal{U}$ to $\mathcal{Y}$, then there exists admissible left-invertible realizations. Conversely, if there are at most $|\mathcal{U}|-1$ disjoint paths from $\mathcal{U}$ to $\mathcal{Y}$, then every admissible realization is not left-invertible.

(If) Let (E, A, B, C, D), with $|sE - A| \neq 0$, be an admissible realization, and suppose there exists a linking of size $|\mathcal{U}|$ from $\mathcal{U}$ to $\mathcal{Y}$. Notice that $|\mathcal{Y}| \geq |\mathcal{U}|$, and select $|\mathcal{U}|$ outputs on a linking of size $|\mathcal{U}|$ from $\mathcal{U}$ to $\mathcal{Y}$ (let $\bar{C}$ and $\bar{D}$ be the submatrices of C and D associated with the smaller set of outputs). Observe that left-invertibility of $(E, A, B, \bar{C}, \bar{D})$ implies left-invertibility of (E, A, B, C, D). For the left-invertibility of $(E, A, B, \bar{C}, \bar{D})$ we need

$$\left| \begin{bmatrix} sE - A & -B \\ \bar{C} & \bar{D} \end{bmatrix} \right| = |sE - A| \left| \bar{D} + \bar{C}(sE - A)^{-1} B \right| \neq 0,$$

and hence we need $\left|\bar{D} + \bar{C}(sE - A)^{-1}B\right| \neq 0$. Notice that $\bar{D} + \bar{C}(sE - A)^{-1}B$ corresponds to the transfer matrix of the cyber-physical system. Since there are $|\mathcal{U}|$ independent paths from $\mathcal{U}$ to $\mathcal{Y}$, the matrix $\bar{D} + \bar{C}(sE - A)^{-1}B$ can be made nonsingular and diagonal by removing some connection lines from the network. In particular, for a given linking of size $|\mathcal{U}|$ from $\mathcal{U}$ to $\mathcal{Y}$, a nonsingular and diagonal transfer matrix is obtained by setting to zero the entries of E and A corresponding to the edges not in the linking. Then there exist admissible left-invertible realizations, and thus the systems $([E], [A], [D], [\bar{C}], [\bar{D}])$ and $([E], [A], [D], [C], [D])$ are structurally left-invertible.

(Only if) Take any subset of $|\mathcal{U}|$ output vertices, and let the maximum size of a linking from $\mathcal{U}$ to $\mathcal{Y}$ be smaller than $|\mathcal{U}|$. Let $[\bar{E}]$ and $[\bar{A}]$ be such that $s[\bar{E}] - [\bar{A}] = \left[\begin{smallmatrix} s[E]-[A] & [B] \\ [\bar{C}] & [\bar{D}] \end{smallmatrix}\right]$, where $[\bar{C}]$ and $[\bar{D}]$ are the structured output matrices corresponding to the chosen $|\mathcal{U}|$ output vertices. Consider the graph $G(s[\bar{E}] - [\bar{A}])$, that consists of $N = |\mathcal{X}| + |\mathcal{U}|$ vertices, and an edge from vertex j to i if $\bar{A}_{ij} \neq 0$ or $\bar{E}_{ij} \neq 0$. Notice that a path from $\mathcal{U}$ to $\mathcal{Y}$ in the digraph associated with the structured system corresponds, possibly after relabeling the output variables, to a cycle in involving input/output vertices in $G(s[\bar{E}]-[\bar{A}])$. Observe that there are only $|\mathcal{U}|-1$ such (disjoint) cycles. Hence, there is no cycle family of length N, and the system $([E], [A], [B], [\bar{C}], [\bar{D}])$ fails to be structurally left-invertible [40, Theorem 1]. Since the same reasoning holds for every set of $|\mathcal{U}|$ output vertices, every realization of the pencil $\left[\begin{smallmatrix} s[E]-[A] & [B] \\ [C] & [D] \end{smallmatrix}\right]$ has no invertible minor of size N, and the claimed statement follows. ■

If the system initial state is unknown, then an undetectable attack u is characterized by the existence of a pair of initial conditions x_1 and x_2 such that $y(x_1, 0, t) = y(x_2, u, t)$, or, equivalently, by the existence of invariant zeros for the given cyber-physical system. We will now show that, provided that a cyber-physical system is left-invertible, its invariant zeros can be computed by simply looking at an associated nonsingular state space system. Let the state vector x of the descriptor system (1) be partitioned as $[x_1^\mathsf{T}\ x_2^\mathsf{T}]^\mathsf{T}$, where x_1 corresponds to the dynamic variables. Let the network matrices E, A, B, C, and D be partitioned accordingly, and assume that the descriptor system (1) is given in semi-explicit form, that is, $E = \text{blkdiag}(E_{11}, 0)$, where E_{11} is nonsingular.[2] In this case, the descriptor system (1) reads as

$$\begin{aligned} E_{11}\dot{x}_1(t) &= A_{11}x_1(t) + B_1u(t) + A_{12}x_2(t)\,, \\ 0 &= A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t)\,, \\ y(t) &= C_1x_1(t) + C_2x_2(t) + Du(t)\,. \end{aligned} \tag{3}$$

Consider now the associated nonsingular state space system which is obtained by regarding $x_2(t)$ as an external input to the descriptor system (3) and the algebraic constraint as output:

$$\begin{aligned} \dot{x}_1(t) &= E_{11}^{-1}A_{11}x_1(t) + E_{11}^{-1}B_1u(t) + E_{11}^{-1}A_{12}x_2(t), \\ \tilde{y}(t) &= \begin{bmatrix} A_{21} \\ C_1 \end{bmatrix} x_1(t) + \begin{bmatrix} A_{22} & B_2 \\ C_2 & D \end{bmatrix} \begin{bmatrix} x_2(t) \\ u(t) \end{bmatrix}\,. \end{aligned} \tag{4}$$

[2] Interesting cyber-physical systems, such as power and mass-transport networks (2), are readily given in semi-explicit form.

Theorem 3.7: ***(Equivalence of invariant zeros)*** Consider the structurally left-invertible system $([E],[A],[B],[C],[D])$. The invariant zeros of every admissible realization (3) coincide with those of the associated nonsingular system (4), except, possibly, for realizations lying on a low dimensional variety of the parameters space.

Proof: In the interest of space we omit the proof, which follows from Theorem 3.5, [41, Proposition 8.4] and a manipulation of the system pencil. ■

Following Theorem 3.7, under the assumption of structural left-invertibility, classical results can be used to investigate the presence of undetectable attacks in structured system with unknown initial state; see [38] for a survey of results on generic properties of linear systems.

IV. Monitor Design for Attack Detection and Identification

We now design centralized and distributed filters for attack detection and identification.

A. Centralized attack detection

The output of the attack detection filters developed in this subsection will be a residual signal $r : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$. If each monitor is equipped with such an attack detection filter and if the attack is detectable, then the outputs of the monitor and the filter are related as follows: $\psi_1 =$ True if and only if $r(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$. We next present a centralized attack detection filter based on a modified Luenberger observer.

Theorem 4.1: ***(Centralized attack detection filter)*** Consider the descriptor system (1) and assume that the attack set K is detectable, and that the network initial state $x(0)$ is known. Consider the *centralized attack detection filter*

$$\begin{aligned} E\dot{w}(t) &= (A + GC)w(t) - Gy(t), \\ r(t) &= Cw(t) - y(t), \end{aligned} \tag{5}$$

where $w(0) = x(0)$ and the output injection matrix $G \in \mathbb{R}^{n \times p}$ is such that the pair $(E, A+GC)$ is regular and Hurwitz.[3] Then $r(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$ if and only if $u_K(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$. Moreover, in the absence of attacks, the filter error $w - x$ is exponentially stable.

[3] For a regular pair (E, A), let $\sigma(E, A) = \{\lambda \ : \ \lambda \in \mathbb{C}, |\lambda| < \infty, |\lambda E - A| = 0\}$. The pair (E, A) is Hurwitz if real$(\lambda) < 0$ for each $\lambda \in \sigma(E, A)$.

Proof: Consider the error $e = w - x$ between the filter (5) and system (1). The error system with output $r(t)$ is

$$\begin{aligned} E\dot{e}(t) &= (A + GC)e(t) - (B_K + GD_K)u_K(t), \\ r(t) &= Ce(t) - D_K u_K(t), \end{aligned} \tag{6}$$

where $e(0) = 0$. To prove the theorem we show that the error system (6) has no invariant zeros, that is, $r(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$ if and only if $u_K(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$. Since the initial condition $x(0)$ and the input u_K are assumed to be consistent (A2) and non-impulsive (A3), the error system (6) has no invariant zeros if and only if [28, Proposition 3.4] there exists no triple $(s, \bar{w}, g_K) \in \mathbb{C} \times \mathbb{R}^n \times \mathbb{R}^p$ satisfying

$$\begin{bmatrix} sE - (A + GC) & B_K + GD_K \\ C & -D_K \end{bmatrix} \begin{bmatrix} \bar{w} \\ g_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{7}$$

The second equation of (7) yields $C\bar{w} = D_K g_K$. By substituting $C\bar{w}$ by $D_K g_K$ in the first equation of (7), we obtain

$$\begin{bmatrix} sE - A & B_K \\ C & -D_K \end{bmatrix} \begin{bmatrix} \bar{w} \\ g_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{8}$$

Note that a solution $(s, -\bar{w}, g_K)$ to (8) would yield an invariant zero, zero state, and zero input for the descriptor system (1). By the detectability assumption, the descriptor system (1) has no invariant zeros and the matrix pencil in (8) necessarily has full rank. It follows that the triple (E, A, C) is observable, G can be chosen to make the pair $(E, A + GC)$ Hurwitz [42], and the error system (6) is stable without zero dynamics. ■

Notice that, if the initial state $x(0)$ is not available, then, an arbitrary initial state $w(0) \in \mathbb{R}^n$ can be chosen. In this case, since $(E, A+GC)$ is Hurwitz, the filter (5) converges asymptotically, and some attacks may remain undetected. Also, if the dynamics and the measurements of (1) are affected by modeling uncertainties and noise with known statistics, then the output injection matrix G should be chosen to optimize the sensitivity of the residual r to attacks versus the effect of noise. Statistical testing techniques can [7] subsequently be used to analyze the residual r. Finally, attacks aligned with the noise statistics may remain undetected.

B. Distributed attack detection

Notice that a direct implementation of the filter (5) requires continuous communication of measurements to a central processor, which needs to integrate the possibly large-scale system

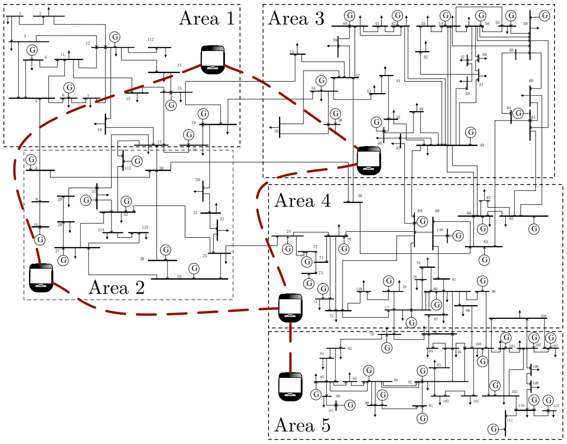


Fig. 3. Partition of IEEE 118 bus system into 5 areas. Each area is monitored and operated by a control center. These control centers cooperate to estimate the state and to assess the functionality of the whole network.

(5). In what follows, we will exploit the sparsity of the filter matrices (E, A, C) to develop a distributed detection filter.

Assume that control centers are geographically deployed in a large scale cyber-physical system to operate the whole plant via distributed computation; see Fig. 3. Let $\mathcal{G}_\text{s} = (\mathcal{V}, \mathcal{E})$ be the directed sparsity graph associated with the pair (E, A), where the vertex set $\mathcal{V} = \mathcal{X}$ corresponds to the system state, and the set of directed edges $\mathcal{E} = \{(x_j, x_i) : e_{ij} \neq 0 \text{ or } a_{ij} \neq 0\}$ is induced by the sparsity pattern of E and A. Let $\mathcal{V}$ be partitioned into N disjoint subsets as $\mathcal{V} = \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_N$, with $|\mathcal{V}_i| = n_i$, and let $\mathcal{G}_\text{s}^i = (\mathcal{V}_i, \mathcal{E}_i)$ be the i-th subgraph of $\mathcal{G}_\text{s}$ with vertices $\mathcal{V}_i$ and edges $\mathcal{E}_i = \mathcal{E} \cap (\mathcal{V}_i \times \mathcal{V}_i)$. According to this partition, and possibly after relabeling the states, the system matrix A in (1) can be written as

$$A = \begin{bmatrix} A_1 & \cdots & A_{1N} \\ \vdots & \vdots & \vdots \\ A_{N1} & \cdots & A_N \end{bmatrix} = A_D + A_C,$$

where $A_i \in \mathbb{R}^{n_i \times n_i}$, $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, and $A_D = \text{blkdiag}(A_1, \ldots, A_N)$. We make the following assumptions:

(A4) the matrices E, C are block-diagonal, that is, $E = \text{blkdiag}(E_1, \ldots, E_N)$, $C = \text{blkdiag}(C_1, \ldots, C_N)$, where $E_i \in \mathbb{R}^{n_i \times n_i}$ and $C_i \in \mathbb{R}^{p_i \times n_i}$; and

(A5) each pair (E_i, A_i) is regular, and each triple (E_i, A_i, C_i) is observable.

Given the above structure and in the absence of attacks, the descriptor system (1) can be written as the interconnection of N subsystems of the form

$$\begin{aligned} E_i\dot{x}_i(t) &= A_i x_i(t) + \sum\nolimits_{j\in\mathcal{N}_i^{\text{in}}} A_{ij}x_j(t), \\ y_i(t) &= C_i x_i(t), \;\; i \in \{1,\dots,N\}, \end{aligned} \tag{9}$$

where x_i and y_i are the state and output of the i-th subsystem and $\mathcal{N}_i^{\text{in}} = \{j \in \{1,\dots,N\} \setminus i \,:\, \|A_{ij}\| \neq 0\}$ are the in-neighbors of subsystem i. We also define the set of out-neighbors as $\mathcal{N}_i^{\text{out}} = \{j \in \{1,\dots,N\} \setminus i \,:\, \|A_{ji}\| \neq 0\}$. We assume the presence of a *control center* in each subnetwork $\mathcal{G}_{\text{s}}^i$ with the following capabilities:

(A6) the i-th control center knows the matrices E_i, A_i, C_i, as well as the neighboring matrices A_{ij}, $j \in \mathcal{N}_i^{\text{in}}$; and

(A7) the i-th control center can transmit an estimate of its state to the j-th control center if $j \in \mathcal{N}_i^{\text{out}}$.

Before presenting our distributed attack detection filter, we need the following result on a decentrally stabilized filter.

Lemma 4.2: ***(Decentralized stabilization of the attack detection filter)*** Consider the descriptor system (1), and assume that the attack set K is detectable and that the network initial state $x(0)$ is known. Consider the attack detection filter

$$\begin{aligned} E\dot{w}(t) &= (A_D + GC)w(t) + A_C w(t) - Gy(t)\,, \\ r(t) &= y(t) - Cw(t)\,. \end{aligned} \tag{10}$$

where $w(0) = x(0)$ and $G = \text{blkdiag}(G_1,\dots,G_N)$ is such that $(E, A_D + GC)$ is regular and Hurwitz. Assume that

$$\rho\left((\mathrm{j}\,\omega E - A_D - GC)^{-1}A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}\,, \tag{11}$$

where $\rho(\cdot)$ denotes the spectral radius operator. Then $r(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$ if and only if $u_K(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$. Moreover, in the absence of attacks, the filter error $w(t) - x(t)$ is exponentially stable.

Proof: The error $e(t) = w(t) - x(t)$ obeys the dynamics

$$\begin{aligned} E\dot{e}(t) &= (A_D + A_C + GC)e(t) - (B_K + GD_K)u_K(t), \\ r(t) &= Ce(t) - D_K u_K(t)\,. \end{aligned} \tag{12}$$

A reasoning analogous to that in the proof of Theorem 4.1 shows the absence of zero dynamics. Hence, for $r(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$ if and only if $u_K(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$.

To show stability of the error dynamics in the absence of attacks, we employ the small-gain approach to large-scale systems and rewrite the error dynamics (12) as the closed-loop interconnection of the two subsystems $\Gamma_1 : E\dot{e}(t) = (A_D + GC)e(t) + v(t)$ and $\Gamma_2 : v(t) = A_C e(t)$. When regarded as input-output systems with respective input/output pairs (v, e) and (e, v), both Γ_1 and Γ_2 are causal and internally stable. Hence, by [43, Theorem 4.11], the overall error dynamics (12) are stable if the loop transfer function $\Gamma_1(\mathrm{j}\,\omega) \cdot \Gamma_2$ satisfies the spectral radius condition $\rho(\Gamma_1(\mathrm{j}\,\omega) \cdot \Gamma_2) < 1$ for all $\omega \in \mathbb{R}$. The latter condition is equivalent to (11). ■

An implementation of the decentrally stabilized filter (10) under assumptions (A1)-(A7) requires the input $A_C w$ and hence continuous communication among control centers. To overcome this continuous communication obstacle we rely on waveform relaxation methods [44], [45] developed for parallel numerical integration. The Gauss-Jacobi waveform relaxation applied to the filter (10) yields the *waveform relaxation iteration*

$$E\dot{w}^{(k)}(t) = (A_D + GC)w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t), \tag{13}$$

where $k \in \mathbb{N}$ denotes the iteration index, $t \in [0, T]$ is the integration interval for some uniform time horizon $T > 0$, and $w^{(k)} : [0, T] \to \mathbb{R}^n$ is a trajectory with initial condition $w^{(k)}(0) = w_0$ for each $k \in \mathbb{N}$. Notice that (13) is a descriptor system with state $w^{(k)}$, and known input $A_C w^{(k-1)}$, since the value of $w(t)$ at iteration $k - 1$ is used. The iteration (13) is initialized with an initial profile $w^{(0)} : [0, T] \to \mathbb{R}^n$.

The iteration (13) is said to be (uniformly) *convergent* if

$$\lim_{k \to \infty} \max_{t \in [0,T]} \left\| w^{(k)}(t) - w(t) \right\|_{\infty} = 0\,, \tag{14}$$

where w is the solution of the non-iterative dynamics (10). In order to obtain a distributed detection scheme, we use the waveform relaxation iteration (13) to iteratively approximate the decentralized filter (10).

Theorem 4.3: ***(Distributed attack detection filter)*** Consider the descriptor system (1) and assume that the attack set K is detectable, and that the network initial state $x(0)$ is known. Let the assumptions (A1) through (A7) be satisfied and consider the *distributed attack detection* *filter*

$$\begin{aligned} E\dot{w}^{(k)}(t) &= \big(A_D + GC\big)w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t), \\ r^{(k)}(t) &= y(t) - Cw^{(k)}(t), \end{aligned} \tag{15}$$

where $k \in \mathbb{N}$, $t \in [0, T]$ for some $T > 0$, $w^{(k)}(0) = x(0)$ for all $k \in \mathbb{N}$, and $G = \text{blkdiag}(G_1, \dots, G_N)$ is such that the pair $(E, A_D + GC)$ is regular, Hurwitz, and

$$\rho\left((\mathrm{j}\,\omega E - A_D - GC)^{-1} A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}\,. \tag{16}$$

Then $\lim_{k\to\infty} \|r^{(k)}(t)\|_\infty = 0$ at all times $t \in [0, T]$ if and only if $u_K(t) = 0$ at all times $t \in [0, T]$. Moreover, in the absence of attacks, the asymptotic filter error $\lim_{k\to\infty}(w^{(k)}(t) - x(t))$ is exponentially stable for $t \in [0, T]$.

Proof: Since $w^{(k)}(0) = x(0)$, it follows from [45, Theorem 5.2] that the solution $w^{(k)}$ of the iteration (15) converges, as $k \to \infty$, to the solution w of (10) if

$$\rho\left(((\sigma + \mathrm{j}\,\omega)E - A_D - GC)^{-1} A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}, \tag{17}$$

where $\sigma = \max\{\alpha, \beta\}$, α is the least upper bound on the real part of the spectrum of (E, A), and β is such that the signal $f : [0, T] \to \mathbb{R}$, $f(t) = y(t)e^{-\beta t}$, and all its derivatives exist and are bounded. Since the pair $(E, A_D + GC)$ is Hurwitz and y is smooth by assumptions (A2) and (A3), we have that $\sigma \leq 0$, and the convergence condition (17) equals condition (16).

Hence, we have uniform convergence (in the sense of (14)) of the solution and output $\big(w^{(k)}, r^{(k)}\big)$ of the distributed filter (15) to the solution and output $\big(w, r\big)$ of the decentrally stabilized filter. Due to the detectability assumption, it follows from Lemma 4.2 that $\lim_{k\to\infty} \|r^{(k)}(t)\|_\infty = 0$ at all times $t \in [0, T]$ if and only if $u_K(t) = 0$ at all times $t \in [0, T]$.

Under condition (16) and due to the Hurwitz assumption, it follows from Lemma 4.2 that the error $e = w - x$ between the state w of the decentralized filter dynamics (10) and the state x of the descriptor model (1) is asymptotically stable in the absence of attacks. This concludes the proof of Theorem 4.3. ■

The waveform relaxation iteration (13) can be implemented in the following distributed fashion. Assume that control center i is able to numerically integrate the descriptor system

$$E_i \dot{w}_i^{(k)}(t) = (A_i + G_i C_i) w_i^{(k)}(t) + \sum\nolimits_{j \in \mathcal{N}_i^{\text{in}}} A_{ij} w_j^{(k-1)}(t) - G_i y_i(t), \tag{18}$$

over a time interval $t \in [0, T]$, with initial condition $w_i^{(k)}(0) = w_{i,0}$, measurements y_i, and the neighboring filter states $w_j^{(k-1)}$ as external inputs. Let $w_j^{(0)}$ be an initial guess of the signal w_j. Each control center performs the following operations assuming $k = 0$ at start:

(1) set $k := k + 1$, and compute the signal $w_i^{(k)}$ by integrating the local filter equation (18);

(2) transmit $w_i^{(k)}$ to the j-th control center if $j \in \mathcal{N}_i^{\text{out}}$;

(3) update the input $w_j^{(k)}$ with the signal received from the j-th control center, with $j \in \mathcal{N}_i^{\text{in}}$, and iterate.

Following Theorem 4.3, for k sufficiently large, the local residuals $r_i^{(k)} = y_i - C_i w_i^{(k)}$ can be used to detect attacks. A related large-scale example is given in Section V-C.

Remark 5: ***(Implementation of distributed attack detection filter)*** When implementing the distributed attack detection filter (15) in the interval $[0, T]$, control center i needs to transmit the signal $w_i^{(k)}(t)$ with $t \in [0, T]$ at each iteration k. In practice, only an approximation or a finite basis representation $\hat{w}_i^{(k)}(t)$ can be transmitted. The error due to this approximation can be characterized, and we refer the reader to [46]. □

C. Complexity of the attack identification problem

In this subsection we study the problem of attack identification, that is, the problem of identifying from measurements the state and output variables corrupted by the attacker. We start our discussion by showing that this problem is generally *NP-hard*. For a vector-valued signal $v : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, let $\|v\|_{\mathcal{L}_0} = |\cup_{t \in \mathbb{R}_{\geq 0}} \operatorname{supp}(v(t))|$, and consider the following cardinality minimization problem: given a descriptor system with matrices $E, A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{p \times n}$ and a measurement signal $y : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$, find the minimum cardinality input signals $v_x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ and $v_y : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$ and an arbitrary initial condition $\xi_0 \in \mathbb{R}^n$ that explain the data y, that is,

$$\begin{aligned} \min_{v_x, v_y, \xi_0} \quad & \|v_x\|_{\mathcal{L}_0} + \|v_y\|_{\mathcal{L}_0} \\ \text{subject to} \quad & E\dot{\xi}(t) = A\xi(t) + v_x(t), \\ & y(t) = C\xi(t) + v_y(t), \\ & \xi(0) = \xi_0 \in \mathbb{R}^n . \end{aligned} \tag{19}$$

Lemma 4.4: ***(Problem equivalence)*** Consider the system (1) with identifiable attack set K. The optimization problem (19) coincides with the problem of identifying the attack set K given the system matrices E, A, C, and the measurements y, where $K = \operatorname{supp}([v_x^\mathsf{T} \; v_y^\mathsf{T}])$.

Proof: Due to the identifiability of K, the attack identification problem consists of finding the smallest attack set capable of injecting an attack $(B_K u_K, D_K u_K)$ that generates the given measurements y for the given dynamics E, A, C, and some initial condition; see Lemma 3.2. The statement follows since $B = [I, 0]$ and $D = [0, I]$ in (1), so that $(B_K u_K, D_K u_K) = (v_x, v_y)$. ■

As it turns out, the optimization problem (19), or equivalently our identification problem, is generally *NP-hard* [47].

*Corollary 4.5: **(Complexity of the attack identification problem)*** Consider the system (1) with identifiable attack set K. The attack identification problem given the system matrices E, A, C, and the measurements y is NP-hard.

Proof: Consider the NP-hard [48] sparse recovery problem $\min_{\bar{\xi}\in\mathbb{R}^n} \|\bar{y} - \bar{C}\bar{\xi}\|_{\ell_0}$, where $\bar{C} \in \mathbb{R}^{p\times n}$ and $\bar{y} \in \mathbb{R}^p$ are given and constant. In order to prove the claimed statement, we show that every instance of the sparse recovery problem can be cast as an instance of (19). Let $E = I$, $A = 0$, $C = \bar{C}$, and $y(t) = \bar{y}$ at all times. Notice that $v_y(t) = \bar{y} - C\xi(t)$ and $\xi(t) = \xi(0) + \int_0^t v_x(\tau)d\tau$. The problem (19) can be written as

$$\min_{v_x,\,\xi} \|v_x\|_{\mathcal{L}_0} + \|\bar{y} - \bar{C}\xi(t)\|_{\mathcal{L}_0} = \min_{v_x(t),\,\bar{\xi}} \|v_x(t)\|_{\mathcal{L}_0} + \|\bar{y} - \bar{C}\bar{\xi} - \bar{C}\int_0^t v_x(\tau)d\tau\|_{\mathcal{L}_0}, \qquad (20)$$

where $\bar{\xi} = \xi(0)$. Notice that there exists a minimizer to problem (20) with $v_x(t) = 0$ for all t. Indeed, since $\|\bar{y} - \bar{C}\bar{\xi} - \bar{C}\int_0^t v_x(\tau)d\tau\|_{\mathcal{L}_0} = |\cup_{t\in\mathbb{R}_{\geq 0}} \text{supp}(\bar{y} - \bar{C}\bar{\xi} - \bar{C}\int_0^t v_x(\tau)d\tau)| \geq |\text{supp}(\bar{y} - \bar{C}\bar{\xi} - \bar{C}\int_0^0 v_x(\tau)d\tau)| = \|\bar{y} - \bar{C}\bar{\xi}\|_{\ell_0}$, problem (20) can be equivalently written as $\min_{\bar{\xi}} \|\bar{y} - \bar{C}\bar{\xi}\|_{\ell_0}$. ■

By Corollary 4.5 the general attack identification problem is combinatorial in nature, and its general solution will require substantial computational effort. In the next subsection we propose a complete identification algorithm.

D. Centralized attack identification

The identification of the attack set K requires a combinatorial procedure, since, a priori, K is one of the $\binom{n+p}{|K|}$ possible attack sets. The following centralized attack identification procedure consists of designing a residual filter to determine whether a predefined set coincides with the attack set. Analogously to the attack detection filter developed in Subsections IV-A and IV-B, the output of the attack identification filter for the attack set K will be a residual signal r_K. If each monitor is equipped with such an attack identification filter and if the attack K is identifiable, then the outputs of the monitor and the filter are related as follows: $\psi_2 = K$ if and only if $r_K(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$.

The design of this residual filter consists of three steps – an input output transformation, a state transformation, and an output injection and definition of a specific residual. We start by showing that the identification problem can be carried out for a modified system without corrupted measurements.

Lemma 4.6: ***(Attack identification with safe measurements)*** Consider the descriptor system (1) with attack set K. The attack set K is identifiable for the descriptor system (1) if and only if it is identifiable for the following descriptor system without corrupted measurements:

$$\begin{aligned} E\dot{x}(t) &= (A - B_K D_K^{\dagger} C)x(t) + B_K(I - D_K^{\dagger} D_K)u_K(t), \\ \tilde{y}(t) &= (I - D_K D_K^{\dagger})Cx(t). \end{aligned} \tag{21}$$

Proof: Due to the identifiability hypothesis, there exists no attack set R with $|R| \leq |K|$ and $R \neq K$, $s \in \mathbb{C}$, $g_K \in \mathbb{R}^{|K|}$, $g_R \in \mathbb{R}^{|R|}$, and $x \in \mathbb{R}^n \setminus \{0\}$ such that

$$\left[\begin{array}{c|c|c} sE - A & -B_K & -B_R \\ \hline C & D_K & D_R \\ \hline C & D_K & D_R \end{array}\right] \begin{bmatrix} x \\ g_K \\ g_R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{22}$$

where we added an additional (redundant) output equation (see Theorem 3.4). A multiplication of equation (22) from the left by the projectors blkdiag$\big(I\,,\ D_K D_K^{\dagger}\,,\ (I - D_K D_K^{\dagger})\big)$ yields

$$\left[\begin{array}{c|c|c} sE - A & -B_K & -B_R \\ \hline D_K D_K^{\dagger} C & D_K & D_K D_K^{\dagger} D_R \\ \hline (I - D_K D_K^{\dagger})C & 0 & (I - D_K D_K^{\dagger})D_R \end{array}\right] \begin{bmatrix} x \\ g_K \\ g_R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The variable g_K can be eliminated in the first redundant (corrupted) output equation according to

$$g_K = -D_K^{\dagger} C x - D_K^{\dagger} D_R g_R + (I - D_K^{\dagger} D_K) g_K.$$

Thus, $P(s)[x^{\mathsf{T}}\ g_K^{\mathsf{T}}\ g_R^{\mathsf{T}}]^{\mathsf{T}} = 0$ has no solution, where $P(s)$ is

$$\left[\begin{array}{c|c|c} sE - A + B_K D_K^{\dagger} C & -B_K(I - D_K^{\dagger} D_K) & -B_R + B_K D_K^{\dagger} D_R \\ \hline (I - D_K D_K^{\dagger})C & 0 & (I - D_K D_K^{\dagger})D_R \end{array}\right]$$

The statement follows. ■

The second design step of our attack identification monitor relies on the concept of *conditioned invariant subspace*. We refer to [24], [25], [28], [49] for a comprehensive discussion of geometric control theory. Let $\mathcal{S}^*$ be the conditioned invariant subspace associated with the system (E, A, B, C, D), that is, the smallest subspace of the state space satisfying

$$\begin{bmatrix} A & B \end{bmatrix} \left(\begin{bmatrix} E^{-1}\mathcal{S}^* \\ \mathbb{R}^m \end{bmatrix} \cap \operatorname{Ker} \begin{bmatrix} C & D \end{bmatrix} \right) \subseteq \mathcal{S}^*, \tag{23}$$

and let L be an output injection matrix satisfying

$$\begin{bmatrix} A + LC & B + LD \end{bmatrix} \begin{bmatrix} E^{-1}\mathcal{S}^* \\ \mathbb{R}^m \end{bmatrix} \subseteq \mathcal{S}^*. \tag{24}$$

Notice that the conditioned invariant $\mathcal{S}^*$ and an output injection L satisfying (23) and (24) always exist (for instance, take $\mathcal{S}^* = \mathbb{R}^n$). We transform the descriptor system (21) into a set of canonical coordinates representing $\mathcal{S}^*$ and its orthogonal complement. For a nonsingular system ($E = I$) such an equivalent state representation can be achieved by a nonsingular transformation of the form $Q^{-1}(sI - A)Q$. However, for a singular system different transformations need to be applied in the domain and codomain such as $P^{\mathsf{T}}(sE - A)Q$ for nonsingular P and Q.

Lemma 4.7: ***(Input decoupled system representation)*** For system (21), let $\mathcal{S}^*$ and L be as in (23) and (24), respectively. Define the unitary matrices $P = \begin{bmatrix} \text{Basis}(\mathcal{S}^*) & \text{Basis}((\mathcal{S}^*)^\perp) \end{bmatrix}$ and $Q = \begin{bmatrix} \text{Basis}(E^{-1}\mathcal{S}^*) & \text{Basis}((E^{-1}\mathcal{S}^*)^\perp) \end{bmatrix}$. Then

$$P^{\mathsf{T}} E Q = \begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ 0 & \tilde{E}_{22} \end{bmatrix}, \; P^{\mathsf{T}} B_K (I - D_K^\dagger D_K) = \begin{bmatrix} \tilde{B}_K(t) \\ 0 \end{bmatrix}, \; P^{\mathsf{T}}(A - B_K D_K^\dagger C + LC) Q = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix},$$

$$(I - D_K D_K^\dagger) C) Q = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}.$$

The attack set K is identifiable for the descriptor system (1) if and only if it is identifiable for the descriptor system

$$\begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ 0 & \tilde{E}_{22} \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \tilde{B}_K(t) \\ 0 \end{bmatrix},$$

$$y(t) = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \tag{25}$$

Proof: Let $\mathcal{L} = E^{-1}\mathcal{S}^*$ and $\mathcal{M} = \mathcal{S}^*$. Notice that $(A + LC)E^{-1}\mathcal{S}^* \subseteq \mathcal{S}^*$ by the invariance property of $\mathcal{S}^*$ [28], [49]. It follows that $\mathcal{L}$ and $\mathcal{M}$ are a pair of *right deflating subspaces* for the matrix pair $(A + LC, E)$ [50], that is, $\mathcal{M} = A\mathcal{L} + E\mathcal{L}$ and $\dim(\mathcal{M}) \leq \dim(\mathcal{L})$. The sparsity pattern in the descriptor and dynamic matrices $\tilde{E}$ and $\tilde{A}$ of (25) arises by construction of the right deflating subspaces P and Q [50, Eq. (2.17)], and the sparsity pattern in the input matrix arises due to the invariance properties of $\mathcal{S}^*$ containing $\mathrm{Im}(B_K)$. The statement follows because the output injection L, the coordinate change $x \mapsto Q^{-1}x$, and the left-multiplication of the dynamics by P^T does not affect the existence of zero dynamics. ∎

For the ease of notation and without affecting generality, the third and final design step of our attack identification filter is presented for the pre-conditioned system (25).

Theorem 4.8: ***(Attack identification for attack set*** K***)*** Consider the preconditioned system (25) associated with the descriptor system (1). Assume that the attack set is identifiable, the network initial state $x(0)$ is known, and the assumptions (A1) through (A3) are satisfied. Consider the *attack identification filter for the attack signature* (B_K, D_K)

$$\begin{aligned} \tilde{E}_{22}\dot{w}_2(t) &= (\tilde{A}_{22} + \tilde{G}(I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2)w_2(t) - \tilde{G}\bar{y}(t), \\ r_K(t) &= (I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2 w_2(t) - \bar{y}(t), \quad \text{with} \quad \bar{y}(t) = (I - \tilde{C}_1\tilde{C}_1^\dagger)y(t), \end{aligned} \tag{26}$$

where $w_2(0) = x_2(0)$, and $\tilde{G}$ is such that $(\tilde{E}_{22}, \tilde{A}_{22} + \tilde{G}(I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2)$ is Hurwitz. Then $r_K(t) = 0$ for all times $t \in \mathbb{R}_{\geq 0}$ if and only if K coincides with the attack set.

Proof: Let $w(t) = [w_1(t)^\mathsf{T}\ w_2(t)^\mathsf{T}]^\mathsf{T}$, where $w_1(t)$ obeys

$$\tilde{E}_{11}\dot{w}_1(t) + \tilde{E}_{12}\dot{w}_2(t) = \tilde{A}_{11}w_1(t) + \tilde{A}_{12}w_2(t).$$

Consider the filter error $e(t) = w(t) - x(t)$, and notice that

$$\begin{aligned} \begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ 0 & E_{22} \end{bmatrix} \begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} &= \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix} \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} - \begin{bmatrix} \tilde{B}_K \\ 0 \end{bmatrix} u_K(t), \\ r_K(t) &= (I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2 e_2(t), \end{aligned}$$

where $\bar{A}_{22} = \tilde{A}_{22} + \tilde{G}(I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2$. Notice that $r_K(t)$ is not affected by the input $u_K(t)$, so that, since $e_2(0) = 0$ due to $w_2(0) = x_2(0)$, the residual $r_K(t)$ is identically zero when K is the attack set. In order to prove the theorem we are left to show that for every set R, with $|R| \leq |K|$ and $R \cap K = \emptyset$, every attack mode u_R results in a nonzero residual r_K. From Theorem 3.4 and the identifiability hypothesis, for any $R \neq K$, there exists no solution to

$$\left[\begin{array}{cc|c|c} s\tilde{E}_{11} - \tilde{A}_{11} & s\tilde{E}_{12} - \tilde{A}_{12} & \tilde{B}_K & -B_{R1} \\ 0 & s\tilde{E}_{22} - \bar{A}_{22} & 0 & -B_{R2} \\ \hline \tilde{C}_1 & \tilde{C}_2 & 0 & D_R \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ g_K \\ g_R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

A projection of the equation $0 = \tilde{C}_1 x_1 + \tilde{C}_2 x_2 + D_R g_R$ onto the image of $\tilde{C}_1$ and its orthogonal complement yields

$$\left[\begin{array}{cc|c|c} s\tilde{E}_{11} - \tilde{A}_{11} & s\tilde{E}_{12} - \tilde{A}_{12} & B_K & -B_{R1} \\ 0 & s\tilde{E}_{22} - \bar{A}_{22} & 0 & -B_{R2} \\ \hline \tilde{C}_1 & \tilde{C}_1\tilde{C}_1^\dagger\tilde{C}_2 & 0 & \tilde{C}_1\tilde{C}_1^\dagger D_R \\ 0 & (I - \tilde{C}_1\tilde{C}_1^\dagger)\tilde{C}_2 & 0 & (I - \tilde{C}_1\tilde{C}_1^\dagger)D_R \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ g_K \\ g_R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{27}$$

Due to the identifiability hypothesis the set of equations (27) features no solution $[x_1^\mathsf{T}\ x_2^\mathsf{T}\ g_K^\mathsf{T}\ g_R^\mathsf{T}]^\mathsf{T}$ with $[x_1^\mathsf{T}\ x_2^\mathsf{T}]^\mathsf{T} = 0$.

Observe that, for every x_2 and g_R, there exists $x_1 \in \operatorname{Ker}(\tilde{C}_1)^\perp$ such that the third equation of (27) is satisfied. Furthermore, for every x_2 and g_R, there exist $x_1 \in \operatorname{Ker}(\tilde{C}_1)$ and g_K such that the first equation of (27) is satisfied. Indeed, since $QE^{-1}\mathcal{S}^* = [\operatorname{Im}(I)\ 0]^\mathsf{T}$ and $P^\mathsf{T}\mathcal{S}^* = [\operatorname{Im}(I)\ 0]^\mathsf{T}$, the invariance of $\mathcal{S}^*$ implies that $\mathcal{S}^* = A(E^{-1}\mathcal{S}^* \cap \operatorname{Ker}(C)) + \operatorname{Im}(B_K)$, or equivalently in new coordinates, $\operatorname{Im}(I) = \tilde{A}_{11}\operatorname{Ker}(\tilde{C}_1) + \operatorname{Im}(\tilde{B}_K)$. Finally note that $[(s\tilde{E}_{11} - \tilde{A}_{11})\operatorname{Ker}(\tilde{C}_1)\ \tilde{B}_K]$ is of full row rank due to the controllability of the subspace $\mathcal{S}^*$ [28]. We conclude that there exist no vectors x_2 and g_R such that $(s\tilde{E}_{22} - \bar{A}_{22})x_2 - B_{R2}g_R = 0$ and $(I - \tilde{C}_1\tilde{C}_1^\dagger)(\tilde{C}_2 x_2 + D_R g_R) = 0$ and the statement follows. ■

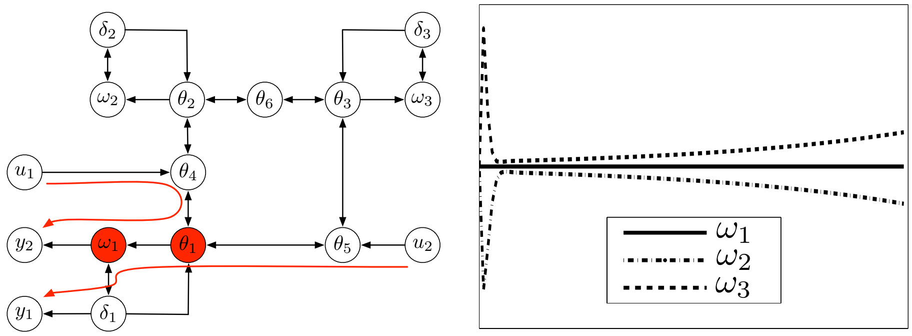
The design of the attack identification filter (26) is summarized as follows:

(1) from system (1) define the system (21);

(2) compute $\mathcal{S}^*$ and L for system (21) as in (23) and (24), and apply L, P, and Q as in Lemma 4.7 leading to system (25);

(3) for system (25), define r_K and apply the output injection $\bar{G}$ as in (26).

Remark 6: ***(Literature comparison)*** Our identification filter extends classical results concerning the design of unknown-input fault detection filters. In particular, our filter generalizes the construction of [6] to descriptor systems with direct feedthrough matrix. Additionally, we guarantee the absence of invariant zeros in the residual dynamics. By doing so, our attack identification filter is sensitive to *every* identifiable attack strategy. Notice that classical fault



(a) Input/output graph $\mathcal{G}_{\text{iso}}$ (b) Effect of an undetectable attack

Fig. 4. In Fig. 4(a), there is no linking of size 2 from the input to the output vertices. Indeed, the vertices θ_1 and ω_1 belong to every path from $\{u_1, u_2\}$ to $\{y_1, y_2\}$. Two input to output paths are depicted in red. In Fig. 4(b), the velocities ω_2 and ω_3 are driven unstable by the inputs u_1 and u_2, which are undetectable from the measurements of ω_1 and δ_1.

detection filters, for instance those presented in [6], are guaranteed to detect and isolate signals that do not excite exclusively zero dynamics. Finally, an equivalent attack identification filter for nonsingular or index-one systems is presented in our previous work [17]. □

Remark 7: ***(Complexity of centralized identification)*** Our centralized identification procedure assumes the knowledge of the cardinality k of the attack set, and it achieves identification by constructing a residual generator for $\binom{n+p}{k}$ possible attack sets. Thus, our procedure constructs $O(n^k)$ filters. If only an upper bound $\bar{k}$ on the cardinality of the attack set is available, identification can be achieved by constructing $\binom{n+p}{\bar{k}}$ filters, and by intersecting the attack sets generating zero residuals. In Section IV-C we show that this non-polynomial complexity is inherent to the identification problem. □

V. Illustrative examples

A. An example of state attack against a power network

Consider the power network model analyzed in Example 1 and illustrated in Fig. 2(a). We consider a load altering attack [18] affecting the power demand P_δ at the load buses 4 and 5. Assume that the variables θ_4 and θ_5 are affected by the unknown and unmeasurable signals u_1 and u_2. Suppose that a monitoring unit measures the state variables of the first generator, that is, $y_1 = \delta_1$ and $y_2 = \omega_1$.

Notice from Fig. 4(a) that the maximum size of a linking from the failure to the output vertices is 1, so that, by Theorem 3.5, there exists a structural vulnerability. In other words, for every choice of the network matrices, there exist nonzero u_1 and u_2 that are not detectable through the measurements.[4]

We now consider a numerical realization of this system. Let the input matrices be $B = [e_8 \; e_9]$ and $D = [0 \; 0]^\mathsf{T}$, the measurement matrix be $C = [e_1 \; e_4]^\mathsf{T}$, and the system matrix A be as in Remark 1 with $M_g = \text{blkdiag}(.125, .034, .016)$, $D_g = \text{blkdiag}(.125, .068, .048)$, and

$$\mathcal{L} = \begin{bmatrix} .058 & 0 & 0 & -.058 & 0 & 0 & 0 & 0 & 0 \\ 0 & .063 & 0 & 0 & -.063 & 0 & 0 & 0 & 0 \\ 0 & 0 & .059 & 0 & 0 & -.059 & 0 & 0 & 0 \\ -.058 & 0 & 0 & .235 & 0 & 0 & -.085 & -.092 & 0 \\ 0 & -.063 & 0 & 0 & .296 & 0 & -.161 & 0 & -.072 \\ 0 & 0 & -.059 & 0 & 0 & .330 & 0 & -.170 & -.101 \\ 0 & 0 & 0 & -.085 & -.161 & 0 & .246 & 0 & 0 \\ 0 & 0 & 0 & -.092 & 0 & -.170 & 0 & .262 & 0 \\ 0 & 0 & 0 & 0 & -.072 & -.101 & 0 & 0 & .173 \end{bmatrix}.$$

Let $U_1(s)$ and $U_2(s)$ be the Laplace transform of the attack signals u_1 and u_2, and let

$$\begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix} = \begin{bmatrix} \frac{-1.024s^4 - 5.121s^3 - 10.34s^2 - 9.584s - 3.531}{s^4 + 5s^3 + 9.865s^2 + 9.173s + 3.531} \\ 1 \end{bmatrix} \bar{U}(s),$$

for *some arbitrary* nonzero signal $\bar{U}(s)$. Then it can be verified that the attack cannot be detected through the measurements y. In fact, the transfer matrix mapping $\bar{U}(s)$ to $U(s)$ coincides with the null space of the input/output transfer matrix. An example is in Fig. 4(b), where the second and the third generator are driven unstable by the attack, but the first generator does not deviate from its nominal operating condition.

Suppose now that the rotor angle of the first generator and the voltage angle at the 6-th bus are measured, that is, $C = [e_1 \; e_{12}]^\mathsf{T}$. Then, there exists a linking of size 2 from $\mathcal{U}$ to $\mathcal{Y}$, and the system (E, A, B, C) is left-invertible. Following Theorem 3.7, the invariant zeros of the power network can be computed by looking at its reduced system, and they are $-1.6864 \pm 1.8070i$ and $-0.8136 \pm 0.2258i$. Consequently, if the network state is unknown at the failure time, there exists vulnerabilities that an attacker may exploit to affect the network while remaining undetected. Finally, we remark that such state attacks are entirely realizable by cyber attacks [18].

B. An example of output attack against a power network

Consider the IEEE 14 bus power network (Fig. 5) modeled as a descriptor system as in Section II. Following [9], let the measurements $y = Cx$ be given by the real power injections at all buses, of the real power flows of all branches, and one generator rotor angle (or one bus

[4] When these ouput-nulling inputs u_1, u_2 are regarded as additional loads, then they are entirely sustained by the second and third generator.

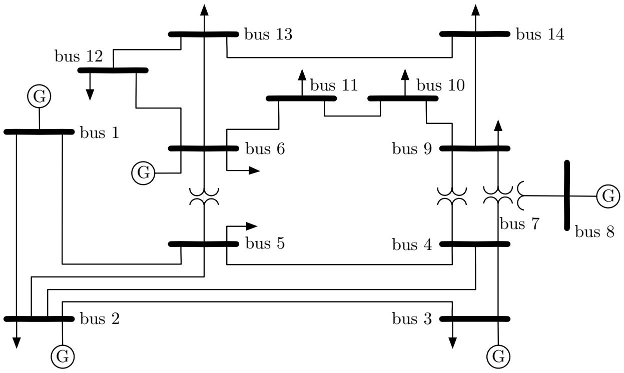


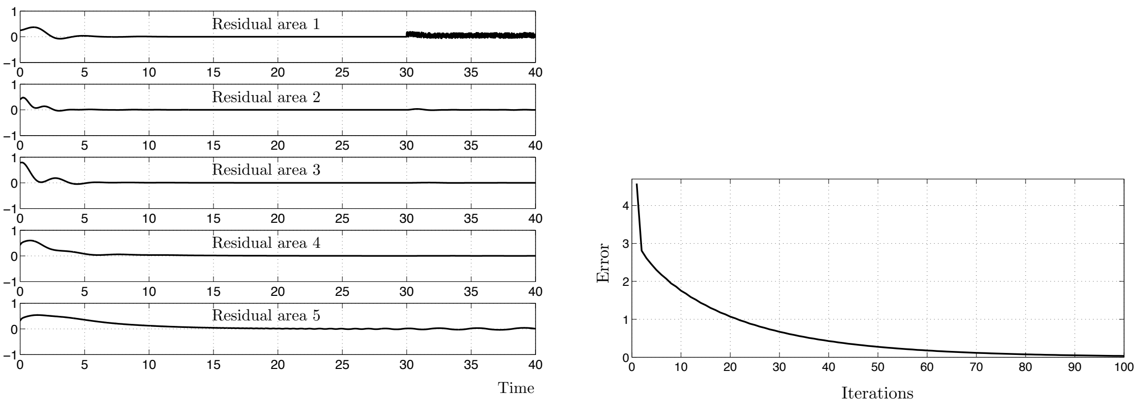
Fig. 5. For the IEEE 14 bus system in Fig. 5, if the voltage angle of one bus is measured exactly, then a cyber attack against the measurements data is always detectable by our dynamic detection procedure. In contrary, as shown in [9], a cyber attack may remain undetected by a static procedure if it compromises as few as four measurements.

angle). We assume that an attacker can compromise all the measurements, independently of each other, except for one referring to the rotor angle.

Let $k \in \mathbb{N}_0$ be the cardinality of the attack set. It is known that an attack undetectable to a static detector exists if $k \geq 4$ [9]. In other words, due to the sparsity pattern of C, there exists a signal $u_K(t)$, with (the same) four nonzero entries at all times, such that $Du_K(t) \in \operatorname{Im}(C)$ at all times. Hence the attack set K remains undetected by a static detector through the attack input u_K. On the other hand, following Theorem 3.3, it can be verified that, for the same output matrix C, and independent of the value of k, there exists *no* undetectable (output) attacks for a dynamic monitor. It should be noticed that this result relies on the fact that the rotor angle measurement is known to be correct, because, for instance, it is protected using sophisticated and costly security methods [29]. Since the state of the IEEE 14 bus system can be reconstructed by means of this measurement only (the system turns out to be observable by measuring one generator rotor angle), the output attack Du is easily identified as $Du = y - C\hat{x}$, where $\hat{x} = x$ is the reconstructed system state at time t.

C. An example of distributed detection

The IEEE 118 bus system shown in Fig. 3 is composed of 118 buses and 54 generators, and its parameters can be found in [51]. Following Section II, a linear continuous-time descriptor model of the system under attack takes the form (1).



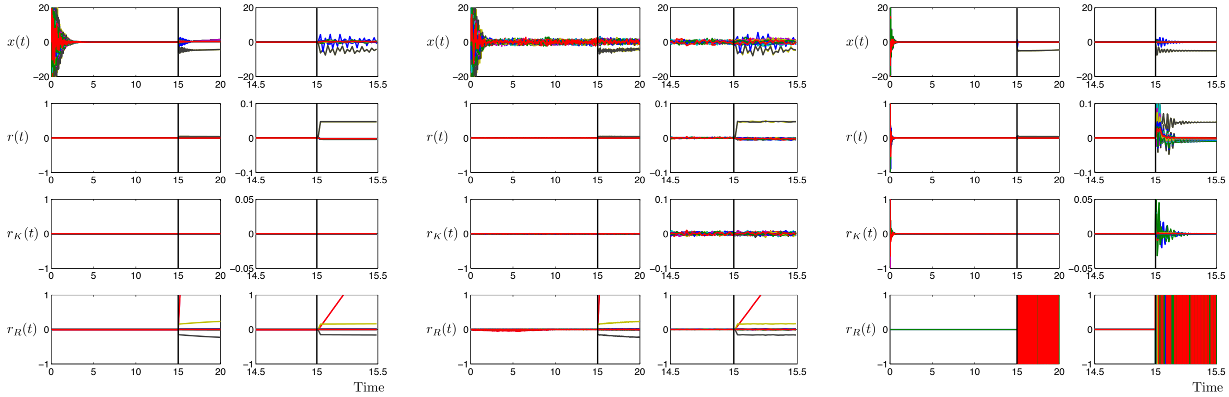
(a) Residual functions computed by distributed attack detection filter (15)

(b) Error $\max_{t\in[0,T]} \left\| w^{(k)}(t) - w(t) \right\|_\infty$ induced by waveform relaxation

Fig. 6. Distributed detection of an output attack in the IEEE 118 system: The attacker compromises the measurements of all generators in area 1 from time 30s with a signal uniformly distributed in the interval $[0, 0.5]$. The residuals in Fig. 6(a) show that the attack is correctly detected, because the residual functions do not decay to zero. For the simulation, we run $k = 100$ iterations of the attack detection method. The plot in Fig. 6(b) represents the error of our waveform relaxation based filter (15) with respect to the corresponding decentralized filter. As predicted by Theorem 4.3, the error is convergent.

For estimation and detection purposes, we partition the IEEE 118 system into 5 disjoint areas, we assign a control center to each area, and we implement our detection procedure via the filter (15); see Fig. 3 for a graphical illustration. Suppose that each control center continuously measures the angle of the generators in its area, and suppose that an attacker compromises the measurements of all the generators of the first area. In particular, starting at time 30s, the attacker comprises all measurements in area 1 by adding a signal u_K. It can be verified that the attack set K is detectable, see Theorem 3.3. According to assumption (A3), the attack signal u_K needs to be continuous to guarantee a continuous state trajectory (since the associated descriptor model is of index 1). To show the robustness of our detection filter (15), we let u_K be discontinuous and randomly distributed in the interval $[0, 0.5]\,\mathrm{rad}$.

The control centers implement the distributed attack detection procedure described in (15), with $G = AC^{\mathsf{T}}$. It can be verified that the pair $(E, A_D + GC)$ is Hurwitz stable, and that $\rho\left(\mathrm{j}\,\omega E - A_D - GC)^{-1} A_C\right) < 1$ for all $\omega \in \mathbb{R}$. As predicted by Theorem 4.3, our distributed attack detection filter is convergent; see Fig. 6(a). For completeness, in Fig. 6(b) we illustrate the convergence rate of our waveform relaxation-based filter as a function of the number of iterations k. Notice that the number of iterations directly reflects the communication complexity



(a) Nominal linear system dynamics (b) Linear and noisy system dynamics (c) Nonlinear and noisy system dynamics

Fig. 7. In Fig. 7(a) we report our simulation results for the case of linear network dynamics without noise and for the proposed detection monitor (5) and identification monitor (26), respectively. The state trajectory x consists of the generators angles and frequencies. The detection residual r becomes nonzero after time 15s, and it reveals the presence of the attack. The identification residual r_K is identically zero even after time 15s, and it reveals that the attack set is $K = \{101, 102\}$. The identification residual r_R is nonzero after time 15s, and it reveals that R is not the attack set. In Fig. 7(b) we report our simulation results for the case of linear network dynamics driven by state and measurements noise. For this case, we choose the output injection matrices of the detection and identification filters as the corresponding optimal Kalman gain. Due to the presence of noise, the residuals deviate from their nominal behavior reported in Fig. 7(a). Although the attack is clearly still detectable and identifiable, additional statistical tools such as hypothesis testing [7] may be adopted to analyze the residuals r, r_K, and r_R. In Fig. 7(c) we report our simulation results for the case of nonlinear network dynamics without noise. For this case, the detection and identification filters are designed for the *nominal linearized dynamics* with output injection matrices as the corresponding optimal Kalman gain. Despite the presence of unmodeled nonlinear dynamics, the residuals reflect their nominal behavior reported in Fig. 7(a).

of our detection scheme.

D. An example of detection and identification in the presence of noise and model uncertainties

We apply our centralized attack detection and identification methods to the IEEE RTS96 power network [52]. In particular, we first consider the nominal case, in which the power network dynamics evolve as linear time-invariant descriptor system, as described in Section II. Second, we consider the case of additive state and measurement noise, and we show the robustness of the attack detection and identification monitors. Third, we consider the case of nonlinear differential-algebraic power network dynamics and show the effectiveness of our methods in the presence of unmodeled nonlinear dynamics.

For our numerical studies, we assume the angles and frequencies of every generator to be measured. Additionally, we let the attacker affect the angles of the generators $\{101, 102\}$ with a random signal starting from time 15s. Since the considered power network dynamics are of index one, the filters are implemented using the nonsingular Kron-reduced system representation [17]. The results of our simulations are in Fig. 7(a), Fig. 7(b), and Fig. 7(c). In conclusion, our centralized detection and identification filters appears robust to state and measurements noise and unmodeled dynamics.

VI. Conclusion

For cyber-physical systems modeled by linear time-invariant descriptor systems, we have analyzed fundamental monitoring limitations. In particular, we have characterized undetectable and unidentifiable attacks from a system-theoretic and a graph-theoretic perspective. Additionally, we have designed centralized and distributed monitors.

Future and ongoing work includes (i) a detailed analysis of the convergence of our distributed monitor, (ii) the design of distributed identification monitors, and (iii) the design of monitors robust to system noise and unmodeled dynamics.

References

[1] J. Slay and M. Miller, “Lessons learned from the Maroochy water breach,” *Critical Infrastructure Protection*, vol. 253, pp. 73–82, 2007.

[2] J. P. Conti, “The day the samba stopped,” *Engineering Technology*, vol. 5, no. 4, pp. 46–47, 06 March - 26 March, 2010.

[3] S. Kuvshinkova, “SQL Slammer worm lessons learned for consideration by the electricity sector,” *North American Electric Reliability Council*, 2003.

[4] J. P. Farwell and R. Rohozinski, “Stuxnet and the future of cyber war,” *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[5] G. Richards, “Hackers vs slackers,” *Engineering & Technology*, vol. 3, no. 19, pp. 40–43, 2008.

[6] M.-A. Massoumnia, G. C. Verghese, and A. S. Willsky, “Failure detection and identification,” *IEEE Transactions on Automatic Control*, vol. 34, no. 3, pp. 316–321, 1989.

[7] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.

[8] S. Amin, A. Cárdenas, and S. Sastry, “Safe and secure networked control systems under denial-of-service attacks,” in *Hybrid Systems: Computation and Control*, vol. 5469, Apr. 2009, pp. 31–45.

[9] Y. Liu, M. K. Reiter, and P. Ning, “False data injection attacks against state estimation in electric power grids,” in *ACM Conference on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[10] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. Sastry, “Cyber security analysis of state estimators in electric power systems,” in *IEEE Conf. on Decision and Control*, Atlanta, GA, USA, Dec. 2010, pp. 5991–5998.

[11] Y. Mo and B. Sinopoli, “Secure control against replay attacks,” in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sep. 2010, pp. 911–918.

[12] R. Smith, “A decoupled feedback structure for covertly appropriating network control systems,” in *IFAC World Congress*, Milan, Italy, Aug. 2011, pp. 90–95.

[13] M. Zhu and S. Martínez, “Stackelberg-game analysis of correlated attacks in cyber-physical systems,” in *American Control Conference*, San Francisco, CA, USA, Jul. 2011, pp. 4063–4068.

[14] F. Hamza, P. Tabuada, and S. Diggavi, “Secure state-estimation for dynamical systems under active adversaries,” in *Allerton Conf. on Communications, Control and Computing*, Sep. 2011.

[15] C. L. DeMarco, J. V. Sariashkar, and F. Alvarado, “The potential for malicious control in a competitive power systems environment,” in *IEEE Int. Conf. on Control Applications*, Dearborn, MI, USA, 1996, pp. 462–467.

[16] G. Dan and H. Sandberg, “Stealth attacks and protection schemes for state estimators in power systems,” in *IEEE Int. Conf. on Smart Grid Communications*, Gaithersburg, MD, USA, Oct. 2010, pp. 214–219.

[17] F. Pasqualetti, F. Dörfler, and F. Bullo, “Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design,” in *IEEE Conf. on Decision and Control and European Control Conference*, Orlando, FL, USA, Dec. 2011, pp. 2195–2201.

[18] A.-H. Mohsenian-Rad and A. Leon-Garcia, “Distributed internet-based load altering attacks against smart power grids,” *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 667 –674, 2011.

[19] S. Sridhar, A. Hahn, and M. Govindarasu, “Cyber–physical system security for the electric power grid,” *Proceedings of the IEEE*, vol. 99, no. 1, pp. 1–15, 2012.

[20] S. Sundaram and C. Hadjicostis, “Distributed function calculation via linear iterative strategies in the presence of malicious agents,” *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, 2011.

[21] F. Pasqualetti, A. Bicchi, and F. Bullo, “Consensus computation in unreliable networks: A system theoretic approach,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, 2012.

[22] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, “Stealthy deception attacks on water SCADA systems,” in *Hybrid Systems: Computation and Control*, Stockholm, Sweden, Apr. 2010, pp. 161–170.

[23] D. G. Eliades and M. M. Polycarpou, “A fault diagnosis and security framework for water systems,” *IEEE Transactions on Control Systems Technology*, vol. 18, no. 6, pp. 1254–1265, 2010.

[24] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*, 3rd ed. Springer, 1985.

[25] G. Basile and G. Marro, *Controlled and Conditioned Invariants in Linear System Theory*. Prentice Hall, 1991.

[26] Y. Mo and B. Sinopoli, “False data injection attacks in control systems,” in *First Workshop on Secure Control Systems*, Stockholm, Sweden, Apr. 2010.

[27] J. W. van der Woude, “A graph-theoretic characterization for the rank of the transfer matrix of a structured system,” *Mathematics of Control, Signals and Systems*, vol. 4, no. 1, pp. 33–40, 1991.

[28] T. Geerts, “Invariant subspaces and invertibility properties for singular systems: The general case,” *Linear Algebra and its Applications*, vol. 183, pp. 61–88, 1993.

[29] A. R. Metke and R. L. Ekl, “Security technology for smart grid networks,” *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 99–107, 2010.

[30] E. Scholtz, “Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems,” Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[31] F. Pasqualetti, A. Bicchi, and F. Bullo, “A graph-theoretical characterization of power network vulnerabilities,” in *American Control Conference*, San Francisco, CA, USA, Jun. 2011, pp. 3918–3923.

[32] A. Osiadacz, *Simulation and Analysis of Gas Networks*. Houston, TX, USA: Gulf Publishing Company, 1987.

[33] A. Kumar and P. Daoutidis, *Control of Nonlinear Differential Algebraic Equation Systems*. CRC Press, 1999.

[34] X. Litrico and V. Fromion, *Modeling and Control of Hydrosystems*. Springer, 2009.

[35] J. Burgschweiger, B. Gnädig, and M. C. Steinbach, “Optimization models for operative planning in drinking water networks,” *Optimization and Engineering*, vol. 10, no. 1, pp. 43–73, 2009.

[36] P. F. Boulos, K. E. Lansey, and B. W. Karney, *Comprehensive Water Distribution Systems Analysis Handbook for Engineers and Planners*. American Water Works Association, 2006.

[37] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[38] J. M. Dion, C. Commault, and J. van der Woude, “Generic properties and control of linear structured systems: a survey,” *Automatica*, vol. 39, no. 7, pp. 1125–1144, 2003.

[39] J. Munkres, *Topology*. Prentice Hall, 2000.

[40] K. J. Reinschke, “Graph-theoretic approach to symbolic analysis of linear descriptor systems,” *Linear Algebra and its Applications*, vol. 197, pp. 217–244, 1994.

[41] J. Tokarzewski, *Finite Zeros in Discrete Time Control Systems*, ser. Lecture notes in control and information sciences. Springer, 2006.

[42] L. Dai, *Singular Control Systems*. Springer, 1989.

[43] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control Analysis and Design*, 2nd ed. Wiley, 2005.

[44] E. Lelarasmee, A. E. Ruehli, and A. L. Sangiovanni-Vincentelli, “The waveform relaxation method for time-domain analysis of large scale integrated circuits,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 1, no. 3, pp. 131–145, 1982.

[45] Z. Z. Bai and X. Yang, “On convergence conditions of waveform relaxation methods for linear differential-algebraic equations,” *Journal of Computational and Applied Mathematics*, vol. 235, no. 8, pp. 2790–2804, 2011.

[46] F. Dörfler, F. Pasqualetti, and F. Bullo, “Distributed estimation in continuous time with discrete communication,” *IEEE Transactions on Automatic Control*, Jun. 2012, to submit.

[47] M. R. Garey and D. S. Johnson, *Computers and Intractability*. Springer, 1979.

[48] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[49] F. L. Lewis, “Geometric design techniques for observers in singular systems,” *Automatica*, vol. 26, no. 2, pp. 411–415, 1990.

[50] K. D. Ikramov, “Matrix pencils: Theory, applications, and numerical methods,” *Journal of Mathematical Sciences*, vol. 64, no. 2, pp. 783–853, 1993.

[51] R. D. Zimmerman, C. E. Murillo-Sánchez, and D. Gan, “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[52] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, “The IEEE Reliability Test System - 1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010–1020, 1999.