J. Mod. Power Syst. Clean Energy (2018) 6(5):872–886
https://doi.org/10.1007/s40565-018-0452-y

Set-theoretic detection of data corruption attacks on cyber physical power systems

Efstathios KONTOURAS[1], Anthony TZES[2], Leonidas DRITSAS[3]

Abstract This paper addresses a set-theoretic method for the detection of data corruption cyber-attacks on the load frequency control loop of a networked power system. The system consists of several interconnected control areas forming a power grid. Based on the overall discrete-time network dynamics, a convex and compact polyhedral robust invariant set is extracted and is used as a set-induced anomaly detector. If the state vector exits the invariant set, then an alarm will be activated, and the potential threat is considered disclosed. The attack scenario used to assess the efficiency of the proposed anomaly detector concerns corrupted frequency sensor measurements transmitted to the automatic generation control unit of a compromised control area. Simulation studies highlight the ability of a set-theoretic approach to disclose persistent and intermittent attack patterns even when they occur at the same time with changes in the power load demand.

Keywords Power systems, Load frequency control, Cyber-attacks, Set-theoretic methods

1 Introduction

Modern power grids are presently integrated with an extended digital layer comprised of sensors and smart meters that provide measurements at a fast rate and a high resolution [1]. Several smart devices are also able to transmit measurements via wireless communication channels that although flexible and efficient are generally unprotected and vulnerable to cyber-attacks [2]. Cyber-attackers compromise the integrity of sensitive elements of the network aiming to cause system malfunctions [3]. Common attack scenarios include the delaying or jamming of the acquired sensor data [4] and the corruption of the measurements by injecting false signals [5]. The reliable transmission of the data requires the use of security-enhancing techniques that increase the complexity of the infrastructure, leading to a cyber physical system modeling approach [6].

Attacks on the load frequency control loop of power networks have been studied in [7, 8], whereas in [9] the concept of positive invariance was used to quantify the attack impact on a two-area power plant. The design of stealthy adversaries was addressed in [10] and attack detectors in the form of network monitors were proposed in [11, 12]. Residual-based state estimators are by far the most common way of detecting attacks on networked systems as it is shown in [13, 14]. These detectors rely on the value of the estimation residue in order to decide

CrossCheck date: 31 July 2018

Received: 28 November 2017 / Accepted: 31 July 2018 / Published online: 18 September 2018

✉ Anthony TZES
anthony.tzes@nyu.edu

Efstathios KONTOURAS
kontouras@ece.upatras.gr

Leonidas DRITSAS
dritsas@aspete.gr

[1] Electrical and Computer Engineering Department, University of Patras, 26500 Rio, Greece

[2] Electrical and Computer Engineering Program, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, United Arab Emirates

[3] Department of Electrical and Electronic Engineering Educators, School of Pedagogical and Technological Education, ASPETE, 14121 Athens, Greece

whether or not the system is under attack. If the residue obtains a steady-state value larger than a critical threshold, then an alarm is activated. The motivation of this work emerges from the fact that if the attack pattern forces the state variables to oscillate, then they will never obtain a constant steady-state value and therefore the residual-based estimator will never cause the alarm to be triggered.

The use of a robust invariant set in order to develop a set-theoretic attack detector was first introduced in [15], where the load frequency control loop of a single control area was studied. The key idea was to trigger an alarm whenever the state vector exits the invariant set. It was shown that small persistent bias injected signals corrupting the frequency sensor measurements can pass undetected. Explicit boundary values of the bias injected signal that ensure a stealthy attack were derived in [16]. However, these bounds are consistently small and therefore impractical from an adversarial point of view, since realistic attack scenarios involve larger values of the attack signal.

This paper elaborates on previous results of the authors [17, 18]. In this work, the use of set-theoretic attack detectors is expanded on a networked power system and their efficiency is assessed considering both persistent and intermittent data corruption attack patterns on the frequency measurements. The simulations concern a case study of the benchmark two area power plant and they highlight the ability of the set-theoretic detector to disclose attacks during the transient response of the system, while also in the presence of disturbances; a feat that the traditional residual-based estimators are unable to demonstrate.

The paper is organized as follows. In Section 2 the mathematical model of the power network is established, while in Section 3 the design of the set-theoretic attack detector is presented. In Section 4 the switching signal driving the intermittent attack pattern is developed and Section 5 presents simulation results validating our conceptual approach. In Section 6 we provide concluding remarks, whereas an "Appendix A" explains some of the notations used throughout the paper.

2 System description

The algorithms used for the computation of a robust invariant set require a discrete-time representation of the system dynamics. The discretization process of the power network is performed in three steps. First, we extract the discrete-time equivalent models of all interconnected areas assuming that the power exchanged via the tie lines is an external signal. Then, we discretize the tie line model of each control area separately and finally, we combine the control area and tie line dynamics into a single discrete-time state space model.

2.1 Interconnected control area model

Consider the generic interconnected control area model, subject to a data corruption cyber-attack depicted in Fig. 1. According to [19, 20], a state space model that describes the evolution of the plant in the continuous-time domain is given as:

$$S_i^c : \begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}_i(t) &= \boldsymbol{A}_{c,i}\boldsymbol{x}_i(t) + \boldsymbol{B}_{c,i}u_{c,i}(t) + \boldsymbol{D}_{c,i}\Delta P_{L,i}(t) \\ & \quad + \boldsymbol{E}_{c,i}\Delta P_{tie,i}(t) \\ \boldsymbol{x}_i(0) &= \boldsymbol{x}_{i,0} \\ y_i(t) &= \boldsymbol{C}_i\boldsymbol{x}_i(t) \end{cases}$$

where the subscript $i \in \mathcal{I} = \{1, 2, \ldots, N\}$ denotes the i-th control area of the network and $t \in \mathbb{R}_+$ is the time variable. The state vector $\boldsymbol{x}_i(t) \in \mathbb{R}^2$ encapsulates the deviation of the electrical frequency $\Delta f_i(t)$ and the deviation of the mechanical power in the output of the turbine $\Delta P_{G,i}(t)$, namely:

$$\boldsymbol{x}_i(t) = \begin{bmatrix} \Delta f_i(t) & \Delta P_{G,i}(t) \end{bmatrix}^{\mathrm{T}}$$

We assume that the mechanical power provided to the rotor shaft is equal to the electrical power produced by the generator. The system output $y_i(t) \in \mathbb{R}$ is identical to the first state variable $\Delta f_i(t)$, thus $\boldsymbol{C}_i = [1 \quad 0]$. The control input $u_{c,i}(t) \in \mathbb{R}$ consists of two components, namely the primary frequency control action $\Delta P_{f,i}(t)$ and the automatic generation control law $\Delta P_{c,i}(t)$. According to Fig. 1, the input $u_{c,i}(t)$ is defined as:

$$u_{c,i}(t) = \Delta P_{c,i}(t) + \Delta P_{f,i}(t)$$

and is subject to the saturation hard constraint:

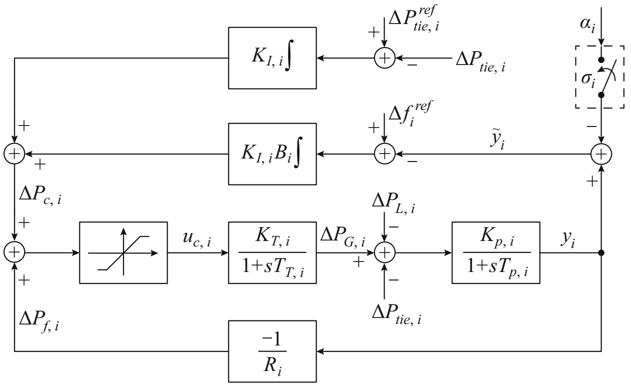


Fig. 1 Load frequency control loop of a generic interconnected control area subject to a data corruption cyber-attack on the frequency sensor measurements (the speed governor dynamics are omitted for brevity)

$$|u_{c,i}(t)| \leq u_{i,\max} \quad \forall t \geq 0$$

The signal $\Delta P_{L,i}(t) \in \mathbb{R}$ is an unknown but bounded disturbance representing the power load deviation due to the demand of the consumers. In contrast to other works [8, 9], we allow the load to change over time according to the power demand, that is $\Delta P_{L,i}(t) \neq 0$, and we assume that it obeys the constraint:

$$|\Delta P_{L,i}(t)| \leq \Delta P_{L,i,\max} \quad \forall t \geq 0$$

The signal $\Delta P_{tie,i}(t) \in \mathbb{R}$ represents the deviation of the electrical power exchanged between the i-th control area and the network through the tie line interconnection, whenever a power load change occurs. The matrix $\boldsymbol{A}_{c,i} \in \mathbb{R}^{2\times 2}$ is defined as:

$$\boldsymbol{A}_{c,i} = \begin{bmatrix} -1/T_{p,i} & K_{p,i}/T_{p,i} \\ 0 & -1/T_{T,i} \end{bmatrix}$$

and the matrices $\boldsymbol{B}_{c,i}, \boldsymbol{D}_{c,i}, \boldsymbol{E}_{c,i} \in \mathbb{R}^{2\times 1}$ are defined as:

$$\boldsymbol{B}_{c,i} = \begin{bmatrix} 0 \\ K_{T,i}/T_{T,i} \end{bmatrix}$$

$$\boldsymbol{D}_{c,i} = \boldsymbol{E}_{c,i} = \begin{bmatrix} -K_{p,i}/T_{p,i} \\ 0 \end{bmatrix}$$

The load changes affecting the control areas cause the electrical frequency and the tie line power to deviate from their nominal values. The speed governor performs the primary frequency control action defined as:

$$\Delta P_{f,i}(t) = -\frac{1}{R_i} y_i(t) \tag{1}$$

where R_i is the speed droop parameter. The remaining steady-state errors are eliminated by the automatic generation control unit, which is usually implemented in terms of an integral controller [21] defined as:

$$\Delta P_{c,i}(t) = K_{I,i} \int_0^t ACE_i(\tau)\mathrm{d}\tau \tag{2}$$

$$ACE_i(t) = \left(\Delta P_{tie,i}^{ref} - \Delta P_{tie,i}(t)\right) + B_i\left(\Delta f_i^{ref} - \tilde{y}_i(t)\right)$$

where $ACE_i(t)$ represents the i-th area control error, the reference signals $\Delta P_{tie,i}^{ref} = 0$ and $\Delta f_i^{ref} = 0$ associate with the tie line power deviation and the electrical frequency deviation respectively and $B_i = 1/R_i$. The signal $\tilde{y}_i(t)$ is defined as:

$$\tilde{y}_i(t) = y_i(t) - \alpha_i \sigma_i(t) \tag{3}$$

where $\alpha_i \in \mathbb{R}$ denotes the attack signal corrupting the measurement channel and $\sigma_i : \mathbb{R}_+ \rightarrow \{0, 1\}$ determines whether or not the i-th area is under attack. The speed governor remains unaffected since it is either mechanically or hydraulically coupled with the generator.

The equivalent discrete-time model of each interconnected control area is extracted by first computing the eigenvalues of the matrices $\boldsymbol{A}_{c,i}$ and then selecting a global sampling frequency f_s at least ten times greater than the frequency of the fastest eigenvalue of the network. We apply the zero-order hold method and obtain a discrete-time state space representation as

$$S_i^d : \begin{cases} \boldsymbol{x}_i[k+1] & = & \boldsymbol{A}_{d,i}\boldsymbol{x}_i[k] + \boldsymbol{B}_{d,i}u_{d,i}[k] + \boldsymbol{D}_{d,i}\Delta P_{L,i}[k] \\ & & + \boldsymbol{E}_{d,i}\Delta P_{tie,i}[k] \\ \boldsymbol{x}_i[0] & = & \boldsymbol{x}_{i,0} \\ y_i[k] & = & \boldsymbol{C}_i\boldsymbol{x}_i[k] \end{cases}$$

where $k \in \mathbb{N}$ is the new time variable.

Finally, since $u_{c,i}(t)$ implements a dynamic control law, we need to determine the equivalent discrete-time controller $u_{d,i}[k]$. Let us set the accumulated time errors

$$z_{1,i}(t) = \frac{1}{f^\circ} \int_0^t \left(\Delta f_i^{ref} - \tilde{y}_i(\tau)\right)\mathrm{d}\tau \tag{4}$$

$$z_{2,i}(t) = \frac{1}{|P_{tie,i}^\circ|} \int_0^t \left(\Delta P_{tie,i}^{ref} - \Delta P_{tie,i}(\tau)\right)\mathrm{d}\tau \tag{5}$$

as the extra state variables augmenting the system due to the existence of the integrator (2). Parameters f° and $P_{tie,i}^\circ$ represent the nominal network frequency and the nominal power exchanged via the i-th tie line respectively. We remark that $P_{tie,i}^\circ$ is considered positive when the power flow is directed from the i-th area towards the network. If we consider (1)-(5), while keeping the sampling frequency and the discretization method unaltered, then we obtain:

$$u_{d,i}[k] = \bar{K}_{I_1,i}z_{1,i}[k] + \bar{K}_{I_2,i}z_{2,i}[k] - \frac{1}{R_i}y_i[k]$$

where the gains $\bar{K}_{I_1,i}$ and $\bar{K}_{I_2,i}$ are defined as:

$$\begin{cases} \bar{K}_{I_1,i} = K_{I,i}B_i f^\circ \\ \bar{K}_{I_2,i} = K_{I,i}|P_{tie,i}^\circ| \end{cases}$$

and the variables $z_{1,i}[k]$, $z_{2,i}[k]$ satisfy the equations:

$$\begin{cases} z_{1,i}[k+1] = z_{1,i}[k] - \dfrac{1}{f_s f^\circ}\tilde{y}_i[k] \\ z_{1,i}[0] = z_{2,i}[0] = 0 \end{cases} \tag{6}$$

$$z_{2,i}[k+1] = z_{2,i}[k] - \frac{1}{f_s|P_{tie,i}^\circ|}\Delta P_{tie,i}[k] \tag{7}$$

The model of the discrete-time closed-loop interconnected control area under attack is written as:

$$S_i^{cl}:\begin{cases}\xi_i[k+1] & = & A_{cl,i}\xi_i[k]+\alpha_i B_{cl,i}\sigma_i[k]\\ & & +D_{cl,i}\Delta P_{L,i}[k]+E_{cl,i}\Delta P_{tie,i}[k]\\ y_i[k] & = & C_{cl,i}\xi_i[k]\end{cases} \tag{8}$$

where the augmented state vector $\xi_i[k]\in\mathbb{R}^4$ is given as:

$$\xi_i[k]=\begin{bmatrix}\Delta f_i[k] & \Delta P_{G,i}[k] & z_{1,i}[k] & z_{2,i}[k]\end{bmatrix}^{\mathrm{T}}$$

the matrix $A_{cl,i}\in\mathbb{R}^{4\times4}$ is given as:

$$A_{cl,i}=\begin{bmatrix}A_{d,i}-(1/R_i)B_{d,i}C_i & B_{d,i}\bar{K}_{I_1,i} & B_{d,i}\bar{K}_{I_2,i}\\ -1/(f_s f^{\circ})C_i & 1 & 0\\ \mathbb{O}_{1\times2} & 0 & 1\end{bmatrix}$$

the matrices $B_{cl,i},D_{cl,i},E_{cl,i}\in\mathbb{R}^{4\times1}$ are given as:

$$\begin{cases}B_{cl,i}=[\mathbb{O}_{1\times2}\quad 1/(f_s f^{\circ})\quad 0]^{\mathrm{T}}\\ D_{cl,i}=\left[D_{d,i}^{\mathrm{T}}\quad 0\quad 0\right]^{\mathrm{T}}\\ E_{cl,i}=\left[E_{d,i}^{\mathrm{T}}\quad 0\quad -1/\left(f_s|P_{tie,i}^{\circ}|\right)\right]^{\mathrm{T}}\end{cases}$$

and $C_{cl,i}=[C_i\quad 0\quad 0]$.

2.2 Tie line model

Each control area that is connected to the network is able to exchange power with it through a tie line. Whenever a load change occurs, the power flow of each tie line deviates from its nominal value $P_{tie,i}^{\circ}$ according to $\Delta P_{tie,i}(t)$. The linearized tie line dynamics associated with the i-th area are governed by the equation [19, 20]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta P_{tie,i}(t)=\sum_{j=1}^{N}\left[2\pi T_{ij}\left(\Delta f_i(t)-\Delta f_j(t)\right)\right] \tag{9}$$

where T_{ij} denotes the synchronization coefficient between the control areas i and j and $\Delta P_{tie,i}$ encapsulates all existing interconnections of the i-th area with the other areas of the grid.

The synchronization coefficients satisfy the condition $T_{ij}=T_{ji}$ for all $i,j\in\mathcal{I}$ and if two control areas i, j are not interconnected, then by definition we have $T_{ij}=0$. In other words, if we consider the network in terms of a weighted graph, as depicted in Fig. 2, where the nodes a_i represent the control areas and the coefficients T_{ij} indicate the existing interconnections, then the synchronization coefficients T_{ij} are the elements of the adjacency matrix.

In order to extract a discrete-time equivalent model for the tie line, we use the same global sampling frequency f_s and apply again the zero order hold method. The outcome is the difference equation:

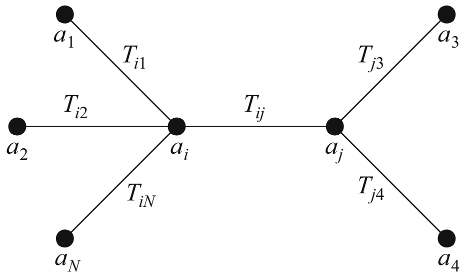


Fig. 2 Graph depiction of a network

$$\Delta P_{tie,i}[k+1]=\Delta P_{tie,i}[k]\ +T_s\sum_{j=1}^{N}\left[2\pi T_{ij}\left(\Delta f_i[k]-\Delta f_j[k]\right)\right] \tag{10}$$

where $T_s=1/f_s$ is the sampling period.

2.3 Network model

If we compute the models of all interconnected control areas along with their corresponding tie lines and then express them in the discrete-time domain, we can directly compute a discrete-time representation of the entire network. In our case, the dynamic evolution of the overall networked power system in the discrete-time domain can be described in augmented form in terms of the following difference equation:

$$S_{net}:\begin{cases}x_{net}[k+1] & = & A_{net}x_{net}[k]+B_{net}[k]+D_{net}\Delta P_{L,net}[k]\\ x_{net}[0] & = & x_{net,0}\\ y_{net}[k] & = & C_{net}x_{net}\end{cases} \tag{11}$$

where the number of the state variables per control area is $n=4$, the vector of the state variables for the entire network $x_{net}\in\mathbb{R}^{(n+1)N}$ is defined as:

$$x_{net}[k]=\begin{bmatrix}\xi_1^{\mathrm{T}}[k]\ \ldots\ \xi_N^{\mathrm{T}}[k] & \Delta P_{tie,1}[k]\ \ldots\ \Delta P_{tie,N}[k]\end{bmatrix}^{\mathrm{T}}$$

and the vector of the power load changes for the entire network $\Delta P_{L,net}\in\mathbb{R}^N$ is defined as:

$$\Delta P_{L,net}[k]=\begin{bmatrix}\Delta P_{L,1}[k] & \Delta P_{L,2}[k]\ \ldots\ \Delta P_{L,N}[k]\end{bmatrix}^{\mathrm{T}}$$

Matrix $A_{net}\in\mathbb{R}^{(n+1)N\times(n+1)N}$ is structured as:

$$A_{net}=\begin{bmatrix}A_{net,11} & A_{net,12}\\ A_{net,21} & A_{net,22}\end{bmatrix}$$

where $A_{net,11}\in\mathbb{R}^{nN\times nN}$ and $A_{net,12}\in\mathbb{R}^{nN\times N}$ associate with the interconnected control area models (8) and are defined as:

$$
\boldsymbol{A}_{net,11} = \begin{bmatrix} \boldsymbol{A}_{cl,1} & \mathbb{O}_{n\times n} & \dots & \mathbb{O}_{n\times n} \\ \mathbb{O}_{n\times n} & \boldsymbol{A}_{cl,2} & \dots & \mathbb{O}_{n\times n} \\ \vdots & \vdots & \vdots & \\ \mathbb{O}_{n\times n} & \mathbb{O}_{n\times n} & \dots & \boldsymbol{A}_{cl,N} \end{bmatrix}
$$

$$
\boldsymbol{A}_{net,12} = \begin{bmatrix} \boldsymbol{E}_{cl,1} & \mathbb{O}_{n\times 1} & \dots & \mathbb{O}_{n\times 1} \\ \mathbb{O}_{n\times 1} & \boldsymbol{E}_{cl,2} & \dots & \mathbb{O}_{n\times 1} \\ \vdots & \vdots & \vdots & \\ \mathbb{O}_{n\times 1} & \mathbb{O}_{n\times 1} & \dots & \boldsymbol{E}_{cl,N} \end{bmatrix}
$$

whereas $\boldsymbol{A}_{net,21} \in \mathbb{R}^{N\times nN}$ and $\boldsymbol{A}_{net,22} \in \mathbb{R}^{N\times N}$ associate with the tie line models (10) are defined as:

$$
\boldsymbol{A}_{net,21} = \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{L}_{12} & \dots & \boldsymbol{L}_{1N} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} & \dots & \boldsymbol{L}_{2N} \\ \vdots & \vdots & \vdots & \\ \boldsymbol{L}_{N1} & \boldsymbol{L}_{N2} & \dots & \boldsymbol{L}_{NN} \end{bmatrix}
$$

$$
\boldsymbol{A}_{net,22} = \mathbb{I}_{N\times N}
$$

The elements $\boldsymbol{L}_{ij} \in \mathbb{R}^{1\times n}$ are vector quantities associated with the tie line synchronization coefficients and are given in terms of the following equations:

$$
\boldsymbol{L}_{ij} = \begin{cases} \left[\sum\limits_{j=1}^{N} (2\pi T_{ij} T_s) \quad \mathbb{O}_{1\times(n-1)} \right], & i = j \\ \left[-2\pi T_{ij} T_s \quad \mathbb{O}_{1\times(n-1)} \right], & i \neq j \end{cases}
$$

We define $\boldsymbol{B}_{net} \in \mathbb{R}^{(n+1)N\times 1}$ and $\boldsymbol{D}_{net} \in \mathbb{R}^{(n+1)N\times N}$ as:

$$
\left\{ \begin{aligned} \boldsymbol{B}_{net}[k] &= \begin{bmatrix} \alpha_1 \boldsymbol{B}_{cl,1}\sigma_1[k] \\ \alpha_2 \boldsymbol{B}_{cl,2}\sigma_2[k] \\ \vdots \\ \alpha_N \boldsymbol{B}_{cl,N}\sigma_N[k] \\ \mathbb{O}_{N\times 1} \end{bmatrix} \\ \boldsymbol{D}_{net} &= \begin{bmatrix} \boldsymbol{D}_{net,11} \\ \mathbb{O}_{N\times N} \end{bmatrix} \\ \boldsymbol{D}_{net,11} &= \begin{bmatrix} \boldsymbol{D}_{cl,1} & \mathbb{O}_{n\times 1} & \dots & \mathbb{O}_{n\times 1} \\ \mathbb{O}_{n\times 1} & \boldsymbol{D}_{cl,2} & \dots & \mathbb{O}_{n\times 1} \\ \vdots & \vdots & & \vdots \\ \mathbb{O}_{n\times 1} & \mathbb{O}_{n\times 1} & \dots & \boldsymbol{D}_{cl,N} \end{bmatrix} \end{aligned} \right.
$$

and the matrix $\boldsymbol{C}_{net} \in \mathbb{R}^{2N\times(n+1)N}$ is defined as:

$$
\boldsymbol{C}_{net} = \begin{bmatrix} \boldsymbol{C}_{cl,1} & \mathbb{O}_{1\times n} & \dots & \mathbb{O}_{1\times n} & \mathbb{O}_{1\times N} \\ \mathbb{O}_{1\times n} & \boldsymbol{C}_{cl,2} & \dots & \mathbb{O}_{1\times n} & \mathbb{O}_{1\times N} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbb{O}_{1\times n} & \mathbb{O}_{1\times n} & \dots & \boldsymbol{C}_{cl,N} & \mathbb{O}_{1\times N} \\ \mathbb{O}_{N\times n} & \mathbb{O}_{N\times n} & \dots & \mathbb{O}_{N\times n} & \mathbb{I}_{N\times N} \end{bmatrix}
$$

For the remainder of this paper we consider that the power network evolves in the discrete-time domain and its dynamic behavior is described in terms of (11).

2.4 Stability analysis

The networked system obtained through our modeling process is Lyapunov stable. Lyapunov stability in the continuous-time domain means that there exist system eigenvalues located on the imaginary axis. Accordingly, in the discrete-time domain it means that there exist system eigenvalues located on the boundary of the unit disc. In our case, it can be shown that the continuous-time system has some eigenvalues located exactly at the origin of the complex plane and that the discrete-time system has some eigenvalues with unit value.

If the network evolves in the absence of an attacker, that is $\sigma_i = 0$ for all $i \in \mathcal{I}$, then $\tilde{y}_i = y_i$ and the system operates normally. Every power load change $\Delta P_{L,i}$ is matched with an equal increase or decrease in the produced power $\Delta P_{G,i}$. At steady-state, all frequency deviations Δf_i converge to zero along with the tie line power deviations $\Delta P_{tie,i}$. The variables $z_{1,i}$ and $z_{2,i}$ converge to some constant nonzero steady-state values and the same holds for $u_{c,i}$ and $u_{d,i}$. This scenario is studied in many textbooks [19, 20] and no instability can occur under these circumstances.

The instability is identified both in the continuous and the discrete-time domain, when the network is affected by an attacker. Suppose that the tie line power deviations $\Delta P_{tie,i}$ reach an equilibrium and converge to constant steady-state values, say $\Delta P_{tie,i,ss}$. This means that at some point we obtain

$$
\frac{\mathrm{d}}{\mathrm{d}t}\Delta P_{tie,i}(t) = 0 \quad \forall i = 1, 2, \dots, N \tag{12}
$$

for the continuous-time domain and

$$
\Delta P_{tie,i}[k+1] = \Delta P_{tie,i}[k] \quad \forall i = 1, 2, \dots, N \tag{13}
$$

for the the discrete-time domain. Expressions (12), (13) along with (9), (10) imply that if an equilibrium is to be reached, then the frequency deviations Δf_i of all control areas must converge to the same steady-state value, that is $\lim_{t\to\infty} \Delta f_i(t) = \Delta f_{i,ss} = \Delta f_{ss}$ for all $i = 1, 2, \dots, N$.

Now, we identify two distinct cases. First, the case where all control areas are affected by the same attack signal and then, the case where the attack signal differs from one control area to another. The scenario where the adversary affects only some control areas and not all of them falls under the second case, where some $\alpha_i = 0$.

Let us consider the case where all control areas are affected by the same attack signal, that is $\alpha_i = \alpha$ for all $i \in \mathcal{I}$. The frequency deviation reference signal is set as $\Delta f_i^{ref} = 0$ and $\tilde{y}_i(t) = y_i(t) - \alpha_i$. Therefore, (4) can be written as:

$$z_{1,i}(t) = \frac{1}{f^\circ} \int_0^t (\alpha_i - y_i(\tau))\mathrm{d}\tau \tag{14}$$

and implies that this attack scenario essentially alters all the reference signals Δf_i^{ref} from zero to α. For the frequency deviations, we have $\lim_{t\to\infty} \Delta f_i(t) = \alpha$ for all $i = 1, 2, \ldots, N$, based either on the continuous-time equation (14) or the discrete-time equation (6). For the tie line power deviations, we have $\lim_{t\to\infty} \Delta P_{tie,i}(t) = 0$ for all $i = 1, 2, \ldots, N$, based either on the continuous-time equations (5), (9), (12) or the discrete-time equations (7), (10), (13). The variables $z_{1,i}$ and $z_{2,i}$ converge to some constant nonzero steady-state values, such that all $u_{c,i}$ and $u_{d,i}$ produce the required $\Delta P_{G,i}$ to match both the load changes $\Delta P_{L,i}$ and the nonzero steady-state frequency deviations. Since all state variables converge to constant values, the system is stable.

Let us now consider the case where the attack signal differs from one control area to another. For a persistent attacker, we have $\lim_{t\to\infty} \Delta P_{tie,i}(t) = \Delta P_{tie,i,ss} \neq 0$. This implies two things. Firstly, as $t \to \infty$, (5) integrates a constant nonzero quantity, therefore all $z_{2,i}$ are forced to increase linearly over time. Secondly, once $\Delta P_{tie,i}$ reaches an equilibrium, all Δf_i must have converged to the same steady-state value, which is nonetheless different than the one that each α_i dictates, that is $\lim_{t\to\infty} \Delta f_i(t) = \lim_{t\to\infty} y_i(t) \neq \alpha_i$. In turn, (14), as $t \to \infty$, integrates a constant nonzero quantity, therefore all $z_{1,i}$ are forced to increase linearly over time as well. The same two results can be obtained from the discrete-time equations (6), (7). Although $z_{1,i}$ and $z_{2,i}$ tend to infinity, their opposite signs drive $u_{c,i}$ and $u_{d,i}$ to some constant steady-state values, causing each generator to produce the necessary power $\Delta P_{G,i}$ that satisfies both the load changes $\Delta P_{L,i}$ and the nonzero steady-state frequency deviations.

In conclusion, the network instability appears when the attack signal differs from one control area to another or when the attacker affects only some control areas of the network. We highlight that instead of $z_{1,i}$ and $z_{2,i}$ we could choose different integral variables, say:

$$w_i(t) = \int_0^t ACE_i(\tau)\mathrm{d}\tau = |P^\circ_{tie,i}| z_{2,i}(t) + B_i f^\circ z_{1,i}(t)$$

and obtain an asymptotically stable system, since the w_i will follow the convergence of $u_{c,i}$ and $u_{d,i}$. However, in this case we lose our main advantage, which is the ability to always detect an adversary, unless all of the control areas are simultaneously corrupted by the same attack signal. The unstable character of $z_{1,i}$ and $z_{2,i}$ ensures that the state trajectory will, sooner or later, exit any given convex and compact robust invariant set, causing the adversary to be disclosed. The trade-off for using the states $z_{1,i}$ and $z_{2,i}$ is the increased complexity in the calculation of the robust invariant set, due to the Lyapunov stable network dynamics.

3 Attack detector design

In order to extract the robust invariant set that will be exploited as a set-theoretic attack detector, our first priority is to determine a set of state constraints that ensures the risk-free behavior of the network. This set can be obtained by enforcing suitable bounds on each state variable based on the standard safety considerations invoked in the literature.

According to [8], large frequency deviations, that may occur during the transients, jeopardize the stability of the grid. Thus, the frequency deviation $\Delta f_i[k]$ should always respect the inequality:

$$|\Delta f_i[k]| \leq \Delta f_{i,\max} = 1.5 \text{ Hz} \quad \forall k \in \mathbb{N}$$

The hard constraints imposed on the control signal imply that similar bounds exist for $\Delta P_{G,i}[k]$. The discretization does not alter the dc-gains of the system and the turbine has a unit dc-gain. Therefore, the bounds of $\Delta P_{G,i}$, $u_{c,i}$ and $u_{d,i}$ are identical with each other and the mechanical power produced in the output of the turbine should always respect the inequality

$$|\Delta P_{G,i}[k]| \leq \Delta P_{G,i,\max} = u_{i,\max} \quad \forall k \in \mathbb{N}$$

In comparison to the bounds of Δf_i, the constraints imposed on $\Delta P_{G,i}$ are hard and can never be violated.

The variables $z_{1,i}[k]$ and $z_{2,i}[k]$ are measured in time units and, according to [19, 22], it is always necessary to limit the deviation of the synchronous clocks driven by the system frequency. Thus, the accumulated time errors should always respect the inequalities:

$$\begin{cases} |z_{1,i}[k]| \leq z_{1,i,\max} = 3 \text{ s} & \forall k \in \mathbb{N} \\ |z_{2,i}[k]| \leq z_{2,i,\max} = 3 \text{ s} & \forall k \in \mathbb{N} \end{cases}$$

Each tie line connects a control area to the network and is designed to transfer a nominal amount of power. After every power load change, the tie line power deviates from its nominal value. Large and persistent oscillations of

$\Delta P_{tie,i}[k]$ are generally undesirable, because they stress the tie line to its thermal limits and threaten the stability of the grid. Therefore, the tie line power deviations should always respect the inequality:

$$|\Delta P_{tie,i}[k]| \leq \Delta P_{tie,i,\max} \quad \forall k \in \mathbb{N}$$

The constraints associated with ξ_i can be expressed in terms of the set:

$$\mathcal{X}_i(\boldsymbol{Q}_i, \boldsymbol{q}_i) = \{\xi_i \in \mathbb{R}^n : \boldsymbol{Q}_i \xi_i \leq \boldsymbol{q}_i\}$$

where $\boldsymbol{Q}_i \in \mathbb{R}^{2n \times n}$ and $\boldsymbol{q}_i \in \mathbb{R}^{2n}$ are given as:

$$\begin{cases} \boldsymbol{Q}_i = \begin{bmatrix} \mathbb{I}_{n \times n} \\ -\mathbb{I}_{n \times n} \end{bmatrix} \\ \boldsymbol{q}_i = \begin{bmatrix} \boldsymbol{q}_{i,\max} \\ \boldsymbol{q}_{i,\max} \end{bmatrix} \end{cases}$$

$$\boldsymbol{q}_{i,\max} = \begin{bmatrix} \Delta f_{i,\max} & \Delta P_{G,i,\max} & z_{1,i,\max} & z_{2,i,\max} \end{bmatrix}^{\mathrm{T}}$$

The constraints associated with $\Delta P_{tie,i}$ can be expressed in terms of the set:

$$\bar{\mathcal{X}}_i(\bar{\boldsymbol{Q}}_i, \bar{\boldsymbol{q}}_i) = \{\Delta P_{tie,i} \in \mathbb{R} : \bar{\boldsymbol{Q}}_i \Delta P_{tie,i} \leq \bar{\boldsymbol{q}}_i\}$$

where $\bar{\boldsymbol{Q}}_i \in \mathbb{R}^{2 \times 1}$ and $\bar{\boldsymbol{q}}_i \in \mathbb{R}^{2 \times 1}$ are given as:

$$\begin{cases} \bar{\boldsymbol{Q}}_i = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ \bar{\boldsymbol{q}}_i = \begin{bmatrix} \Delta P_{tie,i,\max} \\ \Delta P_{tie,i,\max} \end{bmatrix} \end{cases}$$

The constraints associated with $\boldsymbol{x}_{net}$ are determined by combining the sets $\mathcal{X}_i$ and $\bar{\mathcal{X}}_i$ for all $i \in \mathcal{I}$ and can be expressed in terms of the set:

$$\mathcal{X}_{net} = \left\{\boldsymbol{x}_{net} \in \mathbb{R}^{(n+1)N} : \boldsymbol{Q}_{net} \boldsymbol{x}_{net} \leq \boldsymbol{q}_{net}\right\}$$

The matrix $\boldsymbol{Q}_{net}$ is structured as

$$\boldsymbol{Q}_{net} = \begin{bmatrix} \boldsymbol{Q}_{net,11} & \boldsymbol{Q}_{net,12} \\ \boldsymbol{Q}_{net,21} & \boldsymbol{Q}_{net,22} \end{bmatrix}$$

and the vector $\boldsymbol{q}_{net}$ is structured as:

$$\boldsymbol{q}_{net} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}} \ldots \boldsymbol{q}_N^{\mathrm{T}} & \bar{\boldsymbol{q}}_1^{\mathrm{T}} \ldots \bar{\boldsymbol{q}}_N^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

The blocks $\boldsymbol{Q}_{net,11} \in \mathbb{R}^{2nN \times nN}$ and $\boldsymbol{Q}_{net,22} \in \mathbb{R}^{2N \times N}$ associate with the constraints imposed on ξ_i and $\Delta P_{tie,i}$ respectively and are defined as:

$$\boldsymbol{Q}_{net,11} = \begin{bmatrix} \boldsymbol{Q}_1 & \mathbb{O}_{2n \times n} & \ldots & \mathbb{O}_{2n \times n} \\ \mathbb{O}_{2n \times n} & \boldsymbol{Q}_2 & \ldots & \mathbb{O}_{2n \times n} \\ \vdots & \vdots & \vdots & \\ \mathbb{O}_{2n \times n} & \mathbb{O}_{2n \times n} & \ldots & \boldsymbol{Q}_N \end{bmatrix}$$

$$\boldsymbol{Q}_{net,22} = \begin{bmatrix} \bar{\boldsymbol{Q}}_1 & \mathbb{O}_{2 \times 1} & \ldots & \mathbb{O}_{2 \times 1} \\ \mathbb{O}_{2 \times 1} & \bar{\boldsymbol{Q}}_2 & \ldots & \mathbb{O}_{2 \times 1} \\ \vdots & \vdots & \vdots & \\ \mathbb{O}_{2 \times 1} & \mathbb{O}_{2 \times 1} & \ldots & \bar{\boldsymbol{Q}}_N \end{bmatrix}$$

whereas the blocks $\boldsymbol{Q}_{net,12}$ and $\boldsymbol{Q}_{net,21}$ are defined as:

$$\begin{cases} \boldsymbol{Q}_{net,12} = \mathbb{O}_{2nN \times N} \\ \boldsymbol{Q}_{net,21} = \mathbb{O}_{2N \times nN} \end{cases}$$

We may also define the admissible states ξ_i that result in a control law $u_{d,i}[k]$ respecting the hard input constraints in terms of the set:

$$\mathcal{U}_i(\boldsymbol{P}_i, \boldsymbol{p}_i) = \{\xi_i \in \mathbb{R}^n : \boldsymbol{P}_i \xi_i \leq \boldsymbol{p}_i\}$$

where $\boldsymbol{P}_i \in \mathbb{R}^{2 \times n}$ and $\boldsymbol{p}_i \in \mathbb{R}^2$ are given as:

$$\begin{cases} \boldsymbol{P}_i = \begin{bmatrix} -(1/R_i)\boldsymbol{C}_i & \bar{K}_{I_1,i} & \bar{K}_{I_2,i} \\ (1/R_i)\boldsymbol{C}_i & -\bar{K}_{I_1,i} & -\bar{K}_{I_2,i} \end{bmatrix} \\ \boldsymbol{p}_i = \begin{bmatrix} u_{i,\max} \\ u_{i,\max} \end{bmatrix} \end{cases}$$

The set of the admissible values of $\boldsymbol{x}_{net}$ is determined by combining the sets $\mathcal{U}_i$ for all $i \in \mathcal{I}$ and can be expressed as:

$$\mathcal{U}_{net} = \left\{\boldsymbol{x}_{net} \in \mathbb{R}^{(n+1)N} : \boldsymbol{P}_{net} \boldsymbol{x}_{net} \leq \boldsymbol{p}_{net}\right\}$$

The matrix $\boldsymbol{P}_{net}$ is structured as:

$$\boldsymbol{P}_{net} = [\boldsymbol{P}_{net,1} \quad \boldsymbol{P}_{net,2}]$$

and the vector $\boldsymbol{p}_{net}$ is structured as:

$$\boldsymbol{p}_{net} = \begin{bmatrix} \boldsymbol{p}_1^{\mathrm{T}} & \boldsymbol{p}_2^{\mathrm{T}} \ldots \boldsymbol{p}_N^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

The blocks $\boldsymbol{P}_{net,1} \in \mathbb{R}^{2N \times nN}$ and $\boldsymbol{P}_{net,2}$ are defined as:

$$\begin{cases} \boldsymbol{P}_{net,1} = \begin{bmatrix} \boldsymbol{P}_1 & \mathbb{O}_{2 \times n} & \ldots & \mathbb{O}_{2 \times n} \\ \mathbb{O}_{2 \times n} & \boldsymbol{P}_2 & \ldots & \mathbb{O}_{2 \times n} \\ \vdots & \vdots & & \vdots \\ \mathbb{O}_{2 \times n} & \mathbb{O}_{2 \times n} & \ldots & \boldsymbol{P}_N \end{bmatrix} \\ \boldsymbol{P}_{net,2} = \mathbb{O}_{2N \times N} \end{cases}$$

Finally, we can define the set of the admissible network disturbances $\Delta \boldsymbol{P}_{L,net}$ as:

$$\mathcal{W}_{net} = \left\{\Delta \boldsymbol{P}_{L,net} \in \mathbb{R}^N : \boldsymbol{Q}_{net,22} \Delta \boldsymbol{P}_{L,net} \leq \boldsymbol{r}_{net}\right\}$$

where the vector $\boldsymbol{r}_{net} \in \mathbb{R}^{2N}$ is structured as:

$$\begin{cases} \boldsymbol{r}_{net} = \begin{bmatrix} \boldsymbol{r}_1^{\mathrm{T}} & \boldsymbol{r}_2^{\mathrm{T}} \dots \boldsymbol{r}_N^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{r}_i = \begin{bmatrix} \Delta P_{L,i,\max} \\ \Delta P_{L,i,\max} \end{bmatrix} \end{cases}$$

A set is robust invariant when all initial conditions belonging to this set generate state trajectories remaining inside the same set for all future time instances and for all bounded disturbance sequences. A formal definition of robust invariance can be found in [23, 24]. If we consider $\Delta P_{tie,i}$ as an additional disturbance on the model of each control area (8), then we could try to assign robust invariant sets to each control area individually. However, this concept is invalid since (7), (8) imply that a constant $\Delta P_{tie,i}$ will drive $z_{2,i}$ towards infinity, hence the existence of a robust invariant set is immediately denied. In reality, $\Delta P_{tie,i}$ will decay, but the robust approach must consider all potential disturbance sequences. Consequently, the robust invariant set has to be extracted in a centralized manner considering the network dynamics in (11). We remark that since the robust invariant set will be used in order to detect an adversary, it has to be determined considering the dynamics (11) in the absence of an attacker, that is when $\sigma_i[k] = 0$, for all $k \in \mathbb{N}$, $i \in \mathcal{I}$.

According to [15, 16], the input hard constraints do not allow the controllers to perform unsaturated for all states $\boldsymbol{x}_{net} \in \boldsymbol{\mathcal{X}}_{net}$. To solve this problem, we can define the set $\boldsymbol{\mathcal{A}}_{net} = \boldsymbol{\mathcal{X}}_{net} \cap \boldsymbol{\mathcal{U}}_{net}$ and then try to determine the maximal subset of $\boldsymbol{\mathcal{A}}_{net}$ that is robust positively invariant with respect to the network dynamics in (11). This new set, denoted with $\boldsymbol{\mathcal{A}}_{net,\infty}$, is defined as:

$$\begin{aligned} \boldsymbol{\mathcal{A}}_{net,\infty} = & \{\boldsymbol{x}_{net,0} \in \boldsymbol{\mathcal{A}}_{net} : \boldsymbol{A}_{net}\boldsymbol{x}_{net}[k] + \boldsymbol{D}_{net}\Delta \boldsymbol{P}_{L,net}[k] \\ & \in \boldsymbol{\mathcal{A}}_{net,\infty},\ \forall \Delta \boldsymbol{P}_{L,net}[k] \in \boldsymbol{\mathcal{W}}_{net},\ \forall k \in \mathbb{N}\} \end{aligned}$$

An efficient algorithm for the computation of maximal robust invariant subsets was proposed in [25]. However, this algorithm ensures finite time determination of these sets only for systems described by asymptotically stable dynamics. Due to the abundance of the integral control actions, the dynamics of the network are Lyapunov stable. This fact implies that there exist eigenvalues of the matrix $\boldsymbol{A}_{net}$ located exactly on the boundary of the unit disc. In this article, we apply the methods of [25, 26] and we compute an approximation of $\boldsymbol{\mathcal{A}}_{net,\infty}$, based on the structure of the network.

The key idea, is to separate the network dynamics into an asymptotically stable compartment and a Lyapunov stable one. This is always possible through a suitable similarity transformation of the state space coordinates. If we solve the eigensystem $\boldsymbol{A}_{net}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{F}$, then we can compute a diagonal matrix $\boldsymbol{F}$ containing the eigenvalues of $\boldsymbol{A}_{net}$ and an invertible matrix $\boldsymbol{V}$ containing the eigenvectors of the system. For complex eigenvalues, it is trivial to render the matrix $\boldsymbol{F}$ in its equivalent block-diagonal real form and compute the matrix $\boldsymbol{V}$ accordingly. Using the change of variables $\boldsymbol{\psi}[k] = \boldsymbol{V}^{-1}\boldsymbol{x}_{net}[k]$ the network dynamics of (11) can be written as:

$$S_{net}^{(\psi)} : \boldsymbol{\psi}[k+1] = \boldsymbol{F}\boldsymbol{\psi}[k] + \boldsymbol{H}\Delta \boldsymbol{P}_{L,net}[k] \quad \boldsymbol{\psi}[0] = \boldsymbol{\psi}_0$$

where the matrices $\boldsymbol{F}$ and $\boldsymbol{H}$ are given as:

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_S & \mathbb{O}_{s_1 \times s_2} \\ \mathbb{O}_{s_2 \times s_1} & \boldsymbol{F}_L \end{bmatrix}$$

$$\boldsymbol{H} = \boldsymbol{V}^{-1}\boldsymbol{D}_{net} = \begin{bmatrix} \boldsymbol{H}_S \\ \boldsymbol{H}_L \end{bmatrix}$$

for some partitioning indices $s_1, s_2 \in \mathbb{N}^*$ such that $s_1 + s_2 = (n+1)N$. The matrices $\boldsymbol{F}_S \in \mathbb{R}^{s_1 \times s_1}$ and $\boldsymbol{F}_L \in \mathbb{R}^{s_2 \times s_2}$ associate with the asymptotically stable and the Lyapunov stable dynamics respectively, $\boldsymbol{H}_S \in \mathbb{R}^{s_1 \times N}$, $\boldsymbol{H}_L \in \mathbb{R}^{s_2 \times N}$ and the state vector $\boldsymbol{\psi}$ can be split into two compartments as:

$$\boldsymbol{\psi}[k] = \begin{bmatrix} \boldsymbol{\psi}_S^{\mathrm{T}} & \boldsymbol{\psi}_L^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

where $\boldsymbol{\psi}_S \in \mathbb{R}^{s_1}$ and $\boldsymbol{\psi}_L \in \mathbb{R}^{s_2}$. We remark that each vector evolves independently from the other, due to the form of the matrix $\boldsymbol{F}$.

The network model (11) obtained through the discretization process in the previous section has two special characteristics. Firstly, the Lyapunov stable eigenvalues have all unit values, therefore $\boldsymbol{F}_L = \mathbb{I}_{s_2 \times s_2}$ and secondly the matrix $\boldsymbol{D}_{net}$ is sparse, so that even after the change of variables, the matrix $\boldsymbol{H}_L$ satisfies the condition $\boldsymbol{H}_L = \mathbb{O}_{s_2 \times N}$. We remark that if the matrix $\boldsymbol{H}_L$ contained any nonzero elements, then a robust invariant set would not exist.

As the interconnection of two polyhedra, set $\boldsymbol{\mathcal{A}}_{net}$ will have the generic polyhedral representation

$$\boldsymbol{\mathcal{A}}_{net} = \left\{ \boldsymbol{x}_{net} \in \mathbb{R}^{(n+1)N} : \boldsymbol{G}\boldsymbol{x}_{net} \leq \boldsymbol{g} \right\}$$

The change of variables $\boldsymbol{\psi}[k] = \boldsymbol{V}^{-1}\boldsymbol{x}_{net}[k]$ gives the representation of $\boldsymbol{\mathcal{A}}_{net}$ in the $\boldsymbol{\psi}$-domain as:

$$\boldsymbol{\mathcal{A}}_{net}^{(\psi)} = \{ \boldsymbol{\psi} \in \mathbb{R}^{s_1+s_2} : \bar{\boldsymbol{G}}\boldsymbol{\psi} \leq \bar{\boldsymbol{g}} \} \quad \bar{\boldsymbol{G}} = \boldsymbol{G}\boldsymbol{V} \quad \bar{\boldsymbol{g}} = \boldsymbol{g}$$

and its maximal robust invariant subset is defined as:

$$\begin{aligned} \boldsymbol{\mathcal{A}}_{net,\infty}^{(\psi)} = & \Big\{ \boldsymbol{\psi}_0 \in \boldsymbol{\mathcal{A}}_{net}^{(\psi)} : \boldsymbol{F}\boldsymbol{\psi}[k] + \boldsymbol{H}\Delta \boldsymbol{P}_{L,net}[k] \\ & \in \boldsymbol{\mathcal{A}}_{net,\infty}^{(\psi)},\ \forall \Delta \boldsymbol{P}_{L,net}[k] \in \boldsymbol{\mathcal{W}}_{net},\ \forall k \in \mathbb{N} \Big\} \end{aligned}$$

According to [25], a finite time determined approximation of $\boldsymbol{\mathcal{A}}_{net,\infty}^{(\psi)}$ is the set $\hat{\boldsymbol{\mathcal{A}}}_{net,\infty}^{(\psi)}$ defined as:

$$\begin{aligned}\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty} =&\Big\{\boldsymbol{\psi}_0 \in \boldsymbol{\mathcal{A}}^{(\psi)}_{net} : \hat{\boldsymbol{C}}_S\boldsymbol{\psi}_S[k] + \hat{\boldsymbol{C}}_L\boldsymbol{\psi}_L[k] \\ &\in \boldsymbol{\mathcal{L}}' \times \boldsymbol{\mathcal{A}}^{(\psi)}_{net},\ \forall \Delta\boldsymbol{P}_{L,net}[k] \in \boldsymbol{\mathcal{W}}_{net},\ \forall k \in \mathbb{N}\Big\}\end{aligned}$$

where the matrices $\hat{\boldsymbol{C}}_S$ and $\hat{\boldsymbol{C}}_L$ are given as:

$$\begin{cases}\hat{\boldsymbol{C}}_S = \begin{bmatrix}\mathbb{O}_{s_2\times s_1}\\ \boldsymbol{C}_S\end{bmatrix}\\ \hat{\boldsymbol{C}}_L = \begin{bmatrix}\mathbb{I}_{s_2\times s_2}\\ \boldsymbol{C}_L\end{bmatrix}\end{cases}$$

the matrices $\boldsymbol{C}_S \in \mathbb{R}^{(s_1+s_2)\times s_1}$, $\boldsymbol{C}_L \in \mathbb{R}^{(s_1+s_2)\times s_2}$ satisfy the equation $[\boldsymbol{C}_S \quad \boldsymbol{C}_L] = \mathbb{I}_{(s_1+s_2)\times(s_1+s_2)}$, the set $\boldsymbol{\mathcal{L}}'$ is defined as:

$$\boldsymbol{\mathcal{L}}' = \left\{\boldsymbol{\psi}_L \in \mathbb{R}^{s_2} : \boldsymbol{C}_L\boldsymbol{F}^k_L\boldsymbol{\psi}_L \in \boldsymbol{\mathcal{Y}}',\ \forall k \in \mathbb{N}\right\} \tag{15}$$

and the set $\boldsymbol{\mathcal{Y}}' \subset \mathbb{R}^{s_1+s_2}$ is determined via the following standardized procedure [25]. Consider the recursion:

$$\begin{cases}\phi^i_0 = \bar{g}^i\\ \phi^i_{k+1} = \phi^i_k - h_{\boldsymbol{\mathcal{W}}_{net}}((\boldsymbol{C}_S\boldsymbol{F}^k_S\boldsymbol{H}_S)^{\mathrm{T}}\bar{\boldsymbol{G}}_i)\end{cases}$$

The parameters ϕ^i_k and $\bar{g}^i$ denote the i-th element of the vectors $\boldsymbol{\phi}_k$ and $\bar{\boldsymbol{g}}$ respectively, $\bar{\boldsymbol{G}}_i$ stands for the i-th row of the matrix $\bar{\boldsymbol{G}}$, while the mapping:

$$h_{\boldsymbol{\mathcal{W}}_{net}}(\eta) = \sup_{\boldsymbol{x}\in\boldsymbol{\mathcal{W}}_{net}} (\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{x})$$

represents the support function of the set $\boldsymbol{\mathcal{W}}_{net}$. If we specify a scalar $0<\beta<1$, then there exists a k^* such that $\beta\boldsymbol{\phi}_{k^*} - \boldsymbol{\theta}_{k^*} \geq 0$, where the i-th element of the vector $\boldsymbol{\theta}_{k^*}$ is defined as:

$$\boldsymbol{\theta}^i_{k^*} = \lambda\zeta_i(1-\mu)^{-1}\mu^{k^*}$$

The parameter $\mu \in \mathbb{R}^*_+$ is the spectral radius of the matrix $\boldsymbol{F}_S$, the scalar $\lambda \in \mathbb{R}^*_+$ is selected such that $\boldsymbol{\mathcal{W}}_{net} \subset \lambda\boldsymbol{\mathcal{B}}_2(N)$, with $\boldsymbol{\mathcal{B}}_2(N)$ representing the 2-norm unit ball in $\mathbb{R}^N$ and the constants $\zeta_i \in \mathbb{R}^*_+$ can always be determined such that $\|(\boldsymbol{C}_S\boldsymbol{F}^{k^*}_S\boldsymbol{H}_S)^{\mathrm{T}}\bar{\boldsymbol{G}}_i\|_2 \leq \zeta_i\mu^{k^*}$ for all i. Let $\boldsymbol{\phi}'$ satisfy:

$$(1-\beta)\boldsymbol{\phi}_{k^*} < \boldsymbol{\phi}' < \boldsymbol{\phi}_{k^*} - \boldsymbol{\theta}_{k^*}$$

Then, the set $\boldsymbol{\mathcal{Y}}'$ is defined as:

$$\boldsymbol{\mathcal{Y}}' = \left\{\boldsymbol{\psi} \in \mathbb{R}^{s_1+s_2} : \bar{\boldsymbol{G}}\boldsymbol{\psi} \leq \boldsymbol{\phi}'\right\}$$

and since $\boldsymbol{F}_L = \mathbb{I}_{s_2\times s_2}$ the set $\boldsymbol{\mathcal{L}}'$ of (15) is defined as:

$$\boldsymbol{\mathcal{L}}' = \left\{\boldsymbol{\psi}_L \in \mathbb{R}^{s_2} : \bar{\boldsymbol{G}}\boldsymbol{C}_L\boldsymbol{\psi}_L \leq \boldsymbol{\phi}'\right\}$$

Considering the polyhedral nature of the sets $\boldsymbol{\mathcal{L}}'$ and $\boldsymbol{\mathcal{A}}^{(\psi)}_{net}$, their Cartesian product can be computed as:

$$\begin{aligned}\boldsymbol{\mathcal{L}}' \times \boldsymbol{\mathcal{A}}^{(\psi)}_{net} = \Bigg\{ &\begin{bmatrix}\boldsymbol{\psi}_L\\ \boldsymbol{\psi}\end{bmatrix} \in \mathbb{R}^{s_1+2s_2} : \\ &\begin{bmatrix}\bar{\boldsymbol{G}}\boldsymbol{C}_L & \mathbb{O}_{l\times(s_1+s_2)}\\ \mathbb{O}_{l\times s_2} & \bar{\boldsymbol{G}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\psi}_L\\ \boldsymbol{\psi}\end{bmatrix} \leq \begin{bmatrix}\boldsymbol{\phi}'\\ \bar{\boldsymbol{g}}\end{bmatrix}\Bigg\}\end{aligned}$$

and l denotes the number of rows of the matrices $\bar{\boldsymbol{G}}\boldsymbol{C}_L$ and $\bar{\boldsymbol{G}}$. The attack detection mechanism can now be formally introduced in terms of the alarm signal

$$\rho(\boldsymbol{\psi}) = \begin{cases}0 & \boldsymbol{\psi} \in \hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}\\ 1 & \text{otherwise}\end{cases}$$

and it is triggered whenever the vector $\boldsymbol{\psi}$ exits the robust invariant set $\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}$. We assume that the vector $\boldsymbol{x}_{net}[k]$ is available to the control center at any given time instant k in order to allow the real-time computation of the vector $\boldsymbol{\psi}[k]$.

4 Switching signal design

The authors in [15, 16] studied bias injection cyber-attacks on the frequency sensor measurements, when the attack signal α_i was set to a constant value and affected the system indefinitely. It was shown that in the case of a single-area power plant the structural properties of the integral controllers force the frequency deviation to regulate wherever the attack signal dictates. However, the only way for an adversary to regulate the frequency in a networked system is to incorporate a coordinated attack on all interconnected areas using the same attack signal α_i for all $i \in \mathcal{I}$.

A more realistic scenario would be to consider that an attack occurs not on every frequency sensor of the grid at the same time, but only to those sensors that are successfully compromised by the adversary. In this case, only a few of the Lyapunov stable dynamics of the integral controllers will be affected by the attack signal and therefore input to state instability is unavoidable. Specifically, the state variables $z_{1,i}$, $z_{2,i}$ are forced to diverge linearly towards infinity for as long as the attacker remains active and the set-induced anomaly detector will ultimately trigger an alarm.

Since persistent attacks on individual control areas seem inevitably detectable in terms of our set-theoretic approach, the only alternative for the adversary is to prolong their disclosure. The attacker can attempt to remain undetected for a longer period by means of a hysteresis-based switching pattern [27]

$$\sigma_i[k] = \begin{cases}0 & |\tilde{y}_i[k]| > \bar{\alpha}_{i,\max} \quad \text{and} \quad \sigma_i[k-1] = 1\\ 1 & |y_i[k]| < \bar{\alpha}_{i,\min} \quad \text{and} \quad \sigma_i[k-1] = 0\\ \sigma_i[k-1] & \text{otherwise}\end{cases}$$

where $\bar{\alpha}_{i,\max} = \alpha_{i,\max} - \delta$ and $\bar{\alpha}_{i,\min} = \alpha_{i,\min} + \delta$ are the hysteresis bounds and $\delta \in \mathbb{R}_{+}^{*}$ is the tolerance factor ensuring that a switching can occur only inside the frequency zone $y_i \in \left[\Delta f_{i,\min}^{\alpha}, \Delta f_{i,\max}^{\alpha}\right] = \left[\alpha_{i,\min}, \alpha_{i,\max}\right]$.

The only attack resources required in our scenarios are the knowledge of the frequency measurements y_i and the ability to corrupt them. However, the bounds that we used to limit the load changes, the state variables and the control inputs of the system can be easily obtained, since some of them are standard in the literature. We remark that knowledge alone of these bounds is not enough to defeat our detection mechanism. The attacker can remain potentially undetected only if he additionally has full knowledge of the state vector ξ_{net} and of the robust invariant set $\hat{\boldsymbol{\mathcal{A}}}_{net,\infty}^{(\psi)}$. However, to determine this set, the attacker must also know the exact model of the network, namely the matrices $\boldsymbol{A}_{net}$, $\boldsymbol{B}_{net}$, $\boldsymbol{C}_{net}$ and $\boldsymbol{D}_{net}$, along with the design-dependent set $\boldsymbol{\mathcal{L}}'$. Only then the adversary can reproduce the set $\hat{\boldsymbol{\mathcal{A}}}_{net,\infty}^{(\psi)}$ and use it to develop elaborate state-dependent switching patterns that will prevent an alarm activation.

5 Simulation studies

In this section, we study the load frequency control loop of the benchmark two-area power plant considering two distinct attack scenarios. Initially, we address the case where an adversary compromises the frequency sensor measurements of the first control area and corrupts the data transmitted to the automatic generation control unit with an intermittent bias injected attack signal α_1. The simulations indicate that intermittent attack patterns driven by the switching logic developed in the previous section are harder to detect and they cause the state variables to oscillate. In the sequel, we address the cases of detectable and undetectable coordinated cyber attacks occurring simultaneously on both control areas, using two persistent bias injected attack signals α_1, α_2. The simulations highlight that if the attack signals have the same value, then the adversary is able to regulate the frequency deviation of the network to any safety-critical steady-state value. Both scenarios demonstrate the ability of a set-theoretic attack detector to disclose either a persistent or an intermittent adversary, even in the presence of unknown disturbances.

Table 1 Parameter values for control area 1

Parameter	Symbol	Value	Unit
Power base	$P_{B,1}$	2000	MW
Load dependency factor	D_1	16.66	MW/Hz
Speed droop	R_1	1.2×10^{-3}	Hz/MW
Generator inertia constant	H_1	5	s
Turbine static gain	$K_{T,1}$	1	MW/MW
Turbine time constant	$T_{T,1}$	0.3	s
Area static gain	$K_{p,1}$	0.06	Hz/MW
Area time constant	$T_{p,1}$	24	s
Controller static gain	$K_{I,1}$	0.5	1/s
Control input bound	$u_{1,\max}$	600	MW
Power load bound	$\Delta P_{L,1,\max}$	20	MW

Table 2 Parameter values for control area 2

Parameter	Symbol	Value	Unit
Power base	$P_{B,2}$	1500	MW
Load dependency factor	D_2	10.5	MW/Hz
Speed droop	R_2	1.33×10^{-3}	Hz/MW
Generator inertia constant	H_2	4	s
Turbine static gain	$K_{T,2}$	1	MW/MW
Turbine time constant	$T_{T,2}$	0.25	s
Area static gain	$K_{p,2}$	0.095	Hz/MW
Area time constant	$T_{p,2}$	22.85	s
Controller static gain	$K_{I,2}$	0.45	1/s
Control input bound	$u_{2,\max}$	450	MW
Power load bound	$\Delta P_{L,2,\max}$	15	MW

The parameters of the two-area power network that were used in the simulations are provided in the Tables 1 and 2 [19]. For completeness, we also provide the formulas associated with the gains $K_{p,i}$ and the constants $T_{p,i}$ as:

$$\begin{cases} K_{p,i} = \dfrac{1}{D_i} \\ T_{p,i} = \dfrac{2H_i P_{B,i}}{f^{\circ} D_i} \end{cases}$$

where $f^{\circ} = 50\,\text{Hz}$ is the nominal network frequency. We assume that the simulations start at $k = 0$, that the initial condition is $\boldsymbol{x}_{net}[0] = 0$ and that the duration is the time interval $t \in [0, 35]$. For a global sampling frequency $f_s = 100\,\text{Hz}$, we have $k \in [0, 3.5 \times 10^3]$. Furthermore, since our main objective is to assess the efficiency of the set-theoretic attack detector in the presence of disturbances, we assume that the two-area network is subject to the following power load changes:

$$\Delta P_{L,1}(t) = 20\,\text{MW} \qquad t \geq 0\ \text{s}$$

$$\Delta P_{L,2}(t) = \begin{cases} 0\ \text{MW} & 0 \leq t < 10\ \text{s} \\ -5\ \text{MW} & t \geq 10\ \text{s} \end{cases}$$

The tie line is assumed to be lossless, the nominal exchanged power is $P^{\circ}_{tie,1} = -P^{\circ}_{tie,2} = 1000$ MW and the synchronization coefficients are assigned the values $T_{12} = T_{21} = 175$ MW/rad.

The bounds of the load changes $\Delta P_{L,i,\max}$ were selected as small percentages of the power base $P^{\circ}_{B,i}$ of each control area [19, 20]. The bounds of the state variables $\Delta f_{i,\max} = 1.5$ Hz, $\Delta P_{G,i,\max} = u_{i,\max}$ and $z_{1,i,\max} = z_{2,i,\max} = 3$ s are standardized, holding for all $i = 1, 2$. In particular, the bounds $\Delta f_{i,\max}$ are mentioned in [8], whereas the bounds $z_{1,i,\max}$ and $z_{2,i,\max}$ are derived from [22]. The bounds of the tie line power deviations $\Delta P_{tie,i,\max}$ were selected through extensive simulations. Specifically, we observed that even when the maximum admissible power load changes $\Delta P_{L,i,\max}$ occurred, the graphs of $\Delta P_{tie,i}$ never exceeded the values $\Delta P_{tie,i,\max} = 0.5|P^{\circ}_{tie,i}|$. The bounds of the control signals $u_{i,\max}$ were selected in the following manner. First of all, they have to be at least equal to the maximum admissible load changes, in order to be able to service them. Furthermore, they have to amend for potential overshoots during the transient response, so they need to be further increased. The final tuning was performed again through simulations, since the bounds of the control inputs determine to a great extent whether or not a nonempty robust invariant set for the networked system actually exists. Ultimately, they had to be pushed to their current values to ensure the existence of the robust invariant set $\hat{\mathcal{A}}^{(\psi)}_{net,\infty}$.

Regarding the design of the attack detector, the partitioning indices of the matrix $\boldsymbol{F} \in \mathbb{R}^{10\times 10}$ are $s_1 = 7$ and $s_2 = 3$, the set $\boldsymbol{\mathcal{Y}}'$ was determined using $\beta = 0.2$ and $k^* = 60$, the radius λ was selected as:

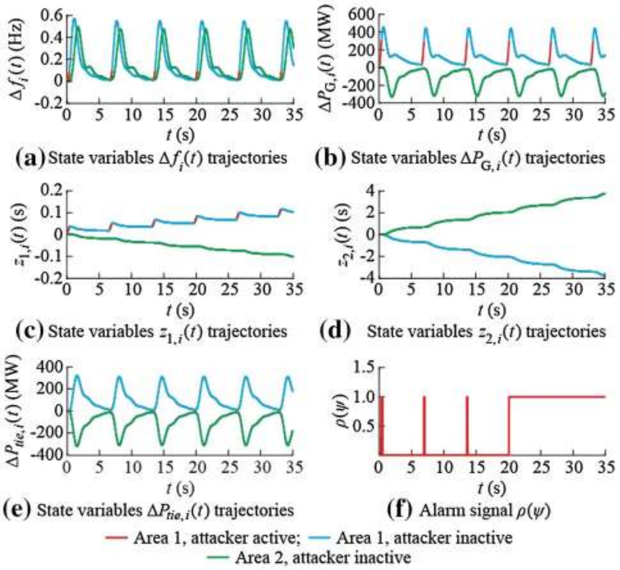
$$\lambda = 1.1\sqrt{\sum_{i=1}^{2}(\Delta P^2_{L,i,\max})}$$

and each element of the vector $\boldsymbol{\phi}'$ was taken as the mean of its boundary values. For the operations involving polyhedral constraints and optimization problems we used the MPT Toolbox 3.0 [28].

The conventions used for the depiction of the state trajectories obey the following rules. The state variables of area 1 are printed in red while an attacker is active (i.e. $\sigma_1[k] = 1$) and in blue while an attacker is inactive (i.e. $\sigma_1[k] = 0$). The state variables of area 2 are always printed in green color and whether the attacker is active (i.e. $\sigma_2[k] = 1$) or not (i.e. $\sigma_2[k] = 0$) is determined in the legend of each figure.

5.1 Individual area attack scenario

For this scenario we consider two separate cases. The first case is presented in Fig. 3 and involves the attack signals $\alpha_1 = 4.5$ Hz and $\alpha_2 = 0$, whereas the second case is presented in Fig. 4 and involves the attack signals $\alpha_1 = 2$ Hz and $\alpha_2 = 0$. In both cases, we study intermittent attack patterns and the switching bounds of $\sigma_1[k]$ are given as $\alpha_{1,\min} = 0.01$ Hz, $\alpha_{1,\max} = 0.1$ Hz and the tolerance $\delta = 10^{-3}$. We remark that the value of $\alpha_{1,\min}$ is meaningful

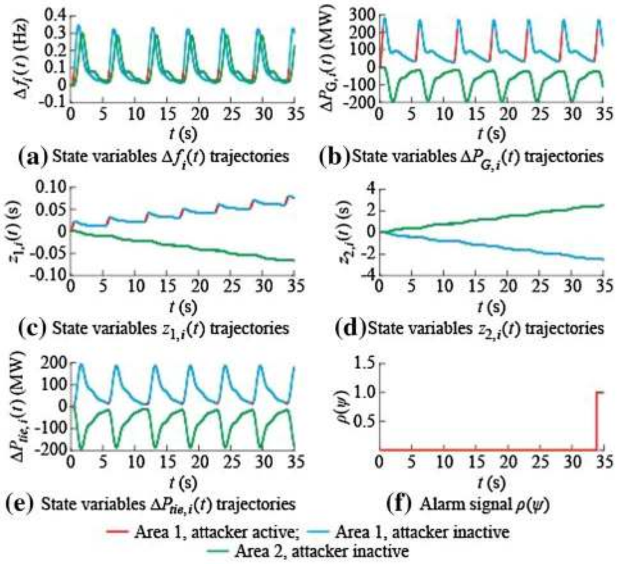


(a) State variables $\Delta f_i(t)$ trajectories (b) State variables $\Delta P_{G,i}(t)$ trajectories

(c) State variables $z_{1,i}(t)$ trajectories (d) State variables $z_{2,i}(t)$ trajectories

(e) State variables $\Delta P_{tie,i}(t)$ trajectories (f) Alarm signal $\rho(\psi)$

Fig. 3 State trajectories for $\alpha_1 = 4.5$ Hz and $\alpha_2 = 0$



(a) State variables $\Delta f_i(t)$ trajectories (b) State variables $\Delta P_{G,i}(t)$ trajectories

(c) State variables $z_{1,i}(t)$ trajectories (d) State variables $z_{2,i}(t)$ trajectories

(e) State variables $\Delta P_{tie,i}(t)$ trajectories (f) Alarm signal $\rho(\psi)$

Fig. 4 State trajectories for $\alpha_1 = 2$ Hz and $\alpha_2 = 0$

only if it is larger than the frequency measurement error ($\sim 10^{-3}$) [19].

Both cases share several common characteristics. First, the input saturation constraints are never activated, since $\Delta P_{G,i}[k] < \Delta P_{G,i,\max} = u_{i,\max}$ for all time instances. This fact is important since it ensures that if an alarm is activated, then this activation did not occur simply because the control input triggered the saturation constraints but rather because the state vector exited the robust invariant set. In addition, the intermittent attack pattern causes the state variables Δf_i to oscillate. Although these discrepancies are not significant, they inflict large persistent and non-decaying oscillations on $\Delta P_{tie,i}$, which stress the tie line and may cause the coupled generators to desynchronize. Finally, we highlight that the attacker is only activated during brief intervals. In fact, during an approximately 5 s oscillation, the switching logic causes the attacker to remain active only for approximately 0.5 s.

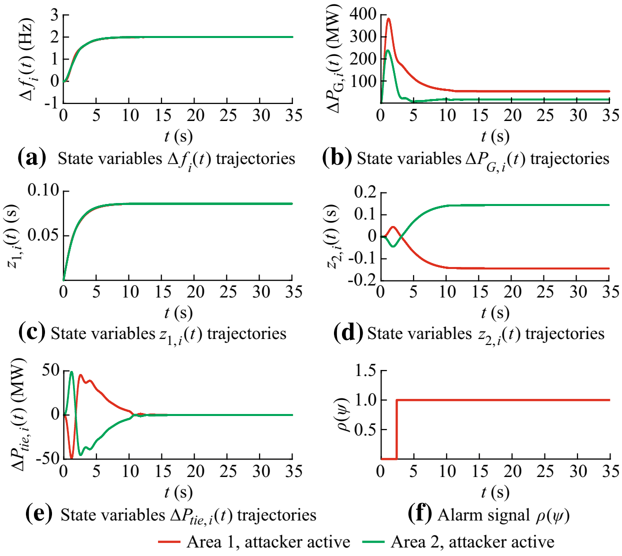
From Fig. 3, we observe that for $\alpha_1 = 4.5$ Hz the set-theoretic attack detector is regularly triggered. In this case, the detection mechanism successfully discloses the adversary on a very early stage. In contrast, Fig. 4 reveals that if the adversary decreases the value of the attack signal to $\alpha_1 = 2$ Hz, then the attack passes undetected for a longer period. Naturally, the state variables $z_{1,i}$ and $z_{2,i}$ slowly increase, starting from a larger value after every activation of the attacker. However, until the alarm is triggered, the stability of the network is already jeopardized due to the tie line power oscillations. This situation is also visible in the first scenario on Fig. 3, where after $t = 20$ s the divergence of the $z_{1,i}$ and $z_{2,i}$ causes the alarm signal to remain constantly active. We remark that a traditional residual-based attack detector would never be able to disclose this adversary, since the intermittent nature of the attack does not allow the residual quantity to obtain a steady-state constant value over time.

5.2 Multiple area attack scenario

For this scenario we consider two separate cases. The first case is presented in Fig. 5 and involves the attack signals $\alpha_1 = \alpha_2 = 2$ Hz, whereas the second case is presented in Fig. 6 and involves the attack signals $\alpha_1 = \alpha_2 = 1$ Hz. In both cases, we assume that the attack is persistent, in the sense that the switching signals $\sigma_1[k] = \sigma_2[k] = 1$ for all $k \geq 0$.

According to the Figs. 5, 6, the adversary is always able to drive the frequency deviation wherever the attack signals dictate. We highlight that the input saturation constraints are never triggered and that after the steady-state is reached, the produced powers $\Delta P_{G,i}$ satisfy both the power load demands and the increase in the network frequency. Although the persistent nature of the attack does not cause any power oscillations on the tie line, the steady-state errors in the frequency deviation are critical for the stability of the grid and may lead the power relays to trip the generators off, thus causing a blackout. Furthermore, we observe that the adversary can remain undetectable as long as the attack signals retain relatively small values. In particular, for $\alpha_1 = \alpha_2 = 2$ Hz the adversary is ultimately disclosed but for $\alpha_1 = \alpha_2 = 1$ Hz the attack is stealthy.

During the stability analysis, we established that the only way to create potentially undetectable attacks is to obtain stable responses of the state variables $z_{1,i}$ and $z_{2,i}$. Since we have explained that the only way to achieve this is to use the same attack signal on all control areas, we can

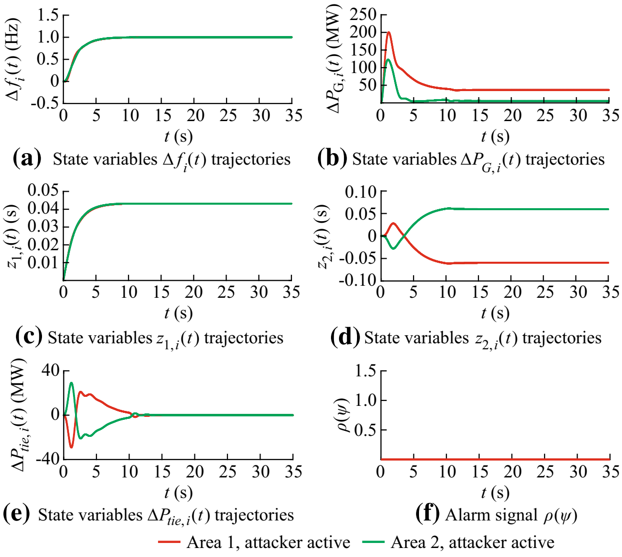


(a) State variables $\Delta f_i(t)$ trajectories (b) State variables $\Delta P_{G,i}(t)$ trajectories

(c) State variables $z_{1,i}(t)$ trajectories (d) State variables $z_{2,i}(t)$ trajectories

(e) State variables $\Delta P_{tie,i}(t)$ trajectories (f) Alarm signal $\rho(\psi)$

Fig. 5 State trajectories for $\alpha_1 = \alpha_2 = 2$ Hz



(a) State variables $\Delta f_i(t)$ trajectories (b) State variables $\Delta P_{G,i}(t)$ trajectories

(c) State variables $z_{1,i}(t)$ trajectories (d) State variables $z_{2,i}(t)$ trajectories

(e) State variables $\Delta P_{tie,i}(t)$ trajectories (f) Alarm signal $\rho(\psi)$

Fig. 6 State trajectories for $\alpha_1 = \alpha_2 = 1$ Hz

now proceed to the attack detection issue. Clearly, an attack that drives the state vector to an equilibrium that belongs to the set $\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}$ remains undetectable when $\Delta P_{L,i} = 0$ for all $i \in \mathcal{I}$. However, even when the power network is affected by nonzero disturbances there is no guarantee that an alarm will be triggered. The robustness property of the set $\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}$ ensures that if the system evolves in the absence of an attacker, then the state vector will remain exclusively inside $\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}$ for any disturbance sequences $\Delta P_{L,i}$ that respect the bounds $\Delta P_{L,i,\max}$. However, when the system is affected both by an attacker and a disturbance, it is mostly dependent on the disturbance whether an alarm will be activated or not. Since the usual disturbances $\Delta P_{L,i}$ have the form of step load changes, it may take a significantly more elaborate disturbance sequence to trigger an alarm.

Consequently, when the adversary uses the same attack signal on all control areas, the key factor that determines whether a detection will occur or not is the magnitude of the attack signal α_i. The larger it is, the greater the chance the state vector will exit the set $\hat{\boldsymbol{\mathcal{A}}}^{(\psi)}_{net,\infty}$ becomes. There is no obvious improvement of the detector from a set-theoretic point of view. We have already calculated the maximal robust invariant set with respect to the networked system dynamics in the absence of an attacker, that is when $\sigma_i = 0$ for all $i \in \mathcal{I}$. Clearly, this is the best approach, in order to ensure that any nonzero attack signal can potentially trigger an alarm.

In contrast to the previous scenario, which involved an attack only on the first control area, this case demonstrates that a set-theoretic anomaly detector may be unable to disclose an adversary as long as the attack occurs simultaneously on every control area and the attack signals have small values. However, compromising every frequency sensor in large power grids consists a highly unrealistic attack scenario. It is more reasonable to consider that only a few areas can be compromised at the same time, but in this case the set-theoretic detector will always be able to disclose a data corruption attack, due to the convex and compact nature of the robust invariant set and due to the linear divergence of the integrator variables $z_{1,i}$ and $z_{2,i}$.

We remark that a traditional residual-based attack detector may or may not be able to disclose a coordinated attack, depending on the value of the critical threshold imposed on the residue. If we take into account the unknown power load changes, then the critical threshold has to be more conservative than usual and an attack may pass undetected.

Based on [13], the critical threshold can be selected as follows. Since the load changes $\Delta P_{L,i}$ are part of the normal operation of the network, they do not pose a threat to the safety of the system. In addition, all $\Delta P_{L,i}$ are bounded signals. Hence, we can calculate the maximum admissible deviation of the estimation residue from zero, say $\delta_{r,\max}$, by considering the behavior of the system when the maximum allowed step load changes $\Delta P_{L,i,\max}$ occur in the absence of an attacker. Now, we can obtain an estimation of the critical threshold as $\delta_{r,\max}$. We stress that, in our case, we neglect the measurement and process noise of the system during the modeling process. Therefore, the threshold $\delta_{r,\max}$ should suffice, since false alarms due to the noise are not about to occur.

Let us now consider the case, when $\Delta P_{L,i} = 0$ and $\alpha_i = \alpha$ for all $i \in \mathcal{I}$. In this case, the estimation residue will ultimately converge to a nonzero constant steady-state value. However, if the attack signals are relatively small, then the steady-sate value of the residue will probably remain below the critical threshold. In other words, the attack signals will be treated by the detector as admissible load changes and the alarm will not be activated. As a matter of fact, a coordinated attack on all control areas with the same attack signal is equally difficult to detect either by a residual-based estimator or by a set-theoretic anomaly detector.

6 Conclusion

This article concerns a security enhancing method for the detection of data corruption attacks on cyber physical power systems. We present the design process of a centralized set-theoretic attack detector using a robust invariant set and apply this concept on the load frequency control loop of a networked power system. The adversarial scenarios studied in this work involve the corruption of the frequency sensor measurements using intermittent and persistent attack patters. Simulation studies on a benchmark two-area power plant demonstrate the ability of a set-theoretic attack detector to disclose an adversary even in the presence of external unknown disturbances.

Appendix A

Regarding notations, the symbols $\mathbb{O}_{m\times n}$ and $\mathbb{I}_{n\times n}$ denote the zero and the identity matrix of appropriate dimensions respectively, the set $\boldsymbol{\mathcal{B}}_p(n) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_p \leq 1\}$ represents the unit ball corresponding to the p-norm in $\mathbb{R}^n$, while all inequalities involving matrices or vectors are assumed to be componentwise.

References

[1] Andersson G, Donalek P, Farmer R et al (2005) Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. IEEE Trans Power Syst 20(4):1922–1928

[2] Mahmood A, Javaid N, Razzaq S (2015) A review of wireless communications for smart grid. Renew Sustain Energy Rev 41:248–260

[3] Wang JW, Ronga LL (2009) Cascade-based attack vulnerability on the US power grid. Saf Sci 47(10):1332–1336

[4] Deka D, Baldick R, Vishwanath S (2016) Jamming aided generalized data attacks: exposing vulnerabilities in secure estimation. In: Proceedings of the 49th Hawaii international conference on system sciences (HICSS), Hawaii, USA, 2016, pp 2556–2565

[5] Teixeira A, Sandberg H, Johansson KH (2010) Networked control systems under cyber attacks with applications to power networks. In: Proceedings of the 2010 American control conference, Baltimore, USA, 30 June–2 July 2010, pp 3690–3696

[6] Akkaya I, Liu Y, Lee EA (2015) Modeling and simulation of network aspects for distributed cyber-physical energy systems, cyber physical systems approach to smart electric power grid, Power Systems. Springer, Berlin, pp 1–23

[7] Franzé G, Tedesco F, Casavola A (2016) A leader-follower architecture for load frequency control purposes against cyber attacks in power grids—parts I and II. In: Proceedings of the 55th IEEE conference on decision and control, Las Vegas, USA, 12-14 December 2016, pp 5128–5139

[8] Esfahani PM, Vrakopoulou M, Margellos K et al (2010) A robust policy for automatic generation control cyber attack in two area power network. In: Proceedings of the 49th IEEE conference on decision and control, Atlanta, USA, 15-17 December 2010, pp 5973–5978

[9] Esfahani PM, Vrakopoulou M, Margellos K et al (2010) Cyber attack in a two-area power system: impact identification using reachability. In: Proceedings of the 2010 American control conference, Baltimore, USA, 30 June-2 July 2010, pp 962–967

[10] Teixeira A, Shames I, Sandberg H et al (2012) Revealing stealthy attacks in control systems. In: Proceedings of the 50th annual Allerton conference on communication, control, and computing (Allerton), Monticello, USA, 1–5 October 2012, pp 1806–1813

[11] Pasqualetti F, Dörfler F, Bullo F (2011) Cyber-physical attacks in power networks: models, fundamental limitations and monitor design. In: Proceedings of the 50th IEEE conference on decision and control and European control conference, Orlando, USA, 12-15 December 2011, pp 2195–2201

[12] Pasqualetti F, Dörfler F, Bullo F (2013) Attack detection and identification in cyber-physical systems. IEEE Trans Autom Control 58(11):2715–2729

[13] Teixeira A, Pérez D, Sandberg H et al (2012) Attack models and scenarios for networked control systems. In: Proceedings of the 1st international conference on high confidence networked systems, Beijing, China, 2012, pp 55–64

[14] Umsonst D, Sandberg H, Cárdenas AA (2017) Security analysis of control system anomaly detectors. In: Proceedings of the 2017 American control conference, Seattle, USA, 24–26 May 2017, pp 5500–5506

[15] Kontouras E, Tzes A, Dritsas L (2017) Cyber-attack on a power plant using bias injected measurements. In: Proceedings of the 2017 American control conference, Seattle, USA, 24–26 May 2017, pp 5507–5512

[16] Kontouras E, Tzes A, Dritsas L (2017) Impact analysis of a bias injection cyber-attack on a power plant. In: Proceedings of the 20th world congress of the international federation of automatic control, Toulouse, France, 2017, pp 11586–11591

[17] Kontouras E, Tzes A, Dritsas L (2018) Set-induced anomaly detectors for networked power systems under bias injection cyber-attacks. In: Proceedings of the European control conference, Limassol, Cyprus, 2018, pp 2472–2475

[18] Kontouras E, Tzes A, Dritsas L (2018) Set-theoretic detection of bias injection cyber-attacks on networked power systems. In: Proceedings of the 2018 American control conference, Milwaukee, USA,27-29 June 2018, pp 165-170

[19] Elgerd OI (1982) Electric energy systems theory: an introduction, 2nd edn. McGraw-Hill, New York

[20] Kundur P (1994) Power system stability and control, 1st edn. McGraw-Hill, New York

[21] Suehiro T, Namerikawa T (2012) Decentralized control of smart grid by using overlapping information. In: Proceedings of the SICE annual conference, Akita, Japan, 20–23 August 2012, pp 125–130

[22] Elgerd OI, Fosha CE Jr (1970) Optimum megawatt-frequency control of multiarea electric energy systems. IEEE Trans Power Appar Syst 89(4):556–563

[23] Blanchini F (1999) Set invariance in control. Automatica 35:1747–1767

[24] Blanchini F, Miani S (2008) Set-theoretic methods in control. Birkhäuser, Basel

[25] Kolmanovsky I, Gilbert EG (1998) Theory and computation of disturbance invariant sets for discrete-time linear systems. Math Probl Eng 4:317–367

[26] Gilbert EG, Kolmanovsky I, Tan KT (1995) Discrete-time reference governors and the nonlinear control of systems with state and control constraints. Int J Robust Nonlinear Control 5(5):487–504

[27] Liberzon D (2003) Switching in systems and control, systems & control: foundations & applications. Birkhäuser, Basel

[28] Herceg M, Kvasnica M, Jones CN et al (2013) Multi-parametric toolbox 3.0. In: Proceedings of the European control conference, Zurich, Switzerland, 17–19 July 2013, pp 502–510

Efstathios KONTOURAS received his Diploma in 2013 from the Department of Electrical and Computer Engineering at the University of Patras (UPAT) in Greece. He is currently working toward his Ph.D. at the same Department in the field of adversarial control tactics in cyber physical systems. His main research interests include power networks, design of cyber attacks and anomaly detectors, optimization theory, robust control and set-theoretic methods.

Anthony TZES received his Diploma in 1985 at University of Patras, Greece and received his Ph.D. (1990) in electrical engineering from the Ohio State University, U.S.A. His research interests are in the field of collaborative control of mobile robots and control of networked cyber physical systems. He has been employed at New York University Tandon School of Engineering and University of Patras. Since 2017, he is a professor with the Engineering Division of the New York University Abu Dhabi, in United Arab Emirates. He has more than 85(250) journal(conference) articles and has been in the organization committees (chairman, program chairman and other positions) of various international conferences and an associate editor in several journals.

Leonidas DRITSAS serves as tenured Lecturer at the Department of Electrical & Electronic Engineering Educators, School of Pedagogical and Technological Education (ASPETE), Athens. Greece. He holds a Diploma Degree in Electrical and Computer Engineering from the University of Patras, Greece, a M.Sc. degree in Electrical and Computer Engineering from Drexel University, USA and a PhD degree from the University of Patras (E.C.E Dept.), Greece. His research interests and most of his recent publications lie in the areas of Robust Networked Control and Time-Delayed Systems with applications to Cyber Physical Systems. He has authored/co-authored ten journal papers, forty conference papers and three Book Chapters. Dr. Dritsas has over thirty years of professional experience in Information Technologies, Industrial Automation and Project Management.