51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Cyber-Physical Security via Geometric Control: Distributed Monitoring and Malicious Attacks

Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo

***Abstract*— Cyber-physical systems are ubiquitous in power systems, transportation networks, industrial processes, and critical infrastructures. These systems need to operate reliably in the face of unforeseen failures and external malicious attacks. This paper summarizes and extends our results on the security of cyber-physical systems based on geometric control theory: (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 design an attack strategy for a group of power generators to physically compromise the functionality of other generators. Novel contributions include a more general framework, the design of novel centralized and distributed identification monitors, and the attack design case study.**

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 [2] in March 2000, multiple recent power blackouts in Brazil [3], the SQL Slammer worm attack on the Davis-Besse nuclear plant in January 2003 [4], the StuxNet computer worm [5] in June 2010, and by various industrial security incidents [6], cyber-physical systems are prone to failures and attacks on their physical infrastructure, as well as 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 [7]. Cyber-physical systems, however, suffer from specific vulnerabilities which do not affect classical control systems, and for which 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 [2].

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. In [9] *false data* injection attacks against static state estimators are introduced. 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. 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. Finally, security issues of specific systems have received considerable attention, such as power networks [13]–[17], linear networks with misbehaving components [18], [19], and water networks [20], [21].

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 [22], [23], our framework includes the *deterministic static detection problem* considered in [9], [10], and the prototypical deception and denial of service [8], stealth [14], (dynamic) false-data injection [24], 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 the considered 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

This material is supported by NSF grants IIS-0904501 and CPS-1135819.

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

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 [25] on structural left-invertibility to regular descriptor systems.

Fourth, we design centralized and distributed monitors (Section IV). Our centralized monitors and our distributed detection monitor are complete, that is they detect and identify every detectable and identifiable attack. Instead, due to the computational complexity of the identification problem, our distributed identification monitor identifies a class of attacks, which we characterize, at a low computational cost.

Fifth, inspired by [13], we consider a competitive power generation scenario (Section V). We exploit our previous findings to characterize all control strategies for a coalition of generators to destabilize other machines involved in the power generation. Finally, we illustrate this technique on a model of the Western North American power grid (Fig. 2).

This paper extends our earlier works [15], [26], [27] in the following ways: (i) we consider a broader class of cyber-physical systems (focusing on general descriptor systems, rather than assuming and working with reduced representations without algebraic constraints), (ii) we propose novel centralized and distributed identification monitors, and (iii) we design novel cooperative attack strategies. All missing proofs and detailed discussions are available online [1].

II. Problem Setup and Preliminary Results

In the present 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, this simplified model has long proven useful in studying stability, faults, and attacks in power networks, sensor networks, and water networks among others. 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 linear time-invariant descriptor system[1]

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

where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$, $y : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$, $E \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, 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. We assume that each state and output variable can be independently compromised. Accordingly, we let $B = [I_{n\times n}\ 0_{n\times p}]$, $D = [0_{p\times n}\ I_{p\times p}]$, and $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n+p}$. Since the attackers strategy cannot, in general, be predicted, the modeling of attacks via unknown inputs is appropriate and convenient for the analysis.

The attack signal u depends upon the specific attack strategy. In the presence of $k \in \mathbb{N}_0$, $k \leq n+p$, attackers indexed by the *attack set* $K \subseteq \{1, \dots, n+p\}$, only and all the entries K of u are nonzero over time, that is, for each $i \in K$, there exists a time t such that $u_i(t) \neq 0$, 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 strategy, 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, denotes the *attack signature*. Hence, $Bu = B_K u_K$, and $Du = D_K u_K$. 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, $(Ax(0) + Bu(0)) \in \mathrm{Im}(E)$; and
(A3) the attack signal u is smooth.

Assumption (A1) assures the existence of a unique solution x to (1). Assumptions (A2) and (A3) guarantee smoothness of the state trajectory x and the measurements y.

Model of monitors. A *monitor* is a deterministic algorithm $\Phi : \Lambda \to \Psi$ 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 $\psi_2 \subseteq \{1, \dots, n+p\}$.

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 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) $\psi_1 = \text{True}$ *only if* the attack set K is nonempty ($\psi_1 =$ False, otherwise),
(ii) $\psi_2 = \emptyset$ *if and only if* $\psi_1 = \text{False}$, 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).

Due to our consistency assumption, monitors do not trigger false-alarms. Examples of monitors can be found in [10], [11], [15]. 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 a monitor Φ if $\psi_2 = K$.

An attack is *undetectable* (respectively *unidentifiable*) if no monitor detects (respectively identifies) the 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 attackers. In this work we consider colluding omniscient attackers with the ability of altering the cyber-physical dynamics through exogenous inputs. In particular,

[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.

we let the attack (Bu, Du) in (1) be designed based on knowledge of the system structure and parameters E, A, C, and the full state x at all times. Additionally, attackers have unlimited computation 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.

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 power (water) supply, line outage, pipe leakages, and failures of sensors and actuators. Possible cyber-physical attacks include measurements corruption [9], [10], [20], and attacks on the control architecture or the physical state itself [2], [13], [16], [17].

III. Fundamental Monitoring Limitations

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 $t \in \mathbb{R}_{\geq 0}$.

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 $t \in \mathbb{R}_{\geq 0}$.

We now elaborate on the above lemmas to derive fundamental detection and identification limitations.

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 $\|(sE - A)x\|_0 + \|Cx\|_0 = k$.

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) [23], [28].

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 $|\bar{K}| \leq 2k$.

In other words, 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 $|\bar{K}| \leq 2k$.

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 [25] associated with the system (1). We associate a directed input/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 *almost every* realization of E and A.

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 [23].

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, and that the system state at the attack initial time is known. 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 gives a characterization of structurally undetectable attacks. Various illustrative examples can be found in [1].

IV. Monitor Design for Attack Detection and Identification

A. Centralized attack detection

In the following result we 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

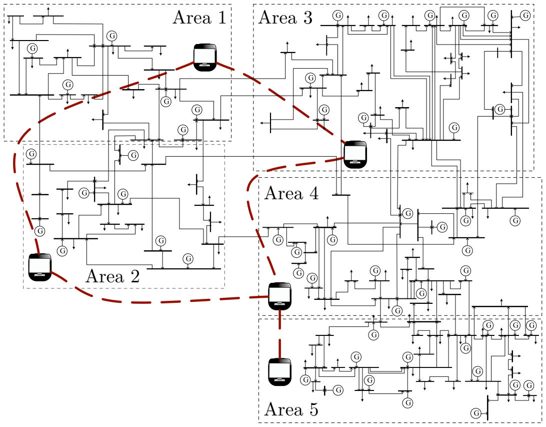

Fig. 1. Partition of IEEE 118 bus system into 5 areas. Each area is monitored and operated by a control center. These control centers cooperate to estimate the state and to assess the functionality of the whole network.

set K is detectable, and that the network initial state $x(0)$ is known. Consider the *centralized attack detection filter*

$$\begin{aligned} E\dot{w}(t) &= (A+GC)w(t) - Gy(t),\\ r(t) &= Cw(t) - y(t), \end{aligned} \tag{2}$$

where $w(0) = x(0)$ and the output injection $G \in \mathbb{R}^{n\times p}$ is such that the pair $(E, A+GC)$ is regular and Hurwitz. 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.

Notice that, if the network initial state is not available, then, since $(E, A+GC)$ is Hurwitz, an arbitrary initial state $w(0) \in \mathbb{R}^n$ can be chosen. Consequently, the filter converges asymptotically, and some attacks may remain undetected or unidentified. 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 in (2) should be chosen as 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, notice that attacks aligned with the noise statistics may be undetectable.

B. Distributed attack detection

Control centers are geographically deployed in a large scale cyber-physical system to operate the whole plant via distributed computation (see Fig. 1). 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 $A_D = \text{blkdiag}(A_1, \dots, A_N)$. We make the following assumptions:

(A4) the matrices E, C are block-diagonal, that is, $E = \text{blkdiag}(E_1, \dots, E_N)$, $C = \text{blkdiag}(C_1, \dots, C_N)$, 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

$$\begin{aligned} E_i\dot{x}_i(t) &= A_i x_i(t) + \sum\nolimits_{j \in \mathcal{N}_i^{\text{in}}} A_{ij} x_j(t),\\ y_i(t) &= C_i x_i(t), \quad i \in \{1, \dots, N\}, \end{aligned} \tag{3}$$

where $x_i : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_i}$ and $y_i : \mathbb{R}_{\geq 0} \to \mathbb{R}^{p_i}$ are the state and output of the i-th subsystem, and $\mathcal{N}_i^{\text{in}} = \{j \in \{1, \dots, N\} \setminus i \mid \|A_{ij}\| \neq 0\}$ are the in-neighbors of subsystem i. We also define the set of out-neighbors as $\mathcal{N}_i^{\text{out}} = \{j \in \{1, \dots, N\} \setminus i \mid \|A_{ji}\| \neq 0\}$. 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 $j \in \mathcal{N}_i^{\text{out}}$.

To derive an attack detection monitor we rely on waveform relaxation methods [29], [30] developed for parallel numerical integration. Consider the *waveform relaxation iteration*

$$E\dot{w}^{(k)}(t) = (A_D + GC)w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t), \tag{4}$$

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 the initial condition $w^{(k)}(0) = w_0$ for each $k \in \mathbb{N}$. Notice that (4) 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.

Theorem 4.2: ***(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*

$$\begin{aligned} E\dot{w}^{(k)}(t) &= \big(A_D + GC\big)w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t),\\ r(t) &= y(t) - Cw^{(k)}(t), \end{aligned} \tag{5}$$

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

$$\rho\left((j\omega E - A_D - GC)^{-1} A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}\,. \tag{6}$$

Then $\lim_{k\to\infty} r^{(k)}(t) = 0$ 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 $t \in [0,T]$.

The waveform relaxation iteration (4) can be implemented in the following distributed fashion. Assume that each control center i is able to numerically integrate the descriptor system

$$\begin{aligned} E_i\dot{w}_i^{(k)}(t) &= (A_i + G_iC_i)w_i^{(k)}(t) \\ &\quad + \sum\nolimits_{j\in\mathcal{N}_i^{\text{in}}} A_{ij}w_j^{(k-1)}(t) - G_iy_i(t), \end{aligned} \tag{7}$$

over a time interval $[0,T]$, 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 (7);
(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 4.2, for k sufficiently large, the local residuals $r_i^{(k)} = y_i - C_iw_i^{(k)}$ can be used to detect attacks. A related large-scale example is in [1].

Remark 1: ***(Implementation of distributed attack detection filter)*** For the implementation of the filter (5), control center i needs to transmit the signal $w_i^{(k)} : [0,T] \to \mathbb{R}^{n_i}$ at each iteration k. In practice, only an approximation or a finite basis representation $\hat{w}_i^{(k)}$ can be transmitted. The error due to this approximation can be characterized; see [31]. □

C. 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. 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 4.3: ***(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:

$$\begin{aligned} E\dot{x}(t) &= (A - B_KD_K^{\dagger}C)x(t) + B_K(I - D_K^{\dagger}D_K)u_K(t), \\ \tilde{y}(t) &= (I - D_KD_K^{\dagger})Cx(t). \end{aligned} \tag{8}$$

The second design step of our attack identification monitor relies on the concept of *conditioned invariant subspace*. We refer to [23], [28], [32] 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

$$\begin{bmatrix} A & B \end{bmatrix}\left(\begin{bmatrix} E^{-1}\mathcal{S}^* \\ \mathbb{R}^m \end{bmatrix} \cap \operatorname{Ker}\begin{bmatrix} C & D \end{bmatrix}\right) \subseteq \mathcal{S}^*, \tag{9}$$

and let L be an output injection matrix satisfying

$$\begin{bmatrix} A + LC & B + LD \end{bmatrix}\begin{bmatrix} E^{-1}\mathcal{S}^* \\ \mathbb{R}^m \end{bmatrix} \subseteq \mathcal{S}^*. \tag{10}$$

We transform the descriptor system (8) into a set of canonical coordinates representing $\mathcal{S}^*$ and its orthogonal complement.

Lemma 4.4: ***(Input decoupled system representation)*** For the system (8), let $\mathcal{S}^*$ and L be as in (9) and (10), 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

$$\begin{aligned} &P^{\mathsf{T}}EQ = \begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ 0 & \tilde{E}_{22} \end{bmatrix}, P^{\mathsf{T}}B_K(I - D_K^{\dagger}D_K) = \begin{bmatrix} \tilde{B}_K(t) \\ 0 \end{bmatrix}, \\ &P^{\mathsf{T}}(A - B_KD_K^{\dagger}C + LC)Q = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \\ &(I - D_KD_K^{\dagger})C)Q = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}. \end{aligned}$$

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

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

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 (11).

Theorem 4.5: ***(Attack identification for attack set K)*** Consider the system (11) 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)

$$\begin{aligned} \tilde{E}_{22}\dot{w}_2(t) &= (\tilde{A}_{22} + \tilde{G}(I - \tilde{C}_1\tilde{C}_1^{\dagger})\tilde{C}_2)w_2(t) - \tilde{G}\bar{y}(t), \\ r_K(t) &= (I - \tilde{C}_1\tilde{C}_1^{\dagger})\tilde{C}_2w_2(t) - \bar{y}(t), \quad \text{with} \\ \bar{y}(t) &= (I - \tilde{C}_1\tilde{C}_1^{\dagger})y(t), \end{aligned} \tag{12}$$

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.

The design of the filter (12) is summarized as follows:

(1) from system (1) define the system (8);
(2) compute $\mathcal{S}^*$ and L for system (8) as in (9) and (10), and apply L, P, and Q as in Lemma 4.4 leading to system (11);
(3) for system (11), define r_K and apply the output injection $\tilde{G}$ as in (12).

Our identification filter extends classical results concerning the design of unknown-input fault detection filters. Finally, an equivalent attack identification filter for nonsingular or index-one systems is presented in our previous work [15].

Remark 2: ***(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. We show in [1] that this non-polynomial complexity is inherent to the identification problem. □

D. Distributed attack identification

Consider the setup presented in Section IV-B with assumptions (A4)-(A7). The subsystems under attack read as

$$\begin{aligned} E_i\dot{x}_i(t) &= A_i x_i(t) + B_i^{\mathrm{b}} f_i(t) + B_{K_i} u_{K_i}(t), \\ y_i(t) &= C_i x_i(t) + D_{K_i} u_{K_i}(t), \quad i \in \{1, \dots, N\}, \end{aligned} \tag{13}$$

where $K_i = (K \cap V_i) \cup K_i^{\mathrm{p}}$ with K the attack set and K_i^{p} the corrupted measurements in the region $\mathcal{G}_{\mathrm{s}}^i$, $B_i^{\mathrm{b}} = [A_{i1} \cdots A_{i,i-1}\, A_{i,i+1} \cdots A_{iN}]$, and $f_i = [x_1^{\mathsf{T}} \cdots x_N^{\mathsf{T}}]^{\mathsf{T}}$. We refer to (13) as the *i-th decoupled system*, and we let $K_i^{\mathrm{b}} \subseteq V_i$ be the set of *boundary nodes* of (13), that is, the nodes $j \in \mathcal{V}_i$ with $A_{jk} \neq 0$ for some $k \in \{1, \dots, n\} \setminus V_i$. Our distributed identification method is based upon a divide and conquer procedure, and it consists of the following three steps.

(S1: estimation and communication) Each control center estimates the state of its own region by means of an *unknown-input observer* for the i-th subsystem subject to the unknown input $B_i^{\mathrm{b}} f_i$. For this task we build upon existing unknown-input estimation algorithms [1]. Let the state x_i be reconstructed modulo some subspace $\mathcal{F}_i$.[2] Let $F_i = \mathrm{Basis}(\mathcal{F}_i)$, and let $x_i = \tilde{x}_i + \hat{x}_i$, where $\hat{x}_i$ is the estimate computed by the i-th control center, and $\tilde{x}_i \in \mathcal{F}_i$. Finally, each control center i transmits the estimate $\hat{x}_i$ and the subspace F_i to control centers $\mathcal{N}_i^{\mathrm{out}}$.

(S2: residual generation) Observe that each input signal $A_{ij}x_j$ can be written as $A_{ij}x_j = A_{ij}\tilde{x}_j + A_{ij}\hat{x}_j$, where $\tilde{x}_j \in \mathcal{F}_j$. Then, after carrying out step (S1), only the inputs $A_{ij}\tilde{x}_j$ are unknown to the i-th control center, while the inputs $A_{ij}\hat{x}_j$ are known to the i-th center due to communication. Let $B_i^{\mathrm{b}} F_i = [A_{i1}F_1 \cdots A_{i,i-1}F_{i-1}\, A_{i,i+1}F_{i+1} \cdots A_{iN}F_N]$, and rewrite the signal $B_i^{\mathrm{b}}\tilde{x}$ as $B_i^{\mathrm{b}}\tilde{x} = B_i^{\mathrm{b}} F_i \bar{f}_i$, for some $\bar{f}_i$. The dynamics of the i-th subsystem read as

$$E_i\dot{x}_i(t) = A_i x_i(t) + B_i^{\mathrm{b}}\hat{x}(t) + B_i^{\mathrm{b}} F_i f_i(t) + B_{K_i} u_{K_i}(t).$$

Analogously to the filter presented in Theorem 4.5 for the attack signature (B_K, D_K), consider now the following filter (in appropriate coordinates) for (13) and $(B_i^{\mathrm{b}} F_i, 0)$

$$\begin{aligned} E_i\dot{w}_i(t) &= (A_i + L_i C_i)w_i(t) - Ly(t) + B_i^{\mathrm{b}}\bar{x}(t), \\ r_i(t) &= Mw_i(t) - Hy(t), \end{aligned} \tag{14}$$

where L_i is the injection matrix associated with the conditioned invariant subspace generated by $B_i^{\mathrm{b}} F_i$, with $(E_i, A_i + L_i C_i)$ Hurwitz, and $\bar{x}$ is the state transmitted to i by its neighbors. Notice that, in the absence of attacks in the regions $\mathcal{N}_i^{\mathrm{in}}$, we have $B_i^{\mathrm{b}}\bar{x} = B_i^{\mathrm{b}}\hat{x}$. Finally, let the matrices M and H in (14) be chosen so that the input $B_i^{\mathrm{b}} F_i f_i$ does not affect the residual r_i.[3]

(S3: cooperative residual analysis) We next state a key result for our distributed identification procedure.

Lemma 4.6: ***(Characterization of nonzero residuals)*** Let each control center implement the distributed identification filter (14) with $w_i(0) = x_i(0)$. Assume that the attack K affects only the i-th subsystem, that is $K = K_i$. Assume that $(E_i, A_i, [B_i^{\mathrm{b}} F_i\, B_{K_i}], C_i)$ and $(E_i, A_i, B_i^{\mathrm{b}}, C_i)$ have no invariant zeros. Then,

(i) $r_i(t) \neq 0$ at some time t; and
(ii) either $r_j(t) = 0$ for all $j \in \mathcal{N}_i^{\mathrm{out}}$ at all times t, or $r_j(t) \neq 0$ for all $j \in \mathcal{N}_i^{\mathrm{out}}$ at some time t.

Following Lemma 4.6 the region under attack can be identified through a distributed procedure. Indeed, the i-th area is safe if either of the following two criteria is satisfied:

(C1) the corresponding residual r_i is identically zero; or
(C2) the neighboring areas $j \in \mathcal{N}_i^{\mathrm{out}}$ feature both zero and nonzero residuals r_j.

Consider now the case of several simultaneously corrupted subsystems. Then, if the graphical distance between any two corrupted areas is at least 2, that is, if there are at least two uncorrupted areas between any two corrupted areas, corrupted areas can be identified through criteria (C1), (C2).

(S4: local identification) Once the corrupted regions have been identified, the identification method in Section IV-C is used to identify the local attack set.

Lemma 4.7: ***(Local identification)*** Consider the decoupled system (13). Assume that the i-th region is under the attack K_i whereas the neighboring regions $\mathcal{N}_i^{\mathrm{out}}$ are uncorrupted. Assume that each control center $j \in \mathcal{N}_i^{\mathrm{in}}$ transmits the estimate $\hat{x}_j(t)$ and the uncertainty subspace F_i to the i-th control center. Then, the attack set K_i is identifiable by the i-th control center if $(E_i, A_i, [B_i^{\mathrm{b}} F_i \;\; B_{K_i} \;\; B_{R_i}], C_i, [D_{K_i} \;\; D_{R_i}])$ has no invariant zeros for any attack set R_i, with $|R_i| \leq |K_i|$.

Notice that in **(S4)** identification is implemented only on the corrupted regions. Consequently, the combinatorial complexity of our distributed identification procedure is $\sum_{i=1}^{\ell} \binom{n_i+p_i}{|K_i|}$, where ℓ is the number of corrupted regions. Hence, the distributed identification procedure greatly reduces the combinatorial complexity of the centralized procedure presented in Subsection IV-C, which requires the implementation of $\binom{n+p}{|K|}$ filters. A related example is in [1].

V. A Case Study: Malicious Coordinated Attacks in Power Networks

Motivated by [13], in this section we study malicious attacks in a competitive power generation environment. In particular, we employ the results developed in the previous sections to cast attacks, rather than to design monitors.

Consider a connected power transmission network with n generators $G_{\mathrm{m}} = \{g_1, \dots, g_n\}$, where the rotor dynamics of each generator are modeled by second-order linear swing equations subject to governor control, and the power flows along lines are modeled by the DC approximation. Assume that a subset $K = \{k_1, \dots, k_m\}$ of m generators is driven by an additional control action besides the primary frequency control. After elimination of the load bus variables through Kron reduction, the power network dynamics subject to the additional control u at the generators K read as (see [1])

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{15}$$

[2] For nonsingular systems without feedthrough matrix, $\mathcal{F}_i$ is the largest $(A_i, \mathrm{Im}(B_i^{\mathrm{b}}))$-controlled invariant subspace contained in $\mathrm{Ker}(C_i)$ [23].

[3] See Section IV-C for a detailed construction of this type of filter.

where $x = [\theta^\mathsf{T}, \omega^\mathsf{T}]^\mathsf{T}$ contains the generator rotor angles and frequencies at time t, $A \in \mathbb{R}^{2n\times 2n}$, $C \in \mathbb{R}^{m\times 2n}$, and $B = I_K \in \mathbb{R}^{2n\times m}$, where $I_K = [e_{n+k_1} \cdots e_{n+k_m}]$ with e_i being the i-th canonical vector in $\mathbb{R}^{2n}$.

In [13] the following competitive scenario is considered: the group of generators K form a coalition, one sacrificial machine $\bar{k} \in K$ is selected in the coalition, and a specific coordinated control strategy is proposed for the generators K to destabilize the other machines $G_\mathrm{m} \setminus K$, while maintaining satisfactory performance within the group $K \setminus \{\bar{k}\}$. We now provide a general characterization of *all* possible strategies available to the generators K to compromise the behavior of the machines $G_\mathrm{m} \setminus K$ and, possibly, of a subset $\bar{K} \subseteq K$ of sacrificial machines. We refer the reader to [23] for the notion of *controlled invariant* and *conditioned invariant* subspaces.

Theorem 5.1: ***(Malicious attacks)*** Consider the network-reduced power system model (15) with controlled generators K and sacrificial machines $\bar{K} \subseteq K$. Let $\bar{C} = I^\mathsf{T}_{K\setminus\bar{K}}$, let $\mathcal{V}^*$ be the largest $(A, \mathrm{Im}(B))$-controlled invariant subspace contained in $\mathrm{Ker}(\bar{C})$, let the state feedback F satisfy $(A + BF)\mathcal{V}^* \subseteq \mathcal{V}^*$, let $\bar{B} = \mathrm{Basis}(\mathcal{V}^* \cap \mathrm{Im}(B))$, and let $\mathcal{S}^*$ be the smallest $(A, \mathrm{Ker}(\bar{C}))$-conditioned invariant subspace containing $\mathrm{Im}(B)$. Let $\bar{B} \in \mathbb{R}^{n\times\bar{m}}$. Then,

(i) for every input $v : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, the input $u = Fx + \bar{B}^\dagger v$ does not affect the generators $K \setminus \{\bar{K}\}$;

(ii) the subspace $\mathcal{V}^* \cap \mathcal{S}^*$ denotes the set of states reachable without affecting the generators $K \setminus \{\bar{K}\}$; and

(iii) any state in $\mathcal{V}^* \cap \mathcal{S}^*$ can be reached with an input of the form $u = Fx + \bar{B}^\dagger v$.

Proof: Define the nonsingular transformation matrix $T = [T_1, T_2, T_3]$, with $T_1 = \mathrm{Basis}(\mathcal{V}^* \cap \mathcal{S}^*)$, $T_2 = \mathrm{Basis}(\mathcal{V}^*)$, and T_3 such that T is nonsingular. In the new $z = T^{-1}x$ coordinates, the system matrices are

$$T^{-1}(A + BF)T = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix},\ T^{-1}B = \begin{bmatrix} B_1 \\ 0 \\ B_3 \end{bmatrix},$$

$$CT = \begin{bmatrix} 0 & 0 & C_3 \end{bmatrix},\ \mathrm{Basis}(T^{-1}B\bar{B}^\dagger) = \begin{bmatrix} B_1^\mathsf{T} & 0 & 0 \end{bmatrix}^\mathsf{T}, \quad (16)$$

where the zero pattern is due to the invariance properties of $\mathcal{V}^*$ and $\mathcal{S}^*$. As a consequence of the above decomposition, any input $u = Fx + \bar{B}^\dagger v$ does not affect the output, and therefore it does not affect the state variables associated with the generators $K \setminus \{\bar{K}\}$. Statements (ii) and (iii) are a direct consequence of the above decomposition; see [23]. ■

The following remarks are in order. First, the result in [13] is a special case of Theorem 5.1, since a destabilizing state feedback can be obtained by properly choosing v. Second, Theorem 5.1 characterizes the states reachable by a set of malicious generators K. If a specific desired state should be contained within this reachable set, then the set of malicious generators K should be selected accordingly. We leave this interesting aspect of coordinated attack design as the subject of future research. Third, the inputs u in Theorem 5.1 correspond to the attacks that can be cast by K independently of the system state while being undetectable by $K \setminus \bar{K}$; see Theorem 3.3 and the notions of left-invertibility [23]. Fourth, as a consequence of Theorem 3.5, if the set of sacrificial machines $\bar{K}$ is not empty, then there exist attacks as in Theorem 5.1. Finally, the input v can be designed as to optimize some performance function, such as, for instance, the effect of the malicious control on the sacrificial machines, the energy of the malicious control, or the information pattern required to implement the malicious control.

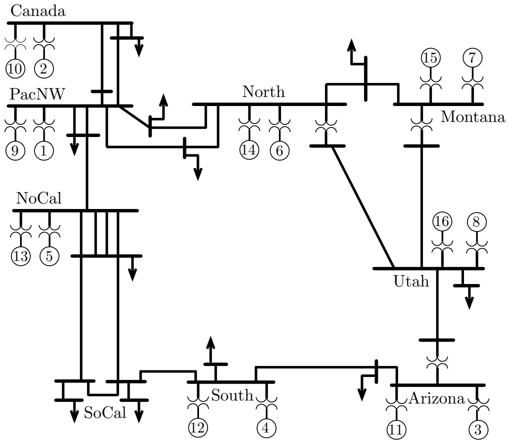

Fig. 2. A schematic diagram of the Western North American power grid.

As an example, consider an aggregated model of the Western North American power grid as illustrated in Fig. 2. This model is often studied [33] in the context of wide-area oscillations. Assume that the generators $\{1, 9\}$ are being controlled, and that generator 9 is the sacrificial machine. Following Theorem 5.1, a malicious attack $u = Fx + \bar{B}^\dagger v$ is cast by the generators $\{1, 9\}$ such that generator 1 is not affected by the attack. Additionally, the input v is optimally chosen such that generator 2 maintains an acceptable working condition even in the presence of the attack, and large frequency deviations are induced at all other generators $G_\mathrm{m} \setminus K$. As a consequence, the linear model (15) is driven far away from the operating point, and the corresponding original nonlinear model eventually loses synchrony. In a real-world scenario the affected generators $G_\mathrm{m} \setminus K$ would be disconnected for safety reasons.

In the above scenario, assume that each generator monitors its own state variables, and that at most two generators may be colluding to disrupt the network. Notice that detectability of the malicious attacks designed in Theorem 5.1 is guaranteed for each generator affected by the attack. Unfortunately, no generator can identify the colluding generators while relying only on its own measurements. To see this, let B_K be the input matrix associated with any set K of two generators, and let $C_i = e_i^\mathsf{T}$ be the output matrix associated with generator i. It can be verified that for every K and i the system (A, B_K, C_i) is right-invertible [23]. Hence, no generator alone can identify the malicious generators, and a coalition of multiple sensors becomes necessary.

VI. Conclusion

For cyber-physical systems modeled by linear time-invariant descriptor systems, we have analyzed fundamental monitoring limitations. In particular, we have character-

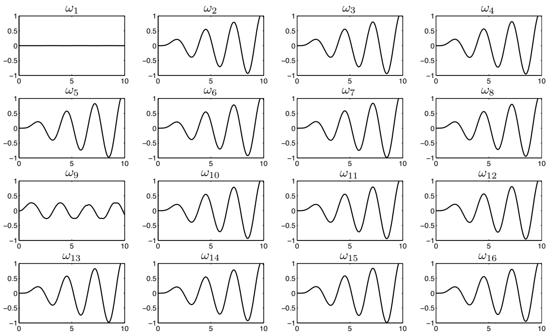

Fig. 3. The deviations of the generators frequencies from their steady state value induced by a malicious attack is here reported. The attack is designed by using the result in Theorem 5.1. In particular, the input v is chosen such that the infinity norm of ω_9 is minimized, subject to the infinity norm of ω_{16} being no less than 1. Notice that generator 1 is not affected by the attack, and that generator 9 maintains satisfactory performance. Instead, the other generators are severely affected by the coordinated attack.

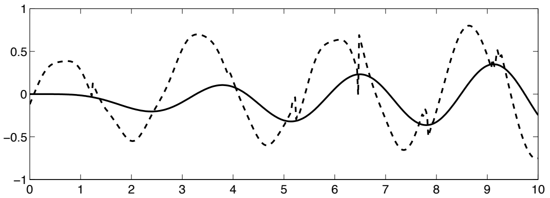

Fig. 4. This figures shows the governor control input injected by generator 1 (solid) and by generator 9 (dashed). Both plots are in p.u. values and for the linear system (15), that is, measured as deviation from the steady state.

ized undetectable and unidentifiable attacks from a system-theoretic and a graph-theoretic perspective. Additionally, we have designed centralized and decentralized monitors. Finally, we have demonstrated the effectiveness of our findings by designing unidentifiable attacks against a simplified model of the Western North American power grid. Interesting future directions include the extensions of the results in this paper to the noisy and nonlinear case, as well as investigating scenarios where attackers have limited capabilities.

REFERENCES

[1] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, Aug. 2012, submitted.

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

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

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

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

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

[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] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

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

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

[25] 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.

[26] 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.

[27] F. Dörfler, F. Pasqualetti, and F. Bullo, "Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach," in *Allerton Conf. on Communications, Control and Computing*, Allerton, IL, USA, Sep. 2011, pp. 1486–1491.

[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] 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.

[30] 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.

[31] F. Dörfler, F. Pasqualetti, and F. Bullo, "Continuous-time distributed estimation with discrete communication," *Journal of Selected Topics in Signal Processing*, Aug. 2012, Submitted.

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

[33] D. J. Trudnowski, J. R. Smith, T. A. Short, and D. A. Pierre, "An application of Prony methods in PSS design for multimachine systems," *IEEE Transactions on Power Systems*, vol. 6, no. 1, pp. 118–126, 1991.