Distributed Detection of Cyber-Physical Attacks in Power Networks: A Waveform Relaxation Approach

Florian Dörfler, Fabio Pasqualetti, and Francesco Bullo

Abstract— **Future power grids will be required to operate safely and reliably against cyber-physical attacks. The large dimensionality and the difficulty in calibrating dynamical network models precludes the use of centralized attack detection algorithms. This paper proposes a unified modeling framework and an advanced detection procedure whose implementation requires only local network knowledge. We model a power network as a linear time-invariant descriptor system and cyber-physical attacks as unknown inputs. This modeling framework captures, for instance, network components malfunction and measurements corruption. In our detection method the power network is partitioned among geographically deployed control centers, possibly located at transmission substations. Each control center has knowledge of only its respective subarea dynamics, is able to acquire information from neighboring areas, and is capable of performing basic computations. Under these minimal technological requirements and a reasonable observability assumption, we design an entirely distributed detection filter which requires only local network knowledge and yet achieves guaranteed global performance. Our detection filter is based on a sparse residual filter in descriptor form, which can be stabilized via decentralized output injection and implemented distributively via waveform relaxation.**

I. Introduction

Cyber-physical security is a topic of primary concern in the envisioned smart power grid [1]–[3]. Besides failures and attacks on the *physical* power grid infrastructure, the smart grid is also prone to *cyber* attacks on its communication and computation layer. In short, cyber-physical security is a fundamental obstacle challenging the smart grid vision.

Related work. While the security of the electricity network has always been an important research subject, there has been a recent explosion of publications concerning cyber-physical security in smart power grids. Traditionally, state estimation and detection procedures have been designed for static power network models [4]–[6], but the development of security procedures that exploit the power network dynamics has been recognized [7] as an important problem. In this work, we consider the linearized version of the classic *structure-preserving power network model* [8], which is composed by the linearized swing equation for the generator rotor dynamics and the DC power flow equation for the loads. The resulting linear continuous-time descriptor model of a power network has also been studied for dynamic estimation, fault detection, and security assessment in [9]–[13].

The difficulty of obtaining and calibrating accurate wide-area dynamical power system models, the low-bandwidth communication capabilities versus the extremely fast physical dynamics, and the high dimensionality of the electric power network preclude the use of centralized estimation, detection, and identification procedures. One possibility to overcome these issues is to geographically deploy some monitors in the network, each one responsible for a different subarea of the whole system, as illustrated in Fig. 1. Local estimation schemes can successively be used, together with an information exchange mechanism to recover the performance of a centralized scheme. Such *distributed* monitoring paradigm has been applied to static and discrete-time state space power network models [6], [9]. In this paper we consider a dynamic descriptor model and exploit the classic *waveform relaxation* method to develop a fully distributed attack detection procedure. The waveform relaxation method is an extension of the classic relaxation method for systems of algebraic equations, with the difference that the iteration is carried out over functions (or waveforms) rather than vectors. We refer the reader to [14]–[16] for a comprehensive discussion of waveform relaxation methods.

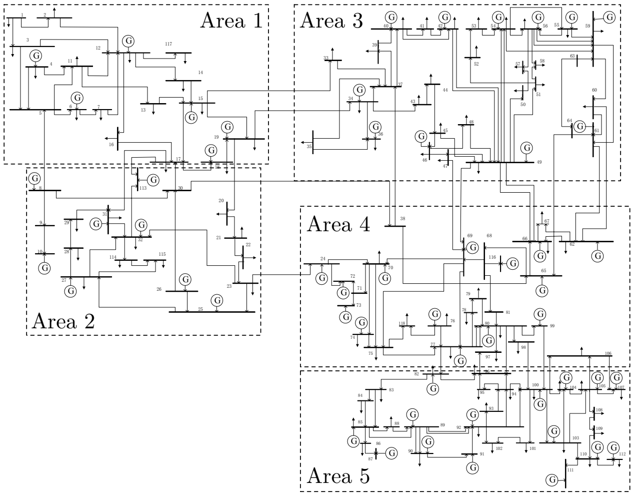

Fig. 1. Partition of IEEE 118 bus system into 5 areas. Each area is monitored and operated by a control center. The control centers cooperate to estimate the state and to assess the functionality of the whole network.

Contributions. This paper features three contributions.

First, we propose a setup for distributed estimation and detection of cyber-physical attacks in large-scale interconnected power networks. Analogously to [10]–[13], we model a power network by a linear time-invariant descriptor system. Cyber-physical attacks on the power network are modeled as unknown inputs affecting the system state and measurements.

This material is based in part upon work supported by NSF grants IIS-0904501 and CPS-1135819.

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

Our model can represent either genuine faults of network components or malicious attacks on sensors, actuators, and the communication infrastructure. Finally, following [6], we propose a block-diagonal partition of the power network model according to geographically deployed control centers equipped with computation and communication capabilities. We remark that the considered descriptor model is very general and features no power network specific assumptions such as invertible algebraic equations (index one). Consequently, our detection methods are also applicable to other large-scale interconnected systems described by descriptor models, such as water, gas, and sensor networks.

Second, starting from the centralized filter presented in our earlier work [13], we develop a fully distributed attack detection filter. In a first design step, we propose a centralized but sparse residual filter in descriptor form to detect cyber-physical attacks. Contrary to the treatment in [13], this attack detection filter is sparse and thus amenable to distributed implementation. Next, we stabilize the attack detection filter based on a decentralized output injection. In a final design step, we distribute the attack detection filter by iterative local computations using the Gauss-Jacobi waveform relaxation technique. In the end, we propose a fully distributed attack detection filter that achieves guaranteed global performance. The implementation of this filter requires communication among the local control centers, local observability of each subnetwork, and a block-diagonal dominance condition to be satisfied. The latter condition can be verified locally and ensures both the decentralized stabilization of the filter as well as the convergence of the waveform relaxation iteration.

Third and finally, we illustrate the performance of our distributed attack detection filter with an example of cyber attack on the IEEE 118 bus system. We show that the control centers cooperatively detect the attack, although none of them knows the network model and measurements entirely.

Paper organization. The remainder of this paper is organized as follows. Section II presents the problem setup and some preliminary results on centralized attack detection. In Section III we develop a distributed attack detection filter, and in Section IV we demonstrate its performance through a numerical example. Finally, Section V concludes the paper.

II. Problem setup and centralized detection

A. Mathematical model of dynamical systems under attack

Consider the linear time-invariant descriptor system

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

where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is the state, $y : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$ is the output with measurement matrix $C \in \mathbb{R}^{p\times n}$, and the state matrices $E \in \mathbb{R}^{n\times n}$ and $A \in \mathbb{R}^{n\times n}$ are assumed to satisfy

(A0) the matrix E is diagonal and, possibly, singular;

(A1) the pair (E, A) is regular, that is, $\det(sE - A)$ does not vanish for all $s \in \mathbb{C}$.

The structural assumption (A1) guarantees the existence of a unique solution $x(t)$, and it is typically satisfied when circuits and power networks are modeled by linear descriptor systems, see [10]–[13]. We refer the reader to [17]–[19] for a comprehensive discussion of descriptor systems. We remark that assumption (A0) is automatically verified for power systems [10], and, although it simplifies the notation, it is not necessary for the derivation of our results.

We allow for the presence of unknown disturbances affecting the behavior of the system (1), which, besides reflecting the genuine failure of system components, can be the effect of an attack against the cyber-physical system. We classify these disturbances into *state attacks*, if they show up in the measurements vector after being integrated through the network dynamics, and *output attacks*, if they corrupt directly the measurements vector. The dynamics of the descriptor system (1) in the presence of an attack can be written as

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \end{aligned} \tag{2}$$

where $B \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{p\times m}$, and $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$. The input signal $u(t)$, as well as the input matrices B and D are assumed to be *unknown* and arbitrary. We let $y(x_0, u, t)$ be the output signal generated from the initial state x_0 under the attack signal $u(t)$. We refer to the triple $(B, D, u(t))$ as *cyber-physical attack*, and make the following assumptions on the descriptor system (2):

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

(A3) the input signal $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ is smooth.

Assumptions (A2) and (A3) simplify the technical presentation in this paper since they guarantee smoothness of the state trajectory $x(t)$, $t \in \mathbb{R}_{\geq 0}$; see [20, Lemma 2.5] for further details. However, we remark that the results in this paper can also be established under weaker assumptions. If the consistency assumption (A2) is dropped, then the additional cases of initial jumps and impulses in the state $x(t \downarrow 0)$ have to be considered that possibly affect the initial measurements $y(t \downarrow 0)$. Hence, in presence of inconsistent initial conditions, the results in paper are valid only for strictly positive times $t > 0$. For power networks models, the smoothness assumption (A3) can actually be replaced by continuity of $u(t)$ (since these models are of index one [12], [13]). If assumption (A3) is further weakened to $u(t)$ belonging to the class of *impulsive smooth distributions*, then a powerful attacker capable of commanding an impulsive input $u(t^*)$ at some time t^* can directly reset the state $x(t^*)$ [20, Theorem 3.2] and, possibly, evade detection.

B. Centralized attack detectability

A cyber-physical attack may remain undetected from the measurements if there exists a normal operating condition of the network under which the output would be the same as under the perturbation due to the attacker.

*Definition 1 (***Undetectable attack***):* For the linear descriptor system (2), the attack $(B, D, u(t))$ is *undetectable* if there exist initial conditions $x_1, x_2 \in \mathbb{R}^n$, such that, for all $t \in \mathbb{R}_{\geq 0}$, $y(x_1, u, t) = y(x_2, 0, t)$.

Necessary and sufficient algebraic conditions for the detectability of cyber-physical attacks are described in [13], while graph-theoretic conditions are given in [12]. In [13] we present a centralized method for the detection of attacks based upon Kron reduction [21] of the algebraic equations in model (2), which results in a convenient but non-sparse state-space detection filter. In what follows, we present a similar but *sparse* centralized attack detection filter. This sparsity will be key to develop a distributed detection method later on.

*Theorem 2.1 (***Centralized attack detection filter):** Consider the descriptor system (2) and assume that the attack is detectable, and that the network initial state $x(0)$ is known. Consider the *centralized attack detection filter*

$$\begin{aligned} E\dot{w}(t) &= (A + GC)w(t) - Gy(t), \\ r(t) &= Cw(t) - y(t), \end{aligned} \tag{3}$$

where $w(0) = x(0)$ and the output injection $G \in \mathbb{R}^{n \times p}$ is such that the generalized eigenvalues of the pair $(E, A+GC)$ have negative real part. Then $r(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$ if and only if $u(t) = 0$ at all times $t \in \mathbb{R}_{\geq 0}$. □

Proof: Consider the error $e(t) = w(t) - x(t)$ between the dynamic states of the filter (3) and the descriptor system (2). The error dynamics with output $r(t)$ are given by

$$\begin{aligned} E\dot{e}(t) &= (A + GC)e(t) - (B + GD)u(t), \\ r(t) &= Ce(t) - Du(t), \end{aligned} \tag{4}$$

where $e(0) = 0$. To prove the theorem we show that the error system (4) has no invariant zeros, that is, $r(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$ if and only if $u(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$. Since the initial condition $x(0)$ and the input $u(t)$ are assumed to be consistent (A2) and non-impulsive (A3), the error system (4) has no invariant zeros if and only if [20, Proposition 3.4] there exists no triple $(s, \bar{w}, g) \in \mathbb{C} \times \mathbb{R}^n \times \mathbb{R}^p$ satisfying

$$\begin{bmatrix} sE - (A + GC) & B + GD \\ C & -D \end{bmatrix} \begin{bmatrix} \bar{w} \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{5}$$

The second equation of (5) yields $C\bar{w} = Dg$. Thus, by substituting $C\bar{w}$ by Dg in the first equation of (5), the set of equations (5) can be equivalently written as

$$\begin{bmatrix} sE - A & B \\ C & -D \end{bmatrix} \begin{bmatrix} \bar{w} \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{6}$$

Finally, note that a solution $(s, -\bar{w}, g)$ to above set of equations would yield an invariant zero, zero state, and zero input for the descriptor system (2). By the detectability assumption,[1] the descriptor model (2) has no zero dynamics and the matrix pencil (6) necessarily has full rank. It follows that the triple (E, A, C) is observable, so that G can be chosen such that the pair $(E, A+GC)$ is Hurwitz [19, Theorem 4.1.1], and the error system (4) has no zero dynamics. This concludes the proof of Theorem 2.1. ■

*Remark 1 (***Detection with unknown initial conditions):** If the network initial state $x(0)$ is not available, then an arbitrary filter initial state $w(0) \in \mathbb{R}^n$ can be chosen. Consequently, the performance of the detection filter (3) becomes asymptotic, and some attacks may remain undetected. For instance, if the generalized eigenvalues of the detection filter pair $(E, A + GC)$ have been assigned to have real part smaller than some constant $c < 0$, then, in the absence of attacks, the filter output $r(t)$ exponentially converges to zero with rate less than c [19, Section 3.1.1]. Hence, only inputs $u(t)$ that vanish faster or equal than e^{-ct} can remain undetected by the filter (3). □

III. Distributed detection

A. Setup for distributed detection

Let $G = (V, \mathcal{E})$ be the directed graph associated with the pair (E, A), i.e., the graph describing the interconnection structure of the state variables. In particular, each element of V corresponds to a system state, and there is a directed edge from vertex j to vertex i if the entry a_{ij} or e_{ij} is nonzero. Assume that V has been partitioned as $V = V_1 \cup \cdots \cup V_N$, and let $G_i = (V_i, \mathcal{E}_i)$, with $i \in \{1, \dots, N\}$, be the subgraph of G with vertices V_i and edges $\mathcal{E} \cap (V_i \times V_i)$. Let $|V_i| = n_i$. According to this partition and possibly after relabeling the nodes, the matrices E and A in (1) can be written as

$$E = \text{blkdiag}(E_1, \dots, E_N), \; A = \begin{bmatrix} A_1 & \cdots & A_{1N} \\ \vdots & \vdots & \vdots \\ A_{N1} & \cdots & A_N \end{bmatrix}.$$

where $E_i, A_i \in \mathbb{R}^{n_i \times n_i}$, and $A_{ij} \in \mathbb{R}^{n_i \times n_j}$. Furthermore, assume that the output matrix C in (2) reads as

$$C = \text{blkdiag}(C_1, \dots, C_N),$$

where $C_i \in \mathbb{R}^{p_i \times n_i}$. Given such a partition and in the absence of attacks, the descriptor system (1) can be written as the interconnection of N subsystems of the form

$$\begin{aligned} E_i\dot{x}_i &= A_i x_i(t) + \sum_{j=1, j\neq i}^{N} A_{ij}x_j(t), \\ y_i(t) &= C_i x_i(t), \quad i \in \{1, \dots, N\}, \end{aligned} \tag{7}$$

where $x_i(t)$ and $y_i(t)$ are the state and output of the i-th subsystem. We assume the presence of a *control center* in each subnetwork G_i with the following capabilities:

(A4) the i-th control center knows only the diagonal matrices E_i, A_i, and C_i, as well as the neighbouring matrices A_{ij}, $j \in \{1, \dots, N\} \setminus \{i\}$;

(A5) the i-th control center can exchange information with control center j if the matrix A_{ij} is non-zero; and

(A6) the pair (E_i, A_i) is regular, and the triple (E_i, A_i, C_i) is observable.

Under assumptions (A4), (A5), and (A6), we consider the problem of designing a distributed algorithm for the control centers to cooperatively detect cyber-physical attacks.

B. Decentralized detection

Before deriving a fully distributed version of the attack detection filter (3), we explore the question of *decentralized stabilization* of the error dynamics of the filter (3). For each subsystem (7), consider the local residual generator

$$\begin{aligned} E_i\dot{w}_i(t) &= (A_i + G_iC_i)w_i(t) + \sum_{j=1, j\neq i}^{N} A_{ij}x_j(t) - G_iy_i(t), \\ r_i(t) &= y_i(t) - C_iw_i(t), \quad i \in \{1, \dots, N\}, \end{aligned} \tag{8}$$

where $w_i(t)$ is the i-th estimate of $x_i(t)$ and $G_i \in \mathbb{R}^{n_i \times p_i}$. In order to derive a compact formulation, let $w(t) =$

[1] Due to linearity of the descriptor system (2), the detectability assumption reads as "the attack $(B, D, u(t))$ is detectable if there exist no initial conditions $x_0 \in \mathbb{R}^n$, such that $y(x_0, u, t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$."

$[w_1^\mathsf{T}(t) \cdots w_N^\mathsf{T}(t)]^\mathsf{T}$ and $r(t) = [r_1^\mathsf{T}(t) \cdots r_N^\mathsf{T}(t)]^\mathsf{T}$, and define the matrices $A_D = \text{blkdiag}(A_1, \dots, A_N)$, $A_C = A - A_D$, and $G = \text{blkdiag}(G_1, \dots, G_N)$. The interconnection structure among the subsystems is described by the matrix A_C and the overall filter dynamics (8) read in vector form as

$$\begin{aligned} E\dot{w}(t) &= (A_D + GC)w(t) + A_C w(t) - Gy(t)\,, \\ r(t) &= y(t) - Cw(t)\,. \end{aligned} \tag{9}$$

Due to the observability assumption (A6) each output injection matrix G_i can be chosen such that $(E_i, A_i - G_i C_i)$ is Hurwitz [19, Theorem 4.1.1]. Notice that if each pair $(E_i, A_i + G_i C_i)$ is regular and Hurwitz, then $(E, A_D + GC)$ is also regular and Hurwitz since the matrices E and $A_D + GC$ are block-diagonal. We are now ready to state a condition for the decentralized stabilization of the filter (9).

Lemma 3.1: **(Decentralized stabilization):** Consider the filter dynamics (9), and let $G = \text{blkdiag}(G_1, \dots, G_N)$ be such that $(E, A_D + GC)$ is regular and Hurwitz. The filter error $x(t) - w(t)$ is asymptotically stable if

$$\rho\left((j\omega E - A_D - GC)^{-1} A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}\,, \tag{10}$$

where $\rho(\cdot)$ denotes the spectral radius operator.

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

$$\begin{aligned} E\dot{e}(t) &= (A_D + A_C + GC)e(t), \\ r(t) &= Ce(t)\,. \end{aligned} \tag{11}$$

We employ a small-gain approach to large-scale interconnected systems [22] and rewrite the error dynamics (11) as the closed-loop interconnection of the two subsystems

$$\begin{aligned} \Gamma_1 : &\quad E\dot{e}(t) = (A_D + GC)e(t) + u(t)\,, \\ \Gamma_2 : &\quad u(t) = A_C e(t)\,. \end{aligned}$$

Since both subsystems Γ_1 and Γ_2 are causal and internally Hurwitz stable, the overall error dynamics (11) are stable if the loop transfer function, say $\Gamma_1(j\omega) \cdot \Gamma_2$, satisfies the spectral radius condition $\rho(\Gamma_1(j\omega) \cdot \Gamma_2) < 1$ for all $\omega \in \mathbb{R}$ [23, Theorem 4.11]. The latter condition is equivalent to (10) ■

It should be observed that, even if each subsystem is assumed to be observable, the stability of the decentralized filter depends on the off-diagonal blocks of the system matrix, and it cannot be always achieved. Moreover notice that, although each control centers can compute the output injection matrix independently of each other, the decentralized attack detection filter (9) requires the control center to continuously exchange their local estimation vector. Hence, this scheme has high communication complexity, and may be applicable only in particular scenarios. A solution to this problem is presented in the next section.

C. Waveform relaxation method

In this subsection we exploit the classic waveform relaxation method to develop a fully distributed variation of the decentralized attack detection filter (9). The Gauss-Jacobi waveform relaxation method applied to the system (9) yields the *waveform relaxation iteration*

$$E\dot{w}^{(k)}(t) = A_D w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t)\,, \tag{12}$$

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 the initial condition at each iteration is $w^{(k)}(0) = w_0$. Notice that (12) is a descriptor system in the variable $w^k(t)$ and the vector $A_C w^{(k-1)}(t)$ is a known input, since the value of $w(t)$ at iteration $k-1$ is used. The iteration (12) is said to be *convergent* if

$$\lim_{k\to\infty} w^{(k)}(t) - w(t) = 0\,, \quad t \in [0, T]\,,$$

where $w(t)$ is the solution of the non-iterative dynamics (9). In order to obtain a low-complexity distributed detection scheme, we use the waveform relaxation iteration (12) to iteratively approximate the decentralized filter (9). We start by presenting a convergence condition for the iteration (12).

Recall that a function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$ is said to be of *exponential order* β if there exists $\beta \in \mathbb{R}$ such that the exponentially scaled function $\tilde{f} : \mathbb{R}_{\geq 0} \to \mathbb{R}^p$, $\tilde{f}(t) = f(t)e^{-\beta t}$ and all its derivatives exist and are bounded. An elegant analysis of the waveform relaxation iteration (12) can be carried out in the Laplace domain [24], where the operator mapping $w^{(k-1)}(t)$ to $w^{(k)}(t)$ is given by $(sE - A_D - GC)^{-1} A_C$. As in the analysis of the regular Gauss-Jacobi iteration, convergence of the waveform relaxation iteration (12) follows from contractivity of the iteration operator.

Lemma 3.2: **(Convergence of the waveform relaxation [24, Theorem 5.2]):** Consider the waveform relaxation iteration (12). Assume that the pair $(E, A_D + GC)$ is regular and the initial condition w_0 is consistent. Let $y(t)$, with $t \in [0, T]$, be of exponential order β. Let α be the least upper bound on the real part of the spectrum of (E, A), and define $\sigma = \max\{\alpha, \beta\}$. The waveform relaxation method (12) is convergent if

$$\rho\left(((\sigma + j\omega)E - A_D - GC)^{-1} A_C\right) < 1 \text{ for all } \omega \in \mathbb{R}\,. \tag{13}$$

In the reasonable case of bounded (integrable) measurements $y(t)$, $t \in [0, T]$, and stable filter dynamics, we have that $\sigma = \alpha = \beta = 0$, and the convergence condition (13) for the wave-form relaxation iteration (12) equals the condition (10) for decentralized stabilization of the filter dynamics.

Remark 2 (**Distributed implementation):** The waveform relaxation iteration (12) can be implemented in the following distributed fashion. Assume that each control center $i \in \{1, \dots, N\}$ is able to integrate the descriptor system

$$\begin{aligned} E_i \dot{w}_i^{(k)}(t) =& (A_i + G_i C_i) w_i^{(k)}(t) \\ &+ \sum_{j=1, j\neq i}^{N} A_{ij} w_j^{(k-1)}(t) - G_i y_i(t)\,, \end{aligned} \tag{14}$$

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

(1) set $k := k + 1$, and compute the signal $w_i^{(k)}(t)$ by integrating equation (14),

(2) transmit $w_i^{(k)}(t)$ to the j-th control center if $A_{ij} \neq 0$,

(3) update the input $w_j^{(k)}$ with the signal received from the j-th control center and iterate.

If the waveform relaxation is convergent, then, for k sufficiently large, the residuals $r_i^{(k)}(t) = y_i(t) - C_i w_i^{(k)}(t)$ can be used to detect attacks; see Theorem 3.3. In summary, a distributed implementation of the iteration (12) requires integration capabilities at each control center, knowledge of the measurements $y_i(t)$, $t \in [0, T]$, as well as synchronous communication between neighboring control centers. □

D. Distributed attack detection filter

We now propose our distributed attack detection filter.

Theorem 3.3: (**Distributed attack detection filter):** Consider the descriptor system (2) and assume that the attack is detectable, and that the network initial state $x(0)$ is known. Let the assumptions (A1) through (A6) be satisfied and consider the *distributed attack detection filter*

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

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 $G = \operatorname{blkdiag}(G_1, \dots, G_N)$ is such that the pair $(E, A_D + GC)$ is regular, Hurwitz, and

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

Then $\lim_{k\to\infty} r^{(k)}(t) = 0$ at all times $t \in [0, T]$ if and only if $u(t) = 0$ at all times $t \in [0, T]$.

Proof: Since the initial condition $w^{(k)}(0) = x(0)$ is consistent, it follows from Lemma 3.2 that the solution $w^{(k)}(t)$ of the iteration (15) converges, as $k \to \infty$, to the solution $w(t)$ of the non-iterative filter dynamics (9) if condition (13) is satisfied with $\sigma = 0$ (due to integrability of $y(t)$, $t \in [0, T]$, and since the pair $(E, A_D + GC)$ is Hurwitz). The latter condition is equivalent to condition (16).

Under condition (16) and due to the Hurwitz assumption, it follows from Lemma 3.1 that the error $e(t) = x(t) - w(t)$ between the state $x(t)$ of the descriptor model (2) and the state $w(t)$ of the decentralized filter dynamics (9) is asymptotically stable. Thus, the pair $(E, A_D + A_C + GC) = (E, A + GC)$ is Hurwitz. Due to the detectability assumption and by analogous reasoning as in the proof of Theorem 2.1, it follows that the error dynamics $e(t)$ have no invariant zeros. This concludes the proof of Theorem 3.3. ■

It should be observed that the distributed attack detection filter (15) needs to be implemented in a receding-horizon fashion. Indeed, the control centers collect measurements and check for attacks every time window of length T.

*Remark 3 (***Distributed filter design):** As discussed in Remark 2, the filter (15) can be implemented in a distributed fashion. In fact, it is also possible to design the filter (15), i.e., the output injections G_i, in an entirely distributed way. Since $\rho(A) \leq \|A\|_p$ for any matrix A and any induced p-norm, condition (16) can be relaxed by the small gain criterion to

$$\left\|(j\omega E - A_D - GC)^{-1} A_C\right\|_p < 1 \text{ for all } \omega \in \mathbb{R}\,. \tag{17}$$

With $p = \infty$, in order to satisfy condition (17), it is sufficient for each control center i to verify the *quasi-block diagonal dominance* condition [25]

$$\left\|(j\omega E_i - A_i - G_i C_i)^{-1} \sum\nolimits_{k=1}^{n} A_{ik}\right\|_\infty < 1 \text{ for all } \omega \in \mathbb{R}. \tag{18}$$

Notice that condition (18) can be checked with only local information, and that, although fully distributed, it is a conservative relaxation of condition (16). In summary, each control center i needs to choose the output injection matrix G_i such that $A_i + G_i C_i$ is Hurwitz stable and the block-diagonal dominance condition (18) is satisfied. □

IV. Illustrative example

The IEEE 118 bus system represents a portion of the American Electric Power System as of December 1962. This test case system is composed of 118 buses, 186 branches, 54 generators, and 99 loads. The IEEE 118 bus system is illustrated in Fig. 1. The network parameters can be found for example in [26]. Following [12], a linear, continuous time, descriptor model of the network dynamics assumes the form

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + P_x(t), \\ y(t) &= Cx(t) + P_y(t), \end{aligned} \tag{19}$$

where, being n (resp. m) the number of generators (resp. loads), $E \in \mathbb{R}^{(2n+m)\times(2n+m)}$, $A \in \mathbb{R}^{(2n+m)\times(2n+m)}$, $C \in \mathbb{R}^{p\times 2n}$, $p \in \mathbb{N}$, and $P_x(t), P_y(t)$ are (known) vector-valued functions of time of appropriate dimension. Due to linearity of the system (2), the known inputs $P_x(t)$ and $P_y(t)$ will been neglected in the forthcoming analysis, since they do not affect the detectability of unknown input attacks.

For estimation and attack detection purposes, we partition, the IEEE 118 bus system into 5 disjoint areas, we assign a control center to each area, and we implement our procedure via the filter (15). See Fig. 1 for a graphical illustration. Suppose that each control center continuously measure 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 30, the attacker adds a signal $u(t)$ to the network measurements, so that the measurements equation becomes

$$y(t) = Cx(t) + Du(t),$$

where, at each time t, each component of the vector $u(t)$ is randomly distributed in the interval $[0, 0.5]$. We assume that the attack $u(t)$ is detectable, and we refer the reader to [13] for a detailed discussion of attack detectability. 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 $\rho\left(j\omega E - A_D - GC)^{-1} A_C\right) < 1$ for all $\omega \in \mathbb{R}$. Hence, as predicted by Theorem 3.3, our distributed attack detection filter is convergent (cf. Fig. 2).

Regarding the identification of the corrupted variables, we remark that a regional identification may be possible by analyzing the residual functions. In this example, for instance, since the residuals associated with the generators of the first area are much larger than the other residuals, the attacker is more likely to have corrupted the measurements of the first area. This important aspect of attack *identification* is left as the subject of future research.

For completeness, in Fig. 3 we illustrate the convergence 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 of our detection scheme.

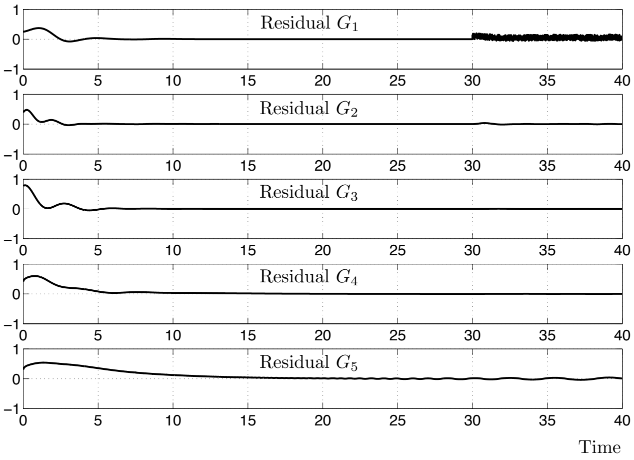

Fig. 2. In this figure we show the residual functions computed through the distributed attack detection filter (15). In particular, G_i represents the residual associated with a generator in the i-th area. The attacker compromises the measurements of all the generators in area 1 from time 30 with a signal uniformly distributed in the interval $[0, 0.5]$. The attack is correctly detected, because the residual functions do not decay to zero. For the simulation, we run 100 iterations of the attack detection method.

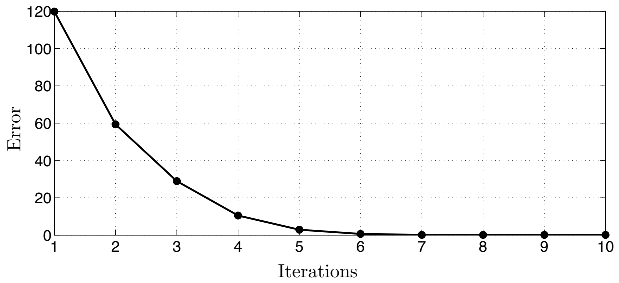

Fig. 3. The plot represents the error of our waveform relaxation based filter (15) with respect to the corresponding decentralized filter (9). On the abscissa axis we plot the infinity norm of the difference of the outputs of the two filters. As predicted by Theorem 3.3, the error is convergent.

V. Conclusions

We presented a fully distributed procedure for the detection of cyber-physical attacks in power networks modeled by linear descriptor systems. Our procedure is based on a sparse residual filter in descriptor form, which can be stabilized via decentralized output injection, and implemented distributively via waveform relaxation.

In future work, we plan to address the extension of the results in this paper to the attack identification problem. Of interest is also the question of optimal network partitioning so as to automatically verify the proposed spectral radius condition for the convergence of our distributed attack detection filter. Furthermore, modeling uncertainties, constraints on communication capabilities, and the presence of noise should be included.

References

[1] H. Khurana, "Cybersecurity: A key smart grid priority," *IEEE Smart Grid Newsletter*, Aug. 2011.

[2] M. Amin, "Guaranteeing the security of an increasingly stressed grid," *IEEE Smart Grid Newsletter*, Feb. 2011.

[3] 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.

[4] F. C. Schweppe and J. Wildes, "Power system static-state estimation, Part I: Exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–125, 1970.

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

[6] F. Pasqualetti, R. Carli, and F. Bullo, "Distributed estimation via iterative projections with application to power network monitoring," *Automatica*, Mar. 2011, to appear.

[7] N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M. G. Lauby, B. F. Wollenberg, and J. N. Wrubel, "On-line power system security analysis," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 262–282, 1992.

[8] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Prentice Hall, 1998.

[9] M. D. Ilić, X. Le, U. A. Khan, and J. M. F. Moura, "Modeling of future cyber-physical energy systems for distributed sensing and control," *IEEE Transactions on Systems, Man & Cybernetics. Part A: Systems & Humans*, vol. 40, no. 4, pp. 825–838, 2010.

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

[11] A. Domınguez-Garcıa and S. Trenn, "Detection of impulsive effects in switched DAEs with applications to power electronics reliability analysis," in *IEEE Conf. on Decision and Control*, Atlanta, GA, USA, Dec. 2010, pp. 5662–5667.

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

[13] 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, to appear.

[14] 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.

[15] J. White, F. Odeh, A. L. Sangiovanni-Vincentelli, and A. Ruehli, "Waveform relaxation: Theory and practice," EECS Department, University of California, Berkeley, Tech. Rep. UCB/ERL M85/65, 1985. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/1985/543.html

[16] M. L. Crow and M. D. Ilić, "The waveform relaxation method for systems of differential/algebraic equations," *Mathematical and Computer Modelling*, vol. 19, no. 12, pp. 67–84, 1994.

[17] F. L. Lewis, "A survey of linear singular systems," *Circuits, Systems, and Signal Processing*, vol. 5, no. 1, pp. 3–36, 1986.

[18] ——, "A tutorial on the geometric analysis of linear time-invariant implicit systems," *Automatica*, vol. 28, no. 1, 1992.

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

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

[21] F. Dörfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *SIAM Review*, Feb. 2011, submitted.

[22] M. Vidyasagar, *Input-Output Analysis of Large-Scale Interconnected Systems: Decomposition, Well-Posedness and Stability*. Springer, 1981.

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

[24] 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.

[25] Y. Ohta, D. Šiljak, and T. Matsumoto, "Decentralized control using quasi-block diagonal dominance of transfer function matrices," *IEEE Transactions on Automatic Control*, vol. 31, no. 5, pp. 420–430, 1986.

[26] 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.