Networked Control Systems under Cyber Attacks with Applications to Power Networks

André Teixeira, Henrik Sandberg, Karl H. Johansson

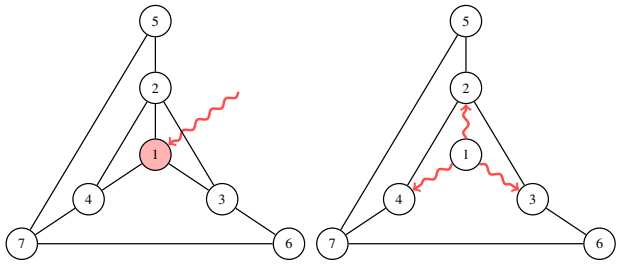
***Abstract*—Networked control systems under certain cyber attacks are analyzed. The communication network of these control systems make them vulnerable to attacks from malicious outsiders. Our work deals with two types of attacks: attacks on the network nodes and attacks on the communication between the nodes. We propose a distributed scheme to detect and isolate the attacks using observers. Furthermore, we discuss how to reduce the number of observer nodes while maintaining the coverage of the entire network. The results are applied to two classes of networked control systems: a network running the consensus protocol and a power network defined by the linearized swing equation. Sufficient conditions for the existence of the proposed attack detection scheme are provided for the first class of systems. For the second class, we provide a necessary condition for the existence of the proposed detection scheme.**

***Index Terms*—Networked Control Systems, Fault Detection, Power Systems**

I. Introduction

Several infrastructure systems are of major importance to society, as they became part of daily life and are now indispensable. Examples include the power grid, telecommunication systems, and water supply. Such systems are referred to as critical infrastructures. The malfunction of critical infrastructures has a great impact not only on people's lives but also on the economy. These systems are operated by means of computers and applications using communication networks to transmit information through wide and local area networks: Measurements are transmitted to the control centers; control data sent to the system's actuators; information exchange between control centers. These communication networks are also used for other purposes, such as office networking, software maintenance, and diagnostics.

Thanks to computers and multi-purpose networks, critical infrastructure systems are vulnerable to cyber attacks [1], [2], which are performed on the information residing and flowing in the IT system. Power networks, for instance, are operated through supervisory control and data acquisition (SCADA) systems. Several cyber attacks on SCADA systems operating power networks have been reported [3], and major blackouts are due to the misuse of the SCADA systems [4]. Power networks, being systems where control loops are closed over the communication network, represent an important class of Networked Control Systems (NCS). In order to increase the robustness of these systems, one needs appropriate tools to first understand and then to protect them against cyber attacks. Some of the literature concerning NCS has already tackled problems such as false data injection in state estimation [5], security constrained control [6], secure computations in networks [7], [8], [9] among others.



(a) Attack on a node (b) Attack on the out-going communications of a node

Figure 1. Two classes of attacks on NCS

This paper will tackle two different attack scenarios in a NCS, shown in Fig. 1: attack on the node dynamics, and attack on the out-going communications of a node. We propose methods to detect and isolate these events in a distributed fashion using a bank of Unknown Input Observers (UIOs) at each node. A similar approach was used to detect attacks on the nodes in the discrete-time consensus problem in [7]. Here we consider more general network models and attacks in continuous time. Sufficient conditions for the feasibility of the UIO will be given, based on the topological properties of the network. Furthermore, we present a method to reduce the number of monitoring nodes, thus reducing the overall complexity of the detection scheme. As examples we consider the continuous-time consensus algorithms, and the linearized swing equations. The swing equations [10] are used to model the active power flow in a network of synchronous generators and motors. Hence this model and the corresponding UIOs can be used to to detect attacks on the power grid. A similar approach can be found in [11].

The structure of the paper is as follows: we introduce the mathematical framework in Sec. II, followed by Sec. III where we formulate the considered problem and present a distributed detection scheme to tackle it. Sec. IV contains some remarks on the complexity of the proposed scheme and a possible method to reduce it. Then we present our results regarding two classes of NCS, the consensus protocol in Sec. V and

The authors are at KTH - Royal Institute of Technology, School of Electrical Engineering, Automatic Control, SE-100 44 Stockholm, Sweden.
{andretei,hsan,kallej}@kth.se

This work was supported in part by the European Commission through the VIKING project.

the power network in Sec. VI. In Sec. VII we summarize our results and present our conclusions.

II. Preliminaries

Before presenting the main results, we give a brief overview of the techniques used throughout this paper.

A. Networked Control Systems

1) Graph Theory: In multi-agent and networked systems, the interactions between agents are usually represented by graphs. We now review some of the basic concepts within graph theory [12].

An *undirected weighted* graph $\mathcal{G}$ is defined by a set of vertices or nodes $\mathcal{V}(\mathcal{G}) = \{1, \cdots, N\}$, a set of edges $\mathcal{E}(\mathcal{G}) \subseteq \mathcal{V}(\mathcal{G}) \times \mathcal{V}(\mathcal{G})$ and a weighted adjacency matrix $\mathcal{A}(\mathcal{G}) \in \mathbb{R}^{N \times N}$ with nonnegative elements. The graph argument will be omitted when the graph is clear within the context. For undirected graphs, two distinct nodes i and j are neighbors if $(i,j) \in \mathcal{E}$, in which case the corresponding entry in the adjacency matrix $[\mathcal{A}]_{ij}$ is positive and reflects the edge weight. A *path* between vertices i and j is a sequence of distinct vertices starting in i and ending in j, such that each consecutive vertices in the sequence are adjacent, were the length of the path is the number of edges it contains. We say the graph $\mathcal{G}$ is connected if there exists a path between any two distinct vertices. The neighbor set of node i is denoted as $\mathcal{N}_i = \{j \in \mathcal{V} \colon (i,j) \in \mathcal{E}\}$ and in this paper we also define the sets $\tilde{\mathcal{N}}_i = \mathcal{N}_i \bigcup i$ and $\bar{N}_i = \mathcal{V} \setminus \tilde{\mathcal{N}}_i$. The out-degree of node i is given by

$$\deg(i) = \sum_{j \in \mathcal{N}_i} [\mathcal{A}]_{ij}$$

and the degree matrix $\Delta(\mathcal{G}) \in \mathbb{R}^{N \times N}$ is a diagonal matrix defined as

$$[\Delta]_{ij} = \begin{cases} \deg(i) & , \ i = j \\ 0 & , \ i \neq j \end{cases}.$$

The Laplacian of $\mathcal{G}$, $\mathcal{L} \in \mathbb{R}^{N \times N}$, is a symmetric positive-semidefinite matrix defined by

$$\mathcal{L} = \Delta - \mathcal{A} \tag{1}$$

and it encodes several interesting properties of the graph, some of which are discussed in [12]. In particular, for connected graphs the Laplacian has a single zero eigenvalue, while for disconnected graphs it has c zero eigenvalues with c being the number of connected components.

A subgraph $\mathcal{G}'$ is said to be *induced* from $\mathcal{G}$ if it is obtained by removing a strict subset of vertices and their respective edges from $\mathcal{G}$.

2) System Dynamics: Throughout this paper we assume that each node has linear dynamics given by

$$\begin{cases} \dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t) \\ w_i(t) &= C_i x_i(t) \\ y_i(t) &= C_i x_i(t) \end{cases}, \tag{2}$$

with $x_i(t) \in \mathbb{R}^{n_i}$ as the node's state, $u_i(t) \in \mathbb{R}^{r_i}$ the control input, $w_i(t) \in \mathbb{R}^{p_i}$ the internal measurement and $y_i(t) \in \mathbb{R}^{p_i}$ the measurement node i transmits to the network. There is a conceptual difference between w_i and y_i: The internal measurement w_i is not shared through the network and so it is considered to be secure. On the other hand, y_i is the measurement transmitted through the communication network, thus being vulnerable to cyber attacks. Hence we envision that w_i may have a role to play in detecting the cyber attack. For the sake of simplicity, the time argument will be omitted and we will assume $n_i = n$, $r_i = r$ and $p_i = p$ for all $i \in \mathcal{V}$. The results presented throughout this paper hold for the general case as well.

By collecting the states of all the nodes into a single vector $\mathbf{x} = \left[x_1^T \cdots x_N^T\right]^T$, we can write the global dynamics of the network as

$$\begin{cases} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} \\ \mathbf{w} &= C\mathbf{x} \\ \mathbf{y} &= C\mathbf{x} \end{cases}, \tag{3}$$

where the matrices A, B and C are block-diagonal matrices defined in terms of $\{A_i\}$, $\{B_i\}$ and $\{C_i\}$, respectively.

Usually in NCS, the control input at each node is computed as a function of the local information available

$$u_i = \mu_i(\mathcal{I}_i). \tag{4}$$

In the classes of systems analyzed in this paper, $\mu_i(\cdot)$ is a linear function and $\mathcal{I}_i$ is the collection of the network's outputs available at node i defined as

$$\mathcal{I}_i = \mathbf{y}_i = \left[w_i^T \ \left\{y_j^T\right\}_{j \in \mathcal{N}_i}\right]^T = J_i \mathbf{x}. \tag{5}$$

Hence the global control input can be written as

$$\mathbf{u} = \mu(\mathbf{y}) = -KC\mathbf{x}, \tag{6}$$

where the structure of K encodes the topology of the network.

B. Unknown Input Observer

We will now present some of the techniques mentioned in [13] to design an observer for a linear time-invariant system affected by an unknown disturbance, described by the following state space equations:

$$\begin{cases} \dot{x} = Ax + Bu + Ed \\ y = Cx \end{cases}, \tag{7}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^r$ is the known input vector, $d \in \mathbb{R}^q$ is the unknown input vector and $y \in \mathbb{R}^p$ is the output vector.

An observer for the dynamical system in (7) is given by:

$$\begin{cases} \dot{z} = Fz + TBu + Ky \\ \hat{x} = z + Hy \end{cases}, \tag{8}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimated state and $z \in \mathbb{R}^n$ is the observer's state.

Definition 2.1 ([13]): A state observer is an unknown input observer (UIO) if the state estimation error e approaches zero asymptotically, regardless of the presence of an unknown input d.

Theorem 2.1 ([13]): Necessary and sufficient conditions for the observer described by (8) to be an UIO for the system in (7) are that the following conditions hold:

i) $\operatorname{rank}(CE) = \operatorname{rank}(E)$
ii) (C, A_1) is a detectable pair, where

$$A_1 = A - HCA.$$

If the conditions above are satisfied, the observer's matrices can be chosen so that the estimation error's dynamics become:

$$\dot{e}(t) = Fe(t), \tag{9}$$

with F being asymptotically stable so that $\lim_{t\to+\infty} e(t) = 0$, regardless of the value of the unknown signal $d(t)$.

III. Cyber Attacks on Networked Control Systems

Consider a NCS with an undirected and weighted graph $\mathcal{G}$ modeling the network and let each node have its dynamics described by (2). The network should have mechanisms to detect possible security breaches. Furthermore, the nature of these breaches should also be identified, thus allowing more effective corrective measures to be taken.

Our proposed scheme for detecting attacks is based on Fault Detection and Isolation (FDI) theory, using the Generalized Observer Scheme (GOS) [13].

In order to keep the detection scheme distributed, let each node have a monitoring system with a bank of UIO observers. From Thm 2.1 i) we see that we can only decouple disturbances which directly affect the measured states. Since each node receives measurements from its neighbors, we expect that it is able to locate disturbances within its neighborhood. Therefore, we propose a similar method as in [7], where each node monitors all its neighbors.

We will now analyze two different types of attacks that can be modeled in terms of unknown disturbances in the node's dynamics and we describe the proposed detection scheme. As an underlying assumption, we assume that there is at most one active attack in the network at any time instant.

A. Attack on a Node

This class of attacks covers scenarios where the normal behavior of a node is affected by an outsider and it no longer follows the distributed control law that governs the entire network. It includes, for instance, denial of service (DoS) or deception attacks on the in-going communications of a node, where the control input is compromised. In this scenario, the dynamics of the attacked node k can be written as

$$\left\{\begin{array}{rcl} \dot{x}_k &=& A_k x_k + B_k u_k + b_f^k f_k \\ w_k &=& C_k x_k \\ y_k &=& C_k x_k \end{array}\right., \tag{10}$$

where $b_f^k \in \mathbb{R}^n$ is the disturbance distribution vector and $f_k \in \mathbb{R}$ the disturbance signal.

Consider the detection scheme in node i. The global dynamics of the network when affected by attacks in all nodes, seen from node i, can be described using (3), (5) and (10)

$$\left\{\begin{array}{rcl} \dot{\mathbf{x}} &=& A\mathbf{x} + B\mathbf{u} + B_f\mathbf{f} \\ \mathbf{y}_i &=& J_i\mathbf{x} \end{array}\right., \tag{11}$$

where $\mathbf{f} = [f_1 \cdots f_N]^T$ and $B_f \in \mathbb{R}^{Nn\times N}$ is a block-diagonal matrix with full-column rank defined in terms of $\left\{b_f^i\right\}$. Note that node i does not know the entire control input vector $\mathbf{u}$ so this model cannot be incorporated in an observer as it is. Therefore, we consider the closed-loop dynamics

$$\left\{\begin{array}{rcl} \dot{\mathbf{x}} &=& \bar{A}\mathbf{x} + B_f\mathbf{f} \\ \mathbf{y}_i &=& J_i\mathbf{x} \end{array}\right., \tag{12}$$

with $\bar{A} = (A - BKC)$. The UIO insensitive to a disturbance in k has the following dynamics

$$\left\{\begin{array}{rcl} \dot{z}_i^k &=& F_i^k z_i^k + T_i^k B\mathbf{u} + K_i^k \mathbf{y}_i \\ \hat{\mathbf{x}}_i^k &=& z_i^k + H_i^k \mathbf{y}_i \end{array}\right. \tag{13}$$

with $\mathbf{u} = 0$. It exists if and only if the conditions in Thm 2.1 are satisfied, considering the system $\left(\bar{A}, 0, J_i\right)$ and $E = B_f^k$, the k^{th} column of B_f.

The detection scheme implemented in node i has one such observer for each of its neighbors and from each observer a residual signal $r_i^k = J_i e_i^k$ is obtained. Since B_f has full-column rank, it can be easily seen that r_i^k is insensitive only to f_k. Therefore, having in mind our assumption that there is at most one active attack, the following threshold logic can be applied:

- No fault is present: $\left\|r_i^k\right\| < T_{f_k}$, $\forall k \in \mathcal{N}_i$
- Neighbor node k has a fault:
$$\left\{\begin{array}{rcl} \left\|r_i^j\right\| &<& T_{f_j} \\ \left\|r_i^k\right\| &\geq& T_{f_k} \end{array}\right., \forall k \neq j \in \mathcal{N}_i$$
- There is a fault in a node $j \notin \tilde{\mathcal{N}}_i$: $\left\|r_i^k\right\| \geq T_{f_k}$, $\forall k \in \mathcal{N}_i$

B. Attack on the Out-Going Communications of a Node

Under this class of attacks, the out-going communication of a node is corrupted while the controls are correctly computed. It covers scenarios where a DoS or deception attack occurs on the broadcasted data of a node. Again it can be modeled in terms of the node's dynamics, corresponding to a disturbance on the information transmitted from node k:

$$\left\{\begin{array}{rcl} \dot{x}_k &=& A_k x_k + B_k u_k \\ w_k &=& C_k x_k \\ y_k &=& C_k x_k + f_{s_k} \end{array}\right., \tag{14}$$

with $f_{s_k} \in \mathbb{R}^p$ being the malicious information. The closed-loop dynamics of the network can be written as

$$\left\{\begin{array}{rcl} \dot{\mathbf{x}} &=& \bar{A}\mathbf{x} + B_f^k f_{s_k} \\ \mathbf{w} &=& C\mathbf{x} \\ \mathbf{y} &=& C\mathbf{x} + c_f^k f_{s_k} \end{array}\right., \tag{15}$$

where $B_f^k = B^{\bar{k}} K c_f^k$, $B^{\bar{k}}$ is obtained from B by replacing the diagonal block B_k by $0_{n\times r}$ and $c_f^k \in \mathbb{R}^{Np\times p}$ defined as

$$c_f^k = [0_{(k-1)p\times p}^T \; I_{p\times p} \; 0_{(k+1)p\times p}^T]^T. \tag{16}$$

Note that the internal measurement is not affected by the attack but now both the communicated measurements and the control inputs are. Partitioning $\bar{A}$ such that

$$\bar{A} = \begin{bmatrix} \bar{A}_{\bar{k}} & \bar{A}_{\bar{k}k} \\ \bar{A}_{k\bar{k}} & \bar{A}_k \end{bmatrix}, \tag{17}$$

the dynamics of the healthy network driven by y_k as an input can be written as

$$\begin{cases} \dot{\mathbf{x}}_{\bar{k}} &= \bar{A}_{\bar{k}}\mathbf{x}_{\bar{k}} + \bar{A}_{\bar{k}k}y_k \\ \mathbf{w}_{\bar{k}} &= C_{\bar{k}}\mathbf{x}_{\bar{k}} \\ \mathbf{y}_{\bar{k}} &= C_{\bar{k}}\mathbf{x}_{\bar{k}} \end{cases}. \tag{18}$$

As it can be seen from (18), the malicious data f_{s_k} may be chosen so that y_k follows any trajectory, including one due to the effect of f_k in (10). In this case, the healthy part of the network will see the same behavior from node k in both attack scenarios.

Theorem 3.1: Given the closed-loop system from (3) and (6), no healthy node i can distinguish between an attack on a node and an attack on all the out-going communication channels of the same node.

Proof: Let i be a healthy node observing the network and let k be the attacked node. Furthermore, assume that i is connected to all nodes in the network, which is equivalent to a centralized approach. The global dynamics in (12) can be rewritten as

$$\begin{cases} \dot{\mathbf{x}}_{\bar{k}} &= \bar{A}_{\bar{k}}\mathbf{x}_{\bar{k}} + \bar{A}_{\bar{k}k}y_k \\ \mathbf{y}_{\bar{k}} &= C_{\bar{k}}\mathbf{x}_{\bar{k}} \\ \dot{x}_k &= \bar{A}_k x_k + \bar{A}_{k\bar{k}}\mathbf{y}_{\bar{k}} + b_f^k f_k \\ y_k &= C_k x_k \end{cases}. \tag{19}$$

Node i will have measurements from all the nodes and so $\mathcal{I}_i = \begin{bmatrix}\mathbf{y}_{\bar{k}}^T & y_k^T\end{bmatrix}^T$. Now consider that all the out-going communications of node k are being compromised and denote $\tilde{y}_k$ as the corrupted measurement transmitted to the network. In this case we have $\tilde{\mathcal{I}}_i = \begin{bmatrix}\mathbf{y}_{\bar{k}}^T & \tilde{y}_k^T\end{bmatrix}^T$. The corrupted measurement can be computed such that it reproduces the effect of f_k, *i.e.* $\tilde{y}_k = y_k$. This way $\mathcal{I}_i = \tilde{\mathcal{I}}_i$ and so we conclude that node i cannot distinguish the nature of both attacks. These arguments are also valid for the general decentralized case. ■

Note that node k should be able to distinguish between an attack on itself and an attack on all its out-going communication channels, due to its own internal measurement w_k. In fact, the network's dynamics seen from node k are described by

$$\begin{cases} \dot{\mathbf{x}} &= \bar{A}\mathbf{x} + B_f^k f_{s_k} \\ \mathbf{y}_k &= J_k\mathbf{x} \end{cases}, \tag{20}$$

which has a similar structure to (12). A possible solution to this problem would then be to add one more observer to each node, which is insensitive only to a fault with distribution vector $B_f^k = B^{\bar{k}}Kc_f^k$, thus enabling each node to detect an attack on its own communications.

IV. On the Complexity of the Detection Scheme

The proposed scheme requires each node in the network to have a bank of observers to monitor each one of its neighbors, resulting in a distributed but computationally heavy FDI scheme. This poses a scalability problem as the order of the network's model increases linearly with the number of nodes.

However, since each node monitors all its neighbors, it is clear that there exists a certain amount of redundancy. Hence, one possible improvement to be made is to reduce the number of monitoring nodes.

Reducing the number of monitoring nodes

Assuming that each node monitors only its neighbors, we say that a FDI system in node i covers the set of nodes $\mathcal{N}_i$. Therefore, the objective is to select a minimum number of observer nodes so that they cover all the nodes in the network, *i.e.*,

$$\begin{aligned} \min_{S_o \subseteq \mathcal{V}} \quad & |S_o| \\ s.t. \quad & \bigcup_{i \in S_o} \mathcal{N}_i = \mathcal{V} \end{aligned}, \tag{21}$$

where S_o is the set of observer nodes.

As it can be seen, this is actually a set cover problem where we wish to determine a *minimum total dominating set*, *i.e.*, a set with minimum cardinality such that all nodes in the graph have at least one neighbor in that set. This is a well studied problem, having been classified as an NP-hard problem and we find two proposed algorithms in [14] that solve it.

Although the number of observers obtained by using $\mathcal{N}_i$ as the set of nodes covered by node i is not minimum, this method has one interesting property: all nodes in S_o are monitored by at least one neighbor. This means that even if an observer node is attacked, there is another observer node in the network that can detect it. Obviously, this decreases the vulnerability of such scheme to faults in the monitoring nodes.

Other interesting properties may also be imposed by modifying the constraints in (21), such as having S_o to be connected, which is related to the *minimum connected dominating set* problem.

V. The Consensus Protocol

Consider a group of N dynamic agents with single-integrator dynamics described by (2) with $x_i, u_i, y_i, w_i \in \mathbb{R}$, $A_i = 0$ and $B_i = C_i = 1$. The agents are connected through a communication network, represented by a graph $\mathcal{G}$, and use the following distributed control law [15]

$$u_i = -\sum_{j \in \mathcal{N}_i} (w_i - y_j). \tag{22}$$

Under this setting, the dynamics of the entire network are given by

$$\dot{\mathbf{x}} = -\mathcal{L}\mathbf{x}. \tag{23}$$

A. Consensus Protocol subject to Attacks on Nodes

The behavior of the attacked node k can be described by rewriting (10) as

$$\left\{\begin{array}{lcl}\dot{x}_k &=& -\sum_{j\in\mathcal{N}_k}(w_k-y_j)+f_k\\ w_k &=& x_k\\ y_k &=& x_k\end{array}\right., \tag{24}$$

and so the global dynamics of the network can be written as

$$\left\{\begin{array}{lcl}\dot{\mathbf{x}} &=& -\mathcal{L}\mathbf{x}+b_f^k f_k\\ \mathbf{y} &=& \mathbf{x}\\ \mathbf{w} &=& \mathbf{x}\end{array}\right., \tag{25}$$

with $b_f^k \in \mathbb{R}^N$ being a vector with the k^{th} component set to 1 and all the others to 0.

B. Consensus Protocol subject to Communication Attacks

The compromised node k having its out-going communications tampered, as in Fig. 1(b), has its dynamics described by:

$$\left\{\begin{array}{lcl}\dot{x}_k &=& -\sum_{j\in\mathcal{N}_k}(w_k-y_j)\\ w_k &=& x_k\\ y_k &=& x_k+f_{s_k}\end{array}\right.. \tag{26}$$

Hence the global dynamics of the network can be written as:

$$\left\{\begin{array}{lcl}\dot{\mathbf{x}} &=& -\mathcal{L}\mathbf{x}+I_{\bar{k}}l^k f_{s_k}\\ \mathbf{y} &=& \mathbf{x}+b_f^k f_{s_k}\\ \mathbf{w} &=& \mathbf{x}\end{array}\right., \tag{27}$$

where $I_{\bar{k}} \in \mathbb{R}^{N\times N}$ is the identity matrix with the k^{th} diagonal entry set to zero, $l^k \in \mathbb{R}^N$ is the k^{th} column of the Laplacian matrix.

C. Detecting the Attacks

We now apply the proposed detection scheme to the consensus protocol, providing sufficient conditions for the existence of the UIOs.

1) Attack on a node: Consider the FDI scheme implemented in node i. The information available at node i is given by (5) and the network's dynamics are described by (25). The UIO insensitive to such fault has dynamics given by (13). The existence conditions for such UIO can be validated using topological properties of the network, according to the following theorem:

Theorem 5.1: There exists a UIO for the system $\left(-\mathcal{L}(\mathcal{G}), b_f^k, J_i\right)$ if the graph $\mathcal{G}$ is connected and $k \in \mathcal{N}_i$.

A similar result was presented in [7] for discrete-time consensus, using a different approach. Before proving the Thm 5.1, we introduce the following lemma:

Lemma 5.2: If an undirected graph $\mathcal{G}$ is connected, then any partition of its Laplacian matrix $\mathcal{L}$, induced by a strict subset of nodes $\bar{F} \subset \mathcal{V}$, is invertible.

Proof of Thm 5.1: It can be easily seen that if the faulty node k is a neighbor of the observer node i, then the first condition of Thm 2.1 is satisfied: since the row corresponding to node k is $J_i^k = {b_f^k}^T$, then it follows that $\operatorname{rank}\left(J_i b_f^k\right) = \operatorname{rank}\left({b_f^k}^T b_f^k\right) = \operatorname{rank}\left(b_f^k\right) = 1$.

As for the second condition in Thm 2.1, this condition is equivalent to say that the transmission zeros of the system $\left(-\mathcal{L}, b_f^k, J_i, 0\right)$ must be stable [13], *i.e.*

$$\begin{bmatrix} sI_N+\mathcal{L} & b_f^k \\ J_i & 0\end{bmatrix},$$

is of full column rank for all s such that $\Re(s) \geq 0$, $I_N \in \mathbb{R}^{N\times N}$ being the identity matrix.

Suppose now that we apply a transformation P to the system $\left(-\mathcal{L}, b_f^k, J_i\right)$ so that $\bar{\mathbf{x}} = P\mathbf{x} = \begin{bmatrix}\mathbf{x}_{\tilde{\mathcal{N}}_i}^T & \mathbf{x}_{\bar{\mathcal{N}}_i}^T\end{bmatrix}^T$ and $\bar{J}_i = \begin{bmatrix} I_{|\tilde{\mathcal{N}}_i|} & 0_{|\tilde{\mathcal{N}}_i|\times|\bar{\mathcal{N}}_i|}\end{bmatrix}$, which consists on a simple permutation operation. After this operation we can write the Laplacian as

$$\bar{\mathcal{L}} = P^{-1}\mathcal{L}P = \begin{bmatrix}\mathcal{L}_{\tilde{\mathcal{N}}_i} & l_{\tilde{\mathcal{N}}_i\bar{\mathcal{N}}_i} \\ l_{\bar{\mathcal{N}}_i\tilde{\mathcal{N}}_i} & \mathcal{L}_{\bar{\mathcal{N}}_i}\end{bmatrix},$$

and hence we have:

$$\begin{bmatrix} sI_N+\mathcal{L} & b_f^k \\ J_i & 0\end{bmatrix} = \begin{bmatrix} sI_{|\tilde{\mathcal{N}}_i|}+\mathcal{L}_{\tilde{\mathcal{N}}_i} & l_{\tilde{\mathcal{N}}_i\bar{\mathcal{N}}_i} & \bar{b}_f^k \\ l_{\bar{\mathcal{N}}_i\tilde{\mathcal{N}}_i} & sI_{|\bar{\mathcal{N}}_i|}+\mathcal{L}_{\bar{\mathcal{N}}_i} & 0_{\bar{\mathcal{N}}_i\times 1} \\ I_{|\tilde{\mathcal{N}}_i|} & 0_{|\bar{\mathcal{N}}_i|\times|\bar{\mathcal{N}}_i|} & 0_{\tilde{\mathcal{N}}_i\times 1}\end{bmatrix},$$

with $\bar{b}_f^k = P^{-1}b_f^k$.

Note that due to the last row of the previous matrix, the first column is independent of the others and furthermore it is of full column rank, thus:

$$\operatorname{rank}\begin{bmatrix} sI_N+\mathcal{L} & b_f^k \\ J_i & 0\end{bmatrix} = |\tilde{\mathcal{N}}_i| + \operatorname{rank}\begin{bmatrix} l_{\tilde{\mathcal{N}}_i\bar{\mathcal{N}}_i} & \bar{b}_f^k \\ sI_{|\bar{\mathcal{N}}_i|}+\mathcal{L}_{\bar{\mathcal{N}}_i} & 0_{|\bar{\mathcal{N}}_i|\times 1}\end{bmatrix}.$$

From Lemma 5.2 we know that any square partition of the Laplacian is invertible if the respective graph is connected, thus $\mathcal{L}_{\bar{\mathcal{N}}_i} \succ 0$ and since $\Re(s) \geq 0$, $sI_{|\bar{\mathcal{N}}_i|}+\mathcal{L}_{\bar{\mathcal{N}}_i}$ is invertible as well and has full rank, following that:

$$\operatorname{rank}\begin{bmatrix} sI_N+\mathcal{L} & b_f^k \\ J_i & 0\end{bmatrix} = |\tilde{\mathcal{N}}_i| + |\bar{\mathcal{N}}_i| + 1 = N+1,$$

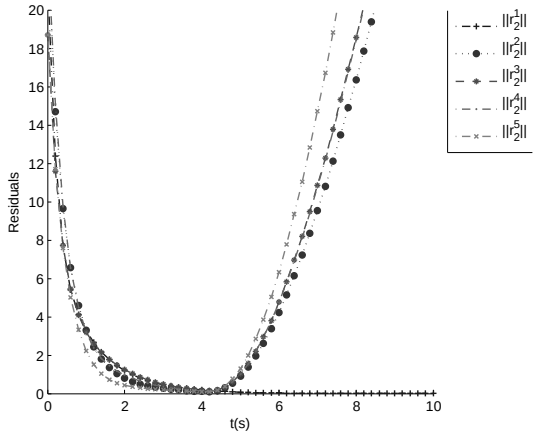
which proves that the transmission zeros are all stable and that a UIO exists. ■

2) Attack on the out-going communications of a node: In this scenario the previous FDI scheme would detect and locate an attack in the compromised agent, but would not identify its nature, given the information available to the healthy part of the network. As discussed before, one solution would be to add to the scheme in each node k one more UIO insensitive to $b_f^k = I_{\bar{k}}l^k$.

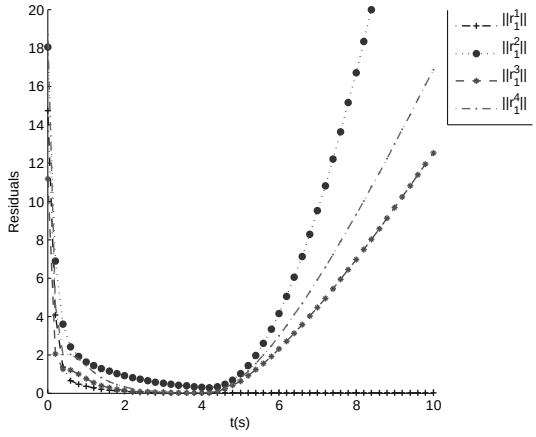
Theorem 5.3: There exists a UIO for the system $\left(-\mathcal{L}(\mathcal{G}), I_{\bar{k}}l^k, J_k\right)$ if the graph $\mathcal{G}$ is connected.

Proof: First note that $b_f^k = I_{\bar{k}}l^k$ is actually the k^{th} column of the Laplacian matrix with the k^{th} entry set to zero. Therefore, if $\mathcal{G}$ is connected, node k has at least one neighbor and $\operatorname{rank}\left(J_k b_f^k\right) = \operatorname{rank}\left(b_f^k\right) = 1$. The rest of the proof uses the same arguments as in the proof of Thm 5.1. ■

3) Simulation results: As an example, we compare the residuals of two different nodes in the graph presented in Fig. 1(b). The network is suffering a deception attack on the out-going communications of node 1. As it can be seen from Fig. 2, node 2 identifies node 1 as being the compromised node using the threshold logic presented in Sec. III-A, since the residual corresponding to node 1 is close to zero while all others are large. However, it cannot identify the nature of the attack. On the other hand, node 1 successfully detects the deception attack on its own out-going communications, as expected.



(a) Residuals at node 2



(b) Residuals at node 1

Figure 2. Deception attack in node 1 at $t = 4s$

VI. Power Network

Power systems are an example of very complex systems in which several elements, such as generators and loads, are dynamically interconnected. They can be seen as a networked system, where each bus is a node. We provide a simple model for the active power flow in a power grid.

A. Modeling the Power Network

The behavior of a bus i can be described by the so-called *swing equation*:

$$M_i\ddot{\delta}_i + D_i\dot{\delta}_i - P_{mi} = -\sum_{j\in\mathcal{N}_i} P_{ij}, \tag{28}$$

where M_i and D_i are the inertia and damping coefficients, respectively, P_{mi} is the mechanical input power and P_{ij} is the active power flow from node i to j. Considering that there are no power losses and letting $V_i = |V_i|\, e^{j\delta_i}$ and δ_i be, respectively, the complex voltage and the phase angle of bus i, the active power flow between bus i and bus j, P_{ij}, is given by:

$$P_{ij} = k_{ij}\sin\left(\delta_i - \delta_j\right), \tag{29}$$

where $k_{ij} = |V_i|\,|V_j|\,b_{ij}$ and b_{ij} is the susceptance of the power line connecting buses i and j.

Since the phase angles usually are close, we can linearize (29), rewriting the dynamics of bus i as:

$$M_i\ddot{\delta}_i + D_i\dot{\delta}_i = u_i, \tag{30}$$

with

$$u_i = -\sum_{j\in\mathcal{N}_i} k_{ij}\left(\delta_i - \delta_j\right) + P_{mi}. \tag{31}$$

Rewriting (30) and (31) in state-space form as (2) and considering $\boldsymbol{\delta} = \begin{bmatrix}\delta_1^T & \cdots & \delta_N^T\end{bmatrix}^T = C_\delta\mathbf{x}$, we can write the network dynamics as

$$\left\{\begin{array}{rcl}\dot{\mathbf{x}} &=& \bar{A}\mathbf{x} + B\mathbf{P_m}\\ \mathbf{y} &=& C\mathbf{x}\\ \mathbf{w} &=& C\mathbf{x}\end{array}\right., \tag{32}$$

where $\bar{A} = A - B\mathcal{L}C_\delta$ and $\mathbf{P_m} = \begin{bmatrix}P_{m1} & \cdots & P_{mN}\end{bmatrix}^T$ is the collection of input power at each bus. These inputs are the generator's setpoints or load power consumptions, which we assume to be known.

B. Power Network subject to Attacks on Buses

Let k be the index of the bus subject to an attack, which we model as a disturbance. The dynamics of the power system under the effect of such fault are described by:

$$\left\{\begin{array}{rcl}\dot{\mathbf{x}} &=& \bar{A}\mathbf{x} + B\mathbf{P_m} + b_f^k f_k\\ \mathbf{y}_i &=& J_i\mathbf{x},\end{array}\right. \tag{33}$$

where b_f^k is the k^{th} column of matrix B. Note that such column has only one non-zero entry, which corresponds to the frequency state of bus k.

C. Power Network subject to Communication Attacks

Let the dynamics of the compromised bus k be described by (14). The closed-loop dynamics of the power network can be written as

$$\left\{\begin{array}{rcl}\dot{\mathbf{x}} &=& \bar{A}\mathbf{x} + B\mathbf{P_m}\\ \mathbf{w} &=& C\mathbf{x}\\ \mathbf{y} &=& C\mathbf{x} + c_f^k f_{s_k}\end{array}\right. . \tag{34}$$

Since the source of interactions between agents is physical rather than induced by communications, a deception attack on a communicated measurement does not cause any effect on the dynamics of the network.

D. Detecting the Attacks

1) Attack on a bus: Let the set of measurements available at bus i be $\mathbf{y}_i = J_i\mathbf{x}$, where $J_i \in \mathbb{R}^{p\times n|\bar{\mathcal{N}}_i|}$. Since we desire an UIO decoupled from b_f^k, we design an UIO insensitive to $E_k = b_f^k$ using the same structure as in (8).

The conditions for such observer to exist are given in Thm 2.1 and we give some comments regarding them. Cond. i) is related only to the available measurements and unknown inputs. From the state-space model, we have that b_f^k has non-zero elements only on the entries corresponding to the frequency offset $\dot{\delta}_k$. Therefore, the measurement set $\mathbf{y}_i$ must include the frequency measurements from bus k in order to satisfy the first condition.

On the other hand, Cond. ii) depends on the structural properties of the system in (33), as it is equivalent to say that the matrix

$$\begin{bmatrix} sI_{nN} - \bar{A} & E_k \\ J_i & 0 \end{bmatrix}$$

is of full column rank for all s such that $\Re(s) \geq 0$. Given the complexity of this kind of interconnected system, such condition is harder to evaluate analytically on this particular system than on the consensus problem.

There are, however, some comments to be made on this matter. First notice that one necessary condition is that the system $\left(J_i, \bar{A}\right)$ is detectable, since otherwise the matrix

$$\begin{bmatrix} sI_{nN} - \bar{A} \\ J_i \end{bmatrix}$$

will not have full column rank for some s such that $\Re(s) \geq 0$, which is a standard test for detectability.

2) Attack on the out-going communications: From (34) we conclude that the deception attack only affects the measurement and in this case detecting the attack can be done by using observer-based sensor fault detection methods [13].

3) Simulation results: These results were obtained from the IEEE 9-bus benchmark [16], in which we consider bus 7 observing its neighborhood, buses 5, 6 and 8. Bus 6 was attacked by $f_6 = 10\sin(t - t_f)\,,\ t \geq t_f$, with $t_f = 6s$ being the time instant at which the attack occurred. As it can be seen, bus 7 successfully detected and located the attack at bus 6, as its corresponding residual is the only converging to zero.

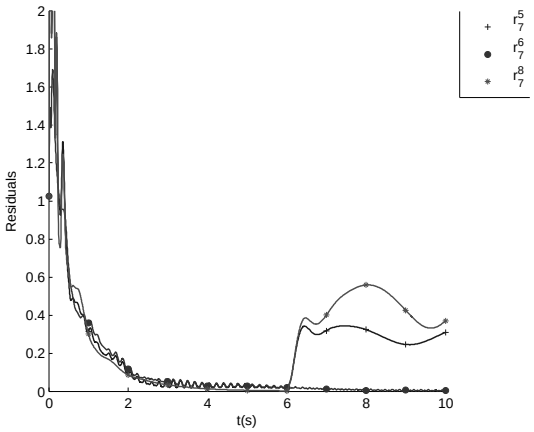


Figure 3. Attack in bus 6 at $t = 6s$ detected by bus 7

VII. Conclusion

We proposed a distributed UIO-based scheme to detect, locate and identify the nature of cyber attacks in NCS. Furthermore, we showed that, when the control loop is closed over a communication network, only the compromised node itself can distinguish the nature of the attack. A possible solution to reduce the complexity of such scheme was also given. We applied the proposed scheme to two classes of NCS and gave sufficient conditions for its feasibility in the first class and necessary conditions for the second class. Simulation results on both systems were also provided.

References

[1] A. Giani, S. Sastry, K. H. Johansson, and H. Sandberg, "The VIKING project: an initiative on resilient control of power networks," in *IEEE Int. Symp. on Resilient Control Systems*, 2009, To appear.

[2] A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *Proceedings of the 3rd USENIX Workshop on Hot topics in security*. USENIX, July 2008, p. Article 6.

[3] "Electricity grid in U.S. penetrated by spies," *The Wall Street Journal*, p. A1, April 8 2009.

[4] "Final report on the August 14th blackout in the United States and Canada," U.S.-Canada Power System Outage Task Force, Tech. Rep., April 2004.

[5] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM conference on Computer and communications security*, Accepted.

[6] S. Amin, A. Cárdenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks." in *HSCC*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds., vol. 5469. Springer, 2009, pp. 31–45.

[7] F. Pasqualetti, A. Bicchi, and F. Bullo, "Distributed intrusion detection for secure consensus computations," in *Proceedings of the IEEE Conf. on Decision and Control*, New Orleans, LA, Dec. 2007, pp. 5594–5599.

[8] S. Sundaram and C. Hadjicostis, "Distributed function calculation via linear iterations in the presence of malicious agents - part i: Attacking the network," in *Proceedings of the American Control Conference*, June 2008, pp. 1350–1355.

[9] ——, "Distributed function calculation via linear iterations in the presence of malicious agents - part ii: Overcoming malicious behavior," in *Proceedings of the American Control Conference*, June 2008, pp. 1356–1361.

[10] P. Kundur, *Power System Stability and Control*. McGraw-Hill Professional, 1994.

[11] E. Scholtz and B. Lesieutre, "Graphical observer design suitable for large-scale DAE power systems," in *Proceedings of the IEEE Conf. on Decision and Control*, Cancun, Dec. 2008, pp. 2955–2960.

[12] C. Godsil and G. Royle, *Algebraic Graph Theory*, 1st ed., ser. Graduate Texts in Mathematics. Springer, 2001.

[13] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers, 1999.

[14] F. Grandoni, "A note on the complexity of minimum dominating set," *J. Discrete Algorithms*, vol. 4, no. 2, pp. 209–214, July 2006.

[15] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," in *Proceedings of the IEEE*, vol. 95, no. 1, Jan. 2007, pp. 215–233.

[16] "Power system test cases," http://psdyn.ece.wisc.edu/IEEE_benchmarks/.