energies

MDPI

Article

Hybrid Detection of Intermittent Cyber-Attacks in Networked Power Systems

Efstathios Kontouras [1], Anthony Tzes [2,*] and Leonidas Dritsas [3]

1 Electrical & Computer Engineering Department, University of Patras, Rio 26500, Greece; kontouras@ece.upatras.gr
2 Electrical & Computer Engineering Program, New York University Abu Dhabi, Abu Dhabi, P.O. Box 129188, UAE; anthony.tzes@nyu.edu
3 Department of Electrical & Electronic Engineering Educators, School of Pedagogical & Technological Education, ASPETE, Athens 14121, Greece; dritsas@aspete.gr
* Correspondence: anthony.tzes@nyu.edu

Received: 30 October 2019; Accepted: 27 November 2019; Published: 4 December 2019

Abstract: This article addresses the concept of a compound attack detection mechanism, that links estimation-based and set-theoretic methods, and is mainly focused on the disclosure of intermittent data corruption cyber-attacks. The detection mechanism is developed as a security enhancing tool for the load-frequency control loop of a networked power system that consists of several interconnected control areas. The dynamics of the power network are derived in observable form in the discrete-time domain, considering that an adversary corrupts the frequency measurements of certain control areas by means of a bias injection cyber-attack. Simulations indicate that an estimation-based detector is unable to discern an intermittent attack, especially when the latter one occurs at the same time as changes in the power load. The detector can be improved by exploiting the safe operation constraints imposed on the state variables of the system. It is shown that the disclosure of intermittent data corruption cyber-attacks in the presence of unknown power load changes is guaranteed only when the estimation-based detector is combined with its set-theoretic counterpart. To this end, a robust invariant set for the networked power system is computed and an alarm is triggered whenever the state vector exits this set. Simulations indicate that the above detectors can operate jointly in terms of a hybrid scheme, which enhances their detection capabilities.

Keywords: cyber-attacks; load-frequency control; power systems; set-theoretic methods; state estimation

1. Introduction

Modern interconnected power systems often transmit their data through unprotected wireless channels [1]. Vulnerabilities of this kind can be exploited by cyber-attackers, that aim to disrupt the normal operation of the power network [2]. Common scenarios concern replay attacks [3], where the adversary records measurements for a certain period of time and then resends them, therefore replacing the real ones, denial of service attacks [4–6], where the adversary jams channels that transmit control commands or sensor measurements, and data corruption cyber-attacks, where the adversary injects system data with biased attack signals [7–9]. Stealthy attacks, that cause system malfunctions, while remaining undetected in the process, have also been extensively studied in [10–13].

Security-enhancing techniques should prevent, mitigate and tolerate cyber-attacks on electric power grids [14]. In [15,16], a generic framework is developed for the detection and identification of different types of cyber-attacks in networked power systems based on state estimation and monitoring methods. In [17], the number of the distributed estimators is reduced, while maintaining the coverage of the entire network, whereas in [18], the authors address the optimal placing of estimation devices in

the power network in order to maximize their utility and increase the security of the system in the case of data corruption cyber-attacks. Finally, in [19,20], the authors develop attack detectors using control-theoretic methods, that involve state estimation through Kalman filtering.

The load-frequency control loop of a networked power system is particularly interesting from a security point of view. This loop depends on an extended digital layer that connects sensors, actuators and physical entities through generally unprotected channels and is designed to operate without human intervention. In addition, this loop is extremely sensitive in the discrepancies of the electrical frequency. Indeed, frequency fluctuations that are caused either by a power load change or by an attack in one control area, affect all other areas, thus jeopardizing the stability of the overall network. Attacks on this loop were studied in [21], where the authors quantify the impact of a cyber-attack on the automatic generation control unit of a two-area power system through a reachability analysis. In this scenario, an attacker replaces the control signal of one control area and causes the abnormal behavior of the overall system. In [22], the authors study a similar scenario and exploit Monte Carlo optimization in order to design an attack signal, which remains robust with respect to the potential parametric uncertainties of the system model.

The security of the load-frequency control loop is addressed in [23], where the authors develop an overlapping networked control architecture, which is then recast into leader-follower configurations, with a time-varying hierarchy. Set-theoretic concepts come also into play, by means of a reachability analysis, that allows for the calculation of one-step ahead controllable sets, based on the input and state constraints of the system. In [24], this framework is tested in the case of data loss cyber-attacks and is used to successfully isolate the control areas under attack.

In this work, we examine a networked power system, where each individual control area is subject to unknown power load changes due to the demand of the consumers. We assume that an adversary can corrupt the frequency measurements, which are transmitted from the sensors to the automatic generation control units, using intermittent bias injected attack signals. The proposed attack scenario is realistic and highly effective. The adversary does not need any knowledge about the model of the system or the detection mechanism and can cause discrepancies on the electrical frequency, which lead to large fluctuations on the tie line power exchanged between the neighboring control areas. These fluctuations are dangerous since they stress the tie line to its thermal limit and can cause the coupled generators to desynchronize. The primary objective is to establish a mathematical link between the classical estimation-based attack detectors [25] and their recently introduced set-theoretic counterparts, proposed in [26,27]. The idea is to combine the best traits of the two methods and develop a hybrid detector, which performs better than each detector alone in the case of intermittent attacks.

An estimation-based detector decides the existence of an attacker based on the value of the estimation residue. If this residue obtains a steady-state value larger than a critical threshold, then an alarm must be activated [25]. It is shown that an estimation-based detector is unable to disclose an intermittent attack, in a guaranteed manner, especially when this attack occurs at the same time with an unknown power load change. This happens because an intermittent attack forces the state vector to oscillate. In this case, the residue is unable to obtain a steady-state value and therefore the alarm signal cannot be activated. An estimation-based detector can also hint the existence of an attacker, when the discrepancies of the estimation residue during the transients are highly intense. However, in this case it is difficult to discern an attack from a load change and false alarms may also occur.

On the other hand, a set-theoretic attack detector relies on the extraction of suitable robust invariant sets, which stem from safety considerations and are used as sets of alarm constraints [26,27]. A set is said to be robust invariant with respect to the dynamics of a system, when the state trajectories emanating from every initial condition that belongs to this set remain within the same set for every future time instant and for every admissible disturbance sequence [28]. The concept of a set-theoretic detector is to activate an alarm signal whenever the state vector exits the robust invariant set, either during the transient response or during the steady-state phase of the system. It is shown that the only way to quantify the discrepancies of the estimation residue is to exploit the safety constraints of the

state variables and resort to set-theoretic methods. In addition, we prove that we are mathematically inclined to determine the robust invariant set based not on the dynamics of the estimator but based on the dynamics of the system itself, inevitably resorting to the detectors developed in [26,27].

Our analysis demonstrates that each detection method complements the other. The two methods are integrated into one via a three-modal system operation, where the individual modes indicate normal operation, alert state or alarm condition, according to the value of the estimation residue and the state vector at every time instant. Thus, emerges the notion of hybrid detection. We highlight that the detection mechanisms that we develop in this article aims to disclose a particular type of cyber-attack. Specifically, the objective of the proposed detectors is to improve the security of the load-frequency control loop of a networked power system against data corruption cyber-attacks. For evaluation purposes, we chose to examine the case where the adversary corrupts the frequency sensor measurements with bias injected attack signals.

This article is organized as follows. In Section 2, we establish the model of a typical power network subject to an intermittent data corruption cyber-attack. In Section 3, we develop an estimation-based detector and we also determine the corresponding steady-state critical threshold. In Section 4, we review the basic aspects of the set-theoretic detectors, indicating the need for a hybrid detection scheme. Finally, in Section 5, we present simulation results for the test case of a two-area power system and in Section 6 we provide some concluding remarks.

Regarding notations, the symbols $\mathbb{O}_{m\times n}$ and $\mathbb{I}_{n\times n}$ are used for the zero and the identity matrix of appropriate dimensions respectively, while all inequalities involving matrices or vectors are assumed to be componentwise.

2. System Description

The algorithms used for the calculation of the robust invariant sets, and thus for the design of the set-theoretic detectors, involve discrete-time systems [26,27]. In this work, we aim to develop an estimation-based detector and then combine it with a set-theoretic one. We adopt a common modeling basis for both detectors, by remaining consistent with the discrete-time approach introduced in [26,27]. However, our original framework requires certain modifications in order to yield an observable state space model of the power grid.

2.1. Interconnected Control Area Model

Let us consider the generic interconnected control area model depicted in Figure 1. According to [29,30] a state space model that describes the evolution of the system in the continuous-time domain is

$$S_i^c : \dot{x}_i(t) = A_{c,i}x_i(t) + B_{c,i}u_{c,i}(t) + D_{c,i}\Delta P_{L,i}(t) + E_{c,i}\Delta P_{tie,i}(t), \quad x_i(0) = x_{i,0},$$

$$y_i(t) = C_i x_i(t),$$

where the index $i \in \mathcal{I} = \{1, 2, ..., N\}$ denotes the i-th area, and $t \in \mathbb{R}_+$ is the time variable.

The state vector $x_i(t) \in \mathbb{R}^2$ is defined as

$$x_i(t) = \begin{bmatrix} \Delta f_i(t) & \Delta P_{G,i}(t) \end{bmatrix}^\top,$$

where $\Delta f_i(t)$ is the deviation of the electrical frequency and $\Delta P_{G,i}(t)$ is the deviation of the mechanical power, which is produced in the output of the turbine. In order to simplify our analysis, we adopt the common assumption that the mechanical power provided to the rotor shaft is equal to the electrical power produced by the generator. We highlight that the system output $y_i(t) \in \mathbb{R}$ coincides with Δf_i, therefore we have $C_i = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $y_i(t) = \Delta f_i(t)$.

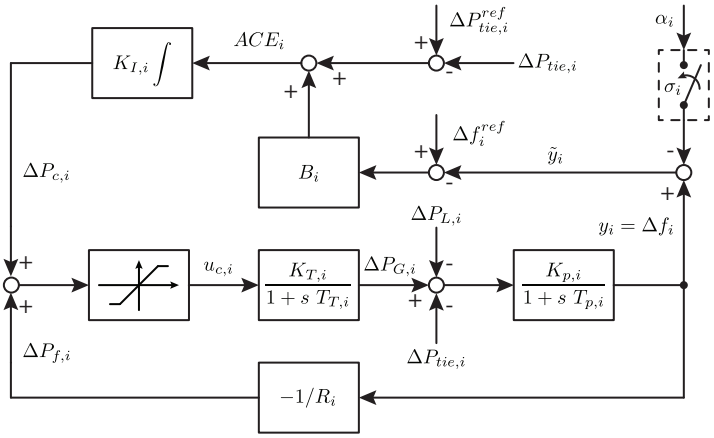


Figure 1. Load-frequency control loop of a control area under a bias injection cyber-attack on the frequency measurements. The speed governor dynamics are omitted. Model used for estimator design.

The system is driven by two individual control actions, namely the primary frequency control $\Delta P_{f,i}(t)$ and the automatic generation control $\Delta P_{c,i}(t)$. The control input $u_{c,i}(t) \in \mathbb{R}$ is defined as

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

and is subject to a saturation hard constraint of the form

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

where the boundary $u_{i,\max} \in \mathbb{R}^*_+$.

The signal $\Delta P_{L,i}(t) \in \mathbb{R}$ encapsulates an unknown but bounded disturbance, that represents in aggregated terms the deviation of the power load, due to the time-varying demand of the consumers, and obeys the constraint

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

where the boundary $\Delta P_{L,i,\max} \in \mathbb{R}^*_+$.

The signal $\Delta P_{tie,i}(t) \in \mathbb{R}$ encapsulates the deviation of the cumulative electrical power exchanged between the i-th control area and the rest of the network through the tie line, whenever a power load change takes place. The dynamics of this signal will be presented later.

The matrix $A_{c,i} \in \mathbb{R}^{2\times 2}$ is defined as

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

and the matrices $B_{c,i}, D_{c,i}, E_{c,i} \in \mathbb{R}^{2\times 1}$ are defined as

$$B_{c,i} = \begin{bmatrix} 0 \\ K_{T,i}/T_{T,i} \end{bmatrix}, \quad D_{c,i} = E_{c,i} = \begin{bmatrix} -K_{p,i}/T_{p,i} \\ 0 \end{bmatrix}.$$

If the balance between the produced and the consumed power is not ensured after each power load change, then the ensuing frequency and tie line power deviations can damage the synchronous generators. The regulation of these quantities is performed by two control loops. The first loop closes over the speed governor, which performs the primary frequency control action and is implemented as a proportional control law 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 second loop closes over the automatic generation control unit and eliminates any remaining steady-state frequency errors through the integral control law [31]

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

The quantity $ACE_i(t)$ represents the i-th area control error and is given as

$$ACE_i(t) = \left(\Delta P_{tie,i}^{ref} - \Delta P_{tie,i}(t)\right) + B_i \left(\Delta f_i^{ref} - \tilde{y}_i(t)\right). \tag{3}$$

The parameters $\Delta P_{tie,i}^{ref} = \Delta f_i^{ref} = 0$ stand for the tie line power deviation and the frequency deviation reference signals respectively, whereas the gain $B_i = 1/R_i$. The signal $\tilde{y}_i(t)$ is defined as

$$\tilde{y}_i(t) = y_i(t) - \alpha_i \sigma_i(t), \tag{4}$$

where $\alpha_i \in \mathbb{R}$ is the attack signal that falsifies the frequency measurements, which are forwarded to the automatic generation control unit, and $\sigma_i : \mathbb{R}_+ \to \{0, 1\}$ is the switching logic that drives the i-th attack pattern. An explicit mathematical expression of $\sigma_i[k]$, describing an intermittent attack, will be derived later. The control $\Delta P_{f,i}$ is unaffected by the cyber-attack, since it is mechanically implemented.

In order to extract an equivalent discrete-time model of each individual control area, first we have to calculate the eigenvalues of all the matrices $A_{c,i}$ and then select a sampling frequency f_s at least ten times greater than the frequency that corresponds to the fastest eigenvalue of the grid. We apply the zero-order hold method and we obtain the discrete-time state space representation of a generic interconnected control area as

$$S_i^d : x_i[k+1] = A_{d,i} x_i[k] + B_{d,i} u_{d,i}[k] + D_{d,i} \Delta P_{L,i}[k] + E_{d,i} \Delta P_{tie,i}[k], \quad x_i[0] = x_{i,0},$$

$$y_i[k] = C_i x_i[k],$$

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

The next step is to define the controller dynamics of each control area and extract an equivalent discrete-time representation. Contrary to our previous works [26,27], here, we use not two, but one additional state variable to describe the integral action of the automatic generation control unit. This modification was deemed necessary, because the networked power system model, as obtained through our framework in [26,27], lacks observability guarantees. In particular, the dependence on two integral variables per control area renders the networked power system model structurally unobservable and, therefore, not suited for the design of an estimation-based attack detector. The simplest way to resolve this problem is to consider only one accumulated time error

$$z_i(t) = \frac{1}{|P_{tie,i}^{\circ}| + B_i f^{\circ}} \int_0^t ACE_i(\tau) d\tau \tag{5}$$

as the extra state variable augmenting the system due to the integrator of the Equation (2). The parameter f° denotes the nominal network frequency, whereas the parameter $P_{tie,i}^{\circ}$ denotes the nominal power, that the i-th control area is scheduled to exchange with the rest of the grid via the connecting tie line. We highlight that the tie line power is assumed to be positive when it flows from the i-th control area towards the rest of the network. Based on the Equations (1)–(5), we can show that

$$u_{d,i}[k] = K'_{I,i} z_i[k] - \frac{1}{R_i} y_i[k], \tag{6}$$

where the normalized gain $K'_{I,i}$ is defined as

$$K'_{I,i} = K_{I,i}\left(|P^{\circ}_{tie,i}| + B_i f^{\circ}\right) \tag{7}$$

and the state variable $z_i[k]$ satisfies the equation

$$z_i[k+1] = z_i[k] + \frac{1}{f_s\left(|P^{\circ}_{tie,i}| + B_i f^{\circ}\right)} ACE_i[k], \quad z_i[0] = 0. \tag{8}$$

Now, the discrete-time closed-loop model of a generic interconnected control area, subject to a data corruption cyber-attack, can be written in augmented form as

$$S_i^{cl} : \xi_i[k+1] = A_{cl,i}\xi_i[k] + \alpha_i B_{cl,i}\sigma_i[k] + D_{cl,i}\Delta P_{L,i}[k] + E_{cl,i}\Delta P_{tie,i}[k], \quad \xi_i[0] = \xi_0, \tag{9}$$

$$y_i[k] = C_{cl,i}\xi_i[k], \tag{10}$$

where the augmented state vector $\xi_i[k] \in \mathbb{R}^3$ is given as

$$\xi_i[k] = \begin{bmatrix} \Delta f_i[k] & \Delta P_{G,i}[k] & z_i[k] \end{bmatrix}^{\top}, \tag{11}$$

the matrix $A_{cl,i} \in \mathbb{R}^{3\times 3}$ is given as

$$A_{cl,i} = \begin{bmatrix} A_{d,i} - (1/R_i)\, B_{d,i}C_i & B_{d,i}K'_{I,i} \\ -1/\left(f_s\left(|P^{\circ}_{tie,i}| + B_i f^{\circ}\right)\right) B_i C_i & 1 \end{bmatrix}, \tag{12}$$

while the matrices $B_{cl,i}, D_{cl,i}, E_{cl,i} \in \mathbb{R}^{3\times 1}$ and $C_{cl,i} \in \mathbb{R}^{1\times 3}$ are given as

$$B_{cl,i} = \begin{bmatrix} \mathbb{O}_{1\times 2} & 1/\left(f_s\left(|P^{\circ}_{tie,i}| + B_i f^{\circ}\right)\right) B_i \end{bmatrix}^{\top}, \tag{13}$$

$$D_{cl,i} = \begin{bmatrix} D_{d,i}^{\top} & 0 \end{bmatrix}^{\top}, \tag{14}$$

$$E_{cl,i} = \begin{bmatrix} E_{d,i}^{\top} & -1/\left(f_s\left(|P^{\circ}_{tie,i}| + B_i f^{\circ}\right)\right) \end{bmatrix}^{\top}, \tag{15}$$

$$C_{cl,i} = \begin{bmatrix} C_i & 0 \end{bmatrix}. \tag{16}$$

The network representation provided by the Equations (9)–(16) is asymptotically stable and will suffice for the needs of the estimators. However, the use of two distinct integral state variables to describe the integral action of the automatic generation control unit remains essential for the design of the set-theoretic detectors. Similarly to the Equation (5), we can define the accumulated time errors

$$z_{1,i}(t) = \frac{1}{f^{\circ}} \int_0^t \left(\Delta f_i^{ref} - \tilde{y}_i(\tau)\right) d\tau, \tag{17}$$

$$z_{2,i}(t) = \frac{1}{|P^{\circ}_{tie,i}|} \int_0^t \left(\Delta P_{tie,i}^{ref} - \Delta P_{tie,i}(\tau)\right) d\tau. \tag{18}$$

According to the stability analysis provided in [26,27], the state variables given by the Equations (17) and (18) play a critical role in the detection of an adversary. The key idea is that the variables $z_{1,i}$, $z_{2,i}$ will demonstrate a linearly unstable behavior in the presence of an attacker, unless the adversary is able to access and corrupt the frequency measurements of every control area, using the same attack signals $\alpha_i = \alpha$ for all $i \in \mathcal{I}$. The latter one is a less realistic scenario, since it requires an excess amount of resources for its implementation. At any rate, the unstable behavior of the

variables $z_{1,i}$, $z_{2,i}$ along with the convex and compact nature of the robust invariant sets, that are used by the set-theoretic detectors, guarantee the detection of an adversary, regardless of the disturbance sequences affecting the power network and regardless of the magnitude of the attack signals affecting the individual control areas. We highlight that the state variables $z_{1,i}$, $z_{2,i}$ are virtual, in the sense that they do not represent natural quantities. This implies that, their unstable behavior during an attack can be exploited by the set-theoretic detectors, without putting at risk the safe operation of the networked power system.

To better understand how the existence of an attacker triggers the unstable response of the state variables $z_{1,i}$, $z_{2,i}$ we consider the following example. Since we always have $\Delta P_{tie,i}^{ref} = \Delta f_i^{ref} = 0$, a constant attack signal $\alpha_i \neq 0$ for some $i \in \mathcal{I}$ forces the corresponding control areas to alter their frequency deviation reference signals from 0 to α_i. In this case, the Equation (17) becomes

$$z_{1,i}(t) = \frac{1}{f^{\circ}} \int_0^t \left(\alpha_i - y_i(\tau)\right) d\tau. \tag{19}$$

However, the frequency deviation of the overall power network must always converge to a constant steady-state value, which is globally identical for every control area. As long as the adversary does not affect every control area, the global steady-state value of the electrical frequency will belong strictly to the range $(\min_{i\in\mathcal{I}}\{\alpha_i\}, \max_{i\in\mathcal{I}}\{\alpha_i\})$. In this case, the quantity under the integral of the Equation (19) is forced to obtain a steady-state constant nonzero value. This in turn implies that the state variables $z_{1,i}$ are forced to diverge linearly towards infinity, for as long as the attacker remains active. It is evident that the adversary will be disclosed the moment when the state vector will exit the convex and compact robust invariant set.

Following the same principles that we applied for the extraction of the Equations (6)–(8), we can show that the control input of the system can be written as

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

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

$$K'_{I_1,i} = K_{I,i} B_i f^{\circ}, \quad K'_{I_2,i} = K_{I,i} |P^{\circ}_{tie,i}|$$

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

$$z_{1,i}[k+1] = z_{1,i}[k] - \frac{1}{f_s f^{\circ}} \tilde{y}_i[k], \quad z_{1,i}[0] = 0,$$

$$z_{2,i}[k+1] = z_{2,i}[k] - \frac{1}{f_s |P^{\circ}_{tie,i}|} \Delta P_{tie,i}[k], \quad z_{2,i}[0] = 0.$$

The Equations (9) and (10), that were previously used to describe the discrete-time model of a generic interconnected control area, are still valid, but they require a few modifications. The augmented state vector $\xi_i[k] \in \mathbb{R}^4$ is now given as

$$\xi_i[k] = \begin{bmatrix} \Delta f_i[k] & \Delta P_{G,i}[k] & z_{1,i}[k] & z_{2,i}[k] \end{bmatrix}^{\top}, \tag{20}$$

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

$$A_{cl,i} = \begin{bmatrix} A_{d,i} - (1/R_i) B_{d,i} C_i & B_{d,i} K'_{I_1,i} & B_{d,i} K'_{I_2,i} \\ -1/(f_s f^{\circ}) C_i & 1 & 0 \\ \mathbb{O}_{1\times 2} & 0 & 1 \end{bmatrix}, \tag{21}$$

while the matrices $B_{cl,i}, D_{cl,i}, E_{cl,i} \in \mathbb{R}^{4\times 1}$ and $C_{cl,i} \in \mathbb{R}^{1\times 4}$ are given as

$$B_{cl,i} = \left[\mathbb{O}_{1\times 2} \quad 1/\left(f_s f^{\circ}\right) \quad 0\right]^{\top}, \tag{22}$$

$$D_{cl,i} = \left[D_{d,i}^{\top} \quad 0 \quad 0\right]^{\top}, \tag{23}$$

$$E_{cl,i} = \left[E_{d,i}^{\top} \quad 0 \quad -1/\left(f_s |P_{tie,i}^{\circ}|\right)\right]^{\top}, \tag{24}$$

$$C_{cl,i} = \left[C_i \quad 0 \quad 0\right]. \tag{25}$$

Let us now address the design of the switching signals σ_i, that are used to drive the attack signals α_i. The purpose of an intermittent attack is to force the state variables to oscillate. The main objective is to create large swingings of the power exchanged through the connecting tie lines in order to endanger the stability of the grid. Such an attack pattern can be developed by employing a hysteresis-based switching logic [32] of the form

$$\sigma_i[k] = \begin{cases} 0, & \text{if} \quad |\tilde{y}_i[k]| > \bar{\alpha}_{i,\max} \quad \text{and} \quad \sigma_i[k-1] = 1 \\ 1, & \text{if} \quad |y_i[k]| < \bar{\alpha}_{i,\min} \quad \text{and} \quad \sigma_i[k-1] = 0 \\ \sigma_i[k-1], & \text{otherwise} \end{cases}, \tag{26}$$

where $\bar{\alpha}_{i,\max} = \alpha_{i,\max} - \delta$ and $\bar{\alpha}_{i,\min} = \alpha_{i,\min} + \delta$ are the hysteresis bounds, that dictate the switching surfaces, and $\delta \in \mathbb{R}_{+}^{*}$ is the tolerance factor ensuring that a switching can occur when the state vector is located strictly inside the frequency zone $y_i \in \left[\Delta f_{i,\min}^{\alpha}, \Delta f_{i,\max}^{\alpha}\right] = [\alpha_{i,\min}, \alpha_{i,\max}]$. We remark that the adversary requires the full knowledge advantage regarding the frequency sensor measurements y_i in order to be able to implement the switching signal presented in the Equation (26).

2.2. Tie Line Model

Whenever a power load change takes place, the power flow of every tie line deviates from its prespecified value, according to the linearized [29,30] dynamical equation

$$\Delta \dot{P}_{tie,i}(t) = \sum_{j=1}^{N} \left(2\pi T_{ij} \left(\Delta f_i(t) - \Delta f_j(t)\right)\right),$$

where T_{ij} is the synchronization coefficient between the areas i and j and $\Delta P_{tie,i}$ is the aggregated electrical power exchanged between the i-th area and every other area of the grid that remains connected with it. By definition we have $T_{ij} = T_{ji}$ for all $i, j \in \mathcal{I}$ and if two areas i, j are not connected with each other, then we have $T_{ij} = 0$.

In order to extract a discrete-time equivalent model for the tie line, we use the global sampling frequency f_s and the zero-order hold method and we obtain the equation

$$\Delta P_{tie,i}[k+1] = \Delta P_{tie,i}[k] + T_s \sum_{j=1}^{N} \left(2\pi T_{ij} \left(\Delta f_i[k] - \Delta f_j[k]\right)\right), \tag{27}$$

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

2.3. Network Model

If we calculate the discrete-time models of all control areas along with the discrete-time models of their tie lines, then we obtain a discrete-time representation of the entire networked power system as

$$S_{net} : x_{net}[k+1] = A_{net} x_{net}[k] + B_{net}[k] + D_{net} \Delta P_{L,net}[k], \quad x_{net}[0] = x_{net,0}, \tag{28}$$

$$y_{net}[k] = C_{net}x_{net}[k], \tag{29}$$

where n is the number of state variables per control area. We set $n = 3$ and we depend on the Equations (11)–(16) in order to extract the networked system model used for the design of the estimation-based detectors. On the other hand, we set $n = 4$ and we depend on the Equations (20)–(25) in order to extract the networked system model used for the design of the set-theoretic detectors.

In both cases, the state vector $x_{net} \in \mathbb{R}^{(n+1)N}$ is defined as

$$x_{net}[k] = \begin{bmatrix} \xi_1^\top[k] & \xi_2^\top[k] & \dots & \xi_N^\top[k] & \Delta P_{tie,1}[k] & \Delta P_{tie,2}[k] & \dots & \Delta P_{tie,N}[k] \end{bmatrix}^\top \tag{30}$$

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

$$\Delta P_{L,net}[k] = \begin{bmatrix} \Delta P_{L,1}[k] & \Delta P_{L,2}[k] & \dots & \Delta P_{L,N}[k] \end{bmatrix}^\top. \tag{31}$$

The matrix $A_{net} \in \mathbb{R}^{(n+1)N \times (n+1)N}$ is defined as

$$A_{net} = \begin{bmatrix} A_{net,11} & A_{net,12} \\ A_{net,21} & A_{net,22} \end{bmatrix}, \tag{32}$$

where $A_{net,11} \in \mathbb{R}^{nN \times nN}$ and $A_{net,12} \in \mathbb{R}^{nN \times N}$ are given as

$$A_{net,11} = \begin{bmatrix} A_{cl,1} & \mathbb{O}_{n\times n} & \dots & \mathbb{O}_{n\times n} \\ \mathbb{O}_{n\times n} & A_{cl,2} & \dots & \mathbb{O}_{n\times n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{n\times n} & \mathbb{O}_{n\times n} & \dots & A_{cl,N} \end{bmatrix}, \quad A_{net,12} = \begin{bmatrix} E_{cl,1} & \mathbb{O}_{n\times 1} & \dots & \mathbb{O}_{n\times 1} \\ \mathbb{O}_{n\times 1} & E_{cl,2} & \dots & \mathbb{O}_{n\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{n\times 1} & \mathbb{O}_{n\times 1} & \dots & E_{cl,N} \end{bmatrix}, \tag{33}$$

whereas $A_{net,21} \in \mathbb{R}^{N \times nN}$ and $A_{net,22} \in \mathbb{R}^{N \times N}$ are given as

$$A_{net,21} = \begin{bmatrix} L_{11} & L_{12} & \dots & L_{1N} \\ L_{21} & L_{22} & \dots & L_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ L_{N1} & L_{N2} & \dots & L_{NN} \end{bmatrix}, \quad A_{net,22} = \mathbb{I}_{N\times N} \tag{34}$$

and the vector elements $L_{ij} \in \mathbb{R}^{1\times n}$ are given as

$$L_{ij} = \begin{cases} \begin{bmatrix} \sum_{j=1}^{N} \left(2\pi T_{ij} T_s\right) & \mathbb{O}_{1\times(n-1)} \end{bmatrix}, & \text{if} \quad i = j \\ \\ \begin{bmatrix} -2\pi T_{ij} T_s & \mathbb{O}_{1\times(n-1)} \end{bmatrix}, & \text{if} \quad i \neq j \end{cases}. \tag{35}$$

Finally, we define the matrices $B_{net} \in \mathbb{R}^{(n+1)N\times 1}$ and $D_{net} \in \mathbb{R}^{(n+1)N\times N}$ as

$$B_{net}[k] = \begin{bmatrix} \alpha_1 B_{cl,1}\sigma_1[k] \\ \alpha_2 B_{cl,2}\sigma_2[k] \\ \vdots \\ \alpha_N B_{cl,N}\sigma_N[k] \\ \mathbb{O}_{N\times 1} \end{bmatrix}, \quad D_{net} = \begin{bmatrix} D_{net,1} \\ \mathbb{O}_{N\times N} \end{bmatrix}, \quad D_{net,1} = \begin{bmatrix} D_{cl,1} & \mathbb{O}_{n\times 1} & \dots & \mathbb{O}_{n\times 1} \\ \mathbb{O}_{n\times 1} & D_{cl,2} & \dots & \mathbb{O}_{n\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{n\times 1} & \mathbb{O}_{n\times 1} & \dots & D_{cl,N} \end{bmatrix} \tag{36}$$

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

$$C_{net} = \begin{bmatrix} C_{cl,1} & \mathbb{O}_{1\times n} & \cdots & \mathbb{O}_{1\times n} & \mathbb{O}_{1\times N} \\ \mathbb{O}_{1\times n} & C_{cl,2} & \cdots & \mathbb{O}_{1\times n} & \mathbb{O}_{1\times N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbb{O}_{1\times n} & \mathbb{O}_{1\times n} & \cdots & C_{cl,N} & \mathbb{O}_{1\times N} \\ \mathbb{O}_{N\times n} & \mathbb{O}_{N\times n} & \cdots & \mathbb{O}_{N\times n} & \mathbb{I}_{N\times N} \end{bmatrix}. \tag{37}$$

For further details on the extraction of the above block matrices and their connection with the physical quantities of the power system, the reader is referred to [26,27].

3. Estimation-Based Detector Design

Conventional attack detectors assume the form of state estimators. The concept is simple, easy to generalize and can be applied both in a centralized and a decentralized manner. In the case of a networked system, the idea of a decentralized application is tempting. However, when it comes to the study of the load-frequency control loop of a power network, the decentralized approach implies that the estimator has to be designed based exclusively on the model of each interconnected control area and the locally available measured quantities. Under this scope, the tie line dynamics are often ignored, since they implicate not only the frequency deviation of one control area, but also the frequency deviations of all the other control areas that remain connected with the first one. As a matter of fact, the Equation (27) is usually disregarded and the estimator is designed based only on the Equations (9) and (10), where the tie line power deviation $\Delta P_{tie,i}[k]$ is considered as an additional externally measured signal.

Although the decentralized approach works effectively in most networked systems, a centralized approach can also be advantageous, since it involves the overall system dynamics and may allow us to obtain a better estimation of the state vector. We highlight that, in the context of a secure-oriented analysis, the estimation is not intended for control purposes, since the state vector of the power system is already available through direct measurements. However, the estimation is still a necessary tool used for the detection of an attacker.

A centralized estimation method can be applied on a networked power system in a meaningful manner, only when the network under consideration can operate in a small scale and, preferably, islanded from the rest of the grid. Such a case was described in the previous section. Further examples include smart grids and micro grids, where, the islanded operation is an essential property of their function. Thus, centralized estimation-based attack detectors can find practical application in such systems, especially if we consider that these systems have a load-frequency control loop, that shares similar traits and vulnerabilities with the one described in this article.

In the sequel, we examine two methods regarding the design of centralized estimation-based attack detectors. In the first method, the estimation residue is calculated based exclusively on the output measurements, whereas in the second improved method we use the entire state vector.

3.1. Output-Based Estimator

Let us consider the discrete-time state space model of the networked power system given by the Equations (28)–(37) for $n = 3$. The discrete-time state space model of the corresponding Luenberger observer is given as

$$\hat{S} : \hat{x}_{net}[k+1] = A_{net}\hat{x}_{net}[k] + L\left(y_{net}[k] - \hat{y}_{net}[k]\right), \quad \hat{x}_{net}[0] = 0, \quad \hat{y}_{net}[0] = 0, \tag{38}$$

$$r[k] = \|y_{net}[k] - \hat{y}_{net}[k]\|_2, \tag{39}$$

where the symbols $\hat{x}_{net}$ and $\hat{y}_{net}$ denote the estimated values of the state vector x_{net} and the output y_{net} respectively, the symbol r denotes the value of the estimation residue and $\|\cdot\|_2$ denotes the Euclidean norm. We highlight that the Equations (38) and (39) implicate the output signal y_{net}, which is derived

from the Equations (28) and (29) and thus it contains the corrupted frequency measurements. On the other hand, the matrix $L \in \mathbb{R}^{(n+1)N \times 2N}$ is selected so as to ensure that the error dynamics of the estimation-based detector respond fast in the presence of disturbances.

According to the Equation (29), the output signal y_{net} is expressed as a linear combination of the state variables implicated in the control signals. If we consider the form of the matrix C_{net}, then the Equations (38) and (39) become

$$\hat{S} : \hat{x}_{net}[k+1] = A_{net}\hat{x}_{net}[k] + LC_{net}\left(x_{net}[k] - \hat{x}_{net}[k]\right), \quad \hat{x}_{net}[0] = 0, \tag{40}$$

$$r[k] = \| C_{net}\left(x_{net}[k] - \hat{x}_{net}[k]\right) \|_2. \tag{41}$$

Now, if we introduce the estimation error

$$e[k] = x_{net}[k] - \hat{x}_{net}[k], \tag{42}$$

then the error dynamics of the estimation-based detector can be obtained from the Equations (40) and (41) as

$$e[k+1] = \left(A_{net} - LC_{net}\right) e[k], \quad e[0] = e_0, \tag{43}$$

$$r[k] = \| C_{net} e[k] \|_2. \tag{44}$$

The gain matrix L is calculated using pole placement methods. A commonly accepted approach is to demand that the eigenvalues of the error dynamics are at least ten times faster than the dynamics of the system. To this end, we can select a matrix L such that

$$eig\{A_{net} - LC_{net}\} = 0.1 \times eig\{A_{net}\}. \tag{45}$$

We remark that an equivalent, but more convenient, way to express the Equation (45) is the equation

$$eig\{A_{net}^{\top} - C_{net}^{\top} L^{\top}\} = 0.1 \times eig\{A_{net}\},$$

which encapsulates the objective of pole placement in a manner that fits nicely in most programming routines.

The design of a Luenberger observer is possible if and only if the system under consideration is observable, that is when the pair (A_{net}, C_{net}) satisfies the condition

$$rank[O] = (n+1)N, \tag{46}$$

where the observability matrix O is defined as

$$O = \begin{bmatrix} C_{net} \\ C_{net}A_{net} \\ C_{net}A_{net}^2 \\ \vdots \\ C_{net}A^{(n+1)N-1} \end{bmatrix}. \tag{47}$$

It is apparent that the property of observability depends not only on the form of the matrix A_{net} but also on the form of the output matrix C_{net}. In general, we assume that the output matrix of any system is formulated so as to ensure that the state variables implicated in the control signals are available through direct measurements. In our case, the state variables implicated in the control signals $u_{d,i}$ are Δf_i and $\Delta P_{tie,i}$. Hence, the output matrix C_{net} has the structure that we established in the Equation (37). Extensive simulations indicated that the model of the networked power system

provided by the Equations (28)–(37) for $n = 3$ remains always observable, regardless of the way that the individual control areas are connected with each other. On the other hand, the Equations (28)–(37) for $n = 4$ yield the structurally unobservable system model that was used for the design of the set-theoretic detectors in [26,27]. Since the observability property is necessary for the design of an estimator, and since this property can only be ensured through the reduction of the state vector in every control area, the modified modeling approach with $n = 3$, that we presented in the previous section, is the preferred way to continue.

An estimation-based detector activates an alarm signal, whenever the steady-state value of the estimation residue $r[k]$ exceeds a critical threshold. In order to determine this critical threshold, we have to calculate the maximum admissible steady-state value of $r[k]$ in the presence of the maximum admissible load disturbances $\Delta P_{L,i}[k]$ and in the absence of an attacker, that is when $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. However, the estimation residue can converge to a constant steady-state value only when the load changes $\Delta P_{L,i}[k]$ are described by step functions. We remark that the assumption about piecewise constant load changes is a realistic one, because the load-frequency control loop is expected to act in a different time scale and much faster than the variation of the power loads. Since we assume that $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$, the primary frequency control $\Delta P_{f,i}$ and the automatic generation control $\Delta P_{c,i}$ ensure that after each step load change, both the frequency and the tie line power deviations converge to their nominal values, that is

$$\lim_{k\to\infty} \Delta f_i[k] = \lim_{k\to\infty} \Delta P_{tie,i}[k] = 0, \quad \forall i \in \mathcal{I}. \tag{48}$$

The matrix C_{net} given by the Equation (37) implies that the estimation residue given by the Equation (44) depends only on the state variables Δf_i and $\Delta P_{tie,i}$ for all $i \in \mathcal{I}$. This fact along with the Equation (48) imply that after every step power load change, the estimation residue is bound to converge to 0 and, therefore, the critical steady-state threshold has to be chosen as

$$r_{crit} = 0. \tag{49}$$

We can infer that the threshold given by the Equation (49) does not suffice for the detection of an intermittent attack. Indeed, an intermittent attack pattern is based on consecutive activations and deactivations of the attack signal. If the activations remain brief and the adversary allows the transients to settle down before reactivating the attack signal, then he will always be able to create the undesirable power oscillations on the tie lines and at the same time ensure that the estimation residue will satisfy the Equation (49). On the other hand, when we tried to exploit the transient behavior of the estimation residue for the detection of an attacker, we concluded that this estimator is unable to discern between the power load changes and the actual attack signals. The latter difficulty will be fully illuminated in the simulation studies.

3.2. Full State-Based Estimator

To improve the performance of the estimation-based detector, we opted for two distinct changes. Firstly, since every state variable is measurable, we decided to use the entire state vector during the calculation of the estimation residue. Secondly, we adjusted the impact of each state variable on the estimation residue, normalizing each state variable according to its maximum admissible value. For the maximum and minimum admissible values of every state variable, we used the standard safety considerations found in the literature. Simulations will later demonstrate that this detector is able to discern an attack from a load change during the transient response of the system.

According to [21,22], the large transient fluctuations of the electrical frequency are generally undesirable since they can cause stability problems in the overall network. It is commonly accepted [21,22] that the deviations of the electrical frequencies must obey the inequalities

$$|\Delta f_i[k]| \leq \Delta f_{i,\max} = 1.5 \text{ [Hz]}, \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}. \tag{50}$$

On the other hand, the saturation constraints imposed on the control signals $u_{c,i}$ and $u_{d,i}$ imply that the maximum available mechanical power $\Delta P_{G,i}$ must also be bounded. According to [26,27], the mechanical power produced in the output of each turbine must obey the inequality

$$|\Delta P_{G,i}[k]| \leq \Delta P_{G,i,\max} = u_{i,\max}, \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}. \tag{51}$$

According to [29,33], the accumulated time errors z_i, $z_{1,i}$ and $z_{2,i}$ need to be limited and they must obey the inequalities

$$|z_i[k]|,\ |z_{1,i}[k]|,\ |z_{2,i}[k]| \leq z_{i,\max} = 3\ [\text{s}], \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}. \tag{52}$$

Finally, the tie line power deviations should not exceed certain limits, otherwise the coupled generators may face desynchronization problems. As a matter of fact, the tie line power deviations must obey theinequalities

$$|\Delta P_{tie,i}[k]| \leq \Delta P_{tie,i,\max}, \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}. \tag{53}$$

We highlight that although the inequalities (51) express hard saturation constraints that can never be violated, the inequalities (50), (52) and (53) express merely soft safety constraints. Hence, the state variables Δf_i, z_i, $z_{1,i}$, $z_{2,i}$ and $\Delta P_{tie,i}$ can also receive values that may not belong to the domains specified by the inequalities (50), (52) and (53) respectively, but this case is regarded as undesirable.

Let us now consider that the output of the networked system comprises of all the individual state variables and also that each state variable is normalized in terms of its corresponding boundary value, given by the inequalities (50)–(53). We define the normalizing vector of boundary values for every individual control area i as

$$q^*_{i,\max} = \begin{bmatrix} 1/\Delta f_{i,\max} & 1/\Delta P_{G,i,\max} & 1/z_{i,\max} \end{bmatrix}^\top \tag{54}$$

and the normalizing vector of boundary values for the tie line power deviations as

$$\bar{q}^*_{\max} = \begin{bmatrix} 1/\Delta P_{tie,1,\max} & 1/\Delta P_{tie,2,\max} & \dots & 1/\Delta P_{tie,N,\max} \end{bmatrix}^\top. \tag{55}$$

Finally, based on the Equations (54) and (55), we define the normalizing vector of boundary values for the networked power system as

$$q^*_{net,\max} = \begin{bmatrix} q^{*\top}_{1,\max} & q^{*\top}_{2,\max} & \dots & q^{*\top}_{N,\max} & \bar{q}^{*\top}_{\max} \end{bmatrix}^\top.$$

Consequently, the output matrix given by the Equation (37) for $n = 3$ can be modified as

$$C_{net} = \mathbb{I}_{(n+1)N \times (n+1)N} q^*_{net,\max}, \tag{56}$$

allowing us to extract every state variable as a system output and at the same time normalize each output with its corresponding maximum admissible or safety value.

The estimation-based detector is again designed based on the standard Equations (40)–(44), where this time the output matrix C_{net} is given by the Equation (56) and the gain matrix $L \in \mathbb{R}^{(n+1)N \times (n+1)N}$. We highlight that, in this case, the observability condition of the pair (A_{net}, C_{net}) is structurally ensured due to the form of the matrix C_{net}. Indeed, since C_{net} is now a diagonal matrix, it possesses exactly $(n+1)N$ linearly independent rows. Hence, the observability matrix O given by the Equation (47) has always full rank and the observability condition given by the Equation (46) is always satisfied.

The next step is to determine the critical steady-state threshold of the estimation residue r_{crit}, using the output matrix C_{net} of the Equation (56). In this case, we have to calculate the maximum admissible steady-state value of $r[k]$ in the presence of the maximum admissible load disturbances

$\Delta P_{L,i}[k]$ and in the absence of an attacker, that is when $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. In order to ensure that the estimation residue ultimately converges to a steady-state value, we must again assume that the disturbances under consideration are described as step load changes. Hence, based on the Equation (44), we can write that

$$r_{crit} = \lim_{k \to \infty} \left(\max_{\Delta P_{L,net}[k] \in \mathcal{W}_{net}} \| C_{net} \left(x_{net}[k] - \hat{x}_{net}[k] \right) \|_2 \right), \tag{57}$$

where the set $\mathcal{W}_{net}$ is defined as

$$\mathcal{W}_{net} = \left\{ \Delta P_{L,net} \in \mathbb{R}^N : R_{net} \Delta P_{L,net} \leq r_{net} \right\},$$

with the matrix $R_{net} \in \mathbb{R}^{2N \times N}$ given as

$$R_{net} = \begin{bmatrix} R'_1 & \mathbb{O}_{2\times 1} & \dots & \mathbb{O}_{2\times 1} \\ \mathbb{O}_{2\times 1} & R'_2 & \dots & \mathbb{O}_{2\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{2\times 1} & \mathbb{O}_{2\times 1} & \dots & R'_N \end{bmatrix}, \quad R'_i = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and the vector $r_{net} \in \mathbb{R}^{2N}$ given as

$$r_{net} = \begin{bmatrix} r_1^\top & r_2^\top & \dots & r_N^\top \end{bmatrix}^\top, \quad r_i = \begin{bmatrix} \Delta P_{L,i,\max} \\ \Delta P_{L,i,\max} \end{bmatrix}.$$

It is evident that the normed quantity of the Equation (57) is maximized when

$$\lim_{k \to \infty} x_{net}[k] = x_{net,\max}, \quad \lim_{k \to \infty} \hat{x}_{net}[k] = \hat{x}_{net,\max}, \tag{58}$$

where $x_{net,\max}$ and $\hat{x}_{net,\max}$ denote the steady-state values of the vectors x_{net} and $\hat{x}_{net}$ respectively, when each individual control area is subjected to its maximum admissible step load change $\Delta P_{L,i,\max}$ and $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. Based on the Equations (58), the Equation (57) becomes

$$r_{crit} = \max_{\Delta P_{L,net,\max} \in \mathcal{W}_{net}} \| C_{net} \left(x_{net,\max} - \hat{x}_{net,\max} \right) \|_2, \tag{59}$$

where $x_{net,\max}$ is obtained from the Equation (28) as

$$x_{net,\max} = \left(\mathbb{I}_{(n+1)N \times (n+1)N} - A_{net} \right)^{-1} D_{net} \Delta P_{L,net,\max} \tag{60}$$

and $\hat{x}_{net,\max}$ is obtained from the Equation (43) as

$$\hat{x}_{net,\max} = \left(\mathbb{I}_{(n+1)N \times (n+1)N} - (A_{net} - L C_{net}) \right)^{-1} L C_{net} x_{net,\max}, \tag{61}$$

while the maximization in the Equation (59) is performed over all vectors $\Delta P_{L,net,\max}$, that represent the vertices of the convex and compact polyhedral set $\mathcal{W}_{net}$.

4. Set-Theoretic Detector Design

The state estimators provide detection guarantees only through the steady-state value of the estimation residue. In the simulations section, we shall demonstrate that the overshoots that occur during the transient response of the system can be exploited for the detection of an attacker, but the criteria have to be decided ad-hoc and they do not offer specific assurances. In order to consider the overall behavior of the system response and still offer detection guarantees, we have to calculate a

robust invariant set for the networked system dynamics given by the Equations (28)–(37) for $n = 4$. We highlight that choosing $n = 4$ is mandatory for the design of an effective centralized set-theoretic detector. Indeed, if the control area models are derived through the Equations (20)–(25), then, according to the stability analysis given in Section 2, the network is bound to respond unstably in the presence of an attacker. This unstable behavior ultimately forces the state vector to exit the convex and compact robust invariant set, thus triggering an alarm and disclosing the attacker.

First, we review the basic aspects of the set-theoretic detectors developed in [27], and in the sequel, we study a link with the estimation-based attack detectors. Finally, we present a hybrid scheme that combines both methods.

4.1. Set-Theoretic Detection Preliminaries

Let us consider the networked power system described by the Equations (28)–(37) for $n = 4$. In order to design a set-theoretic detector, first we have to extract the safety constraints of the overall power network based on the inequalities (50)–(53). The constraints of each vector ξ_i, given by the Equation (20), can be expressed as

$$\mathcal{X}_i = \{\xi_i \in \mathbb{R}^n : Q_i \xi_i \leq q_i\},$$

where $Q_i \in \mathbb{R}^{2n \times n}$ and $q_i \in \mathbb{R}^{2n}$ are given as

$$Q_i = \begin{bmatrix} \mathbb{I}_{n \times n} \\ -\mathbb{I}_{n \times n} \end{bmatrix}, \quad q_i = \begin{bmatrix} q_{i,\max} \\ q_{i,\max} \end{bmatrix},$$

$$q_{i,\max} = \begin{bmatrix} \Delta f_{i,\max} & \Delta P_{G,i,\max} & z_{i,\max} & z_{i,\max} \end{bmatrix}^\top.$$

Accordingly, the safety constraints of each tie line power deviation $\Delta P_{tie,i}$ can be expressed as

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

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

$$Q'_i = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad q'_i = \begin{bmatrix} \Delta P_{tie,i,\max} \\ \Delta P_{tie,i,\max} \end{bmatrix}.$$

Finally, the constraints of x_{net} can be expressed as

$$\mathcal{X}_{net} = \left\{ x_{net} \in \mathbb{R}^{(n+1)N} : Q_{net} x_{net} \leq q_{net} \right\},$$

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

$$q_{net} = \begin{bmatrix} q_1^\top & q_2^\top & \dots & q_N^\top & q_1'^\top & q_2'^\top & \dots & q_N'^\top \end{bmatrix}^\top,$$

where $Q_{net,11} \in \mathbb{R}^{2nN \times nN}$ and $Q_{net,22} \in \mathbb{R}^{2N \times N}$ are given as

$$Q_{net,11} = \begin{bmatrix} Q_1 & \mathbb{O}_{2n \times n} & \dots & \mathbb{O}_{2n \times n} \\ \mathbb{O}_{2n \times n} & Q_2 & \dots & \mathbb{O}_{2n \times n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{2n \times n} & \mathbb{O}_{2n \times n} & \dots & Q_N \end{bmatrix}, \quad Q_{net,22} = \begin{bmatrix} Q'_1 & \mathbb{O}_{2 \times 1} & \dots & \mathbb{O}_{2 \times 1} \\ \mathbb{O}_{2 \times 1} & Q'_2 & \dots & \mathbb{O}_{2 \times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{2 \times 1} & \mathbb{O}_{2 \times 1} & \dots & Q'_N \end{bmatrix},$$

whereas $Q_{net,12}$ and $Q_{net,21}$ are given as

$$Q_{net,12} = \mathbb{O}_{2nN\times N}, \quad Q_{net,21} = \mathbb{O}_{2N\times nN}.$$

The set of all the states ξ_i that yield a control law $u_{d,i}[k]$ that respects the input saturation hard constraints is expressed as

$$\mathcal{U}_i = \{\xi_i \in \mathbb{R}^n : P_i\xi_i[k] \leq p_i, \quad \forall k \geq 0\},$$

where $P_i \in \mathbb{R}^{2\times n}$ and $p_i \in \mathbb{R}^2$ are given as

$$P_i = \begin{bmatrix} -(1/R_i)\,C_i & K'_{I_1,i} & K'_{I_2,i} \\ (1/R_i)\,C_i & -K'_{I_1,i} & -K'_{I_2,i} \end{bmatrix}, \quad p_i = \begin{bmatrix} u_{i,\max} \\ u_{i,\max} \end{bmatrix}.$$

Therefore, the set of the states x_{net} that obey the input saturation constraints is expressed as

$$\mathcal{U}_{net} = \left\{x_{net} \in \mathbb{R}^{(n+1)N} : P_{net}x_{net}[k] \leq p_{net}, \quad \forall k \geq 0\right\},$$

$$P_{net} = \begin{bmatrix} P_{net,1} & P_{net,2} \end{bmatrix}, \quad p_{net} = \begin{bmatrix} p_1^\top & p_2^\top & \dots & p_N^\top \end{bmatrix}^\top,$$

where $P_{net,1} \in \mathbb{R}^{2N\times nN}$ and $P_{net,2} \in \mathbb{R}^{2N\times N}$ are given as

$$P_{net,1} = \begin{bmatrix} P_1 & \mathbb{O}_{2\times n} & \dots & \mathbb{O}_{2\times n} \\ \mathbb{O}_{2\times n} & P_2 & \dots & \mathbb{O}_{2\times n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{2\times n} & \mathbb{O}_{2\times n} & \dots & P_N \end{bmatrix}, \quad P_{net,2} = \mathbb{O}_{2N\times N}.$$

Now, we can specify the set of safety and admissible constraints of x_{net} as $\mathcal{A}_{net} = \mathcal{X}_{net} \cap \mathcal{U}_{net}$ and its maximal robust invariant subset is defined as

$$\mathcal{A}_{net,\infty} = \{x_{net,0} \in \mathcal{A}_{net} : A_{net}x_{net}[k] + D_{net}\Delta P_{L,net}[k] \in \mathcal{A}_{net,\infty}, \quad \forall \Delta P_{L,net}[k] \in \mathcal{W}_{net}, \quad \forall k \geq 0\}. \quad (62)$$

A maximal robust invariant subset is usually extracted in terms of the algorithm proposed in [34]. However, this algorithm returns the desired invariant set after a finite number of iterations only when the system is described by asymptotically stable dynamics. According to [27], the networked system dynamics (28)–(37) for $n = 4$ are Lyapunov stable, that is the system has some eigenvalues located exactly on the boundary of the unit disc. In [27], we developed an alternative method, that allows us to extract an approximation of the desired robust invariant set, say $\hat{\mathcal{A}}_{net,\infty} \simeq \mathcal{A}_{net,\infty}$. The corresponding set-theoretic detection mechanism was summarized in the alarm signal

$$\rho(x_{net}[k]) = \begin{cases} 0, & \text{if} \quad x_{net}[k] \in \hat{\mathcal{A}}_{net,\infty} \\ 1, & \text{otherwise} \end{cases}. \quad (63)$$

For the procedure followed to extract the set $\hat{\mathcal{A}}_{net,\infty}$, the reader is referred to our work in [27], where the centralized nature of the method is also justified. In [26], we developed a second method, which involves not one, but two robust invariant sets. However, for the purposes of this article, where we aim to link estimation-based and set-theoretic attack detectors, the first method will suffice, since it is direct and requires only one set.

4.2. Set-Theoretic Methods and State Estimation

Since we are about to reexamine estimators, we must consider that the networked power system is described by the Equations (28) and (37) for $n = 3$. Our intention here is to apply, if possible, the invariance concept of the Equation (62) on the estimators that we developed in the previous section. To this end, the Equations (43) and (44), describing the error dynamics of the estimator, must be written in a way that will allow us to amend for the unknown but bounded load disturbances $\Delta P_{L,net}$, that is

$$e[k+1] = (A_{net} - LC_{net})\, e[k] + D_{net}\Delta P_{L,net}[k], \quad e[0] = e_0. \tag{64}$$

It is important to highlight that the Equation (64) is only meant for the extraction, if possible, of a robust invariant set for the estimator. The dynamic model of the estimator is always given by the Equations (40) and (41). We remark that although the load disturbances $\Delta P_{L,net}$ do not appear directly in the Equations (40) and (41), they are nonetheless considered in the calculation of the critical steady-state threshold of the estimation residue r_{crit}. The matrices A_{net} and D_{net} implicated in the Equation (64) are given by the Equations (32)–(36) respectively and since we refer to a state estimator, we set $n = 3$. Finally, since we try to extract a robust invariant set, we require the overall state vector. Hence, we must use the full state-based estimator and the matrix C_{net} has to be given by the Equation (56).

To quantify the maximum admissible discrepancies of the estimation residue $r[k]$ during the transient response and during the steady-state phase, first we need to bound the state trajectories of the estimation error $e[k]$ in terms of the robust invariant set

$$\mathcal{E}_\infty = \{e_0 \in \mathcal{E} : (A_{net} - LC_{net})\, e[k] + D_{net}\Delta P_{L,net}[k] \in \mathcal{E}_\infty, \quad \forall \Delta P_{L,net}[k] \in \mathcal{W}_{net}, \quad \forall k \geq 0\}\,.$$

The initial set $\mathcal{E}$ can be defined in a manner similar to the one employed for the set $\mathcal{A}_{net}$, but this time we have to use the estimation error e instead of the state vector x_{net}. We remark that since the matrix $A_{net} - LC_{net}$ is by design asymptotically stable, the corresponding maximal robust invariant set $\mathcal{E}_\infty$ can always be calculated using the standardized algorithm presented in [34].

The problem in this approach is not the calculation of $\mathcal{E}_\infty$ but the definition of the initial set $\mathcal{E}$. Indeed, if we consider the Equation (42), then it becomes apparent that the safety constraints of the estimation error e stem from the safety constraints of the state vectors x_{net} and $\hat{x}_{net}$. Since $\hat{x}_{net}$ is the estimation of x_{net}, it is natural to assume that both state vectors x_{net} and $\hat{x}_{net}$ must satisfy the exact same constraints. Let us now consider, without any loss of generality, the constraints associated with the frequency deviations Δf_i and their estimated values $\hat{\Delta f}_i$. According to the inequality (50), we have

$$-1.5 \leq \Delta f_i[k] \leq 1.5 = \Delta f_{i,\max}, \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}, \tag{65}$$

$$-1.5 \leq \hat{\Delta f}_i[k] \leq 1.5 = \Delta f_{i,\max}, \quad \forall k \geq 0, \quad \forall i \in \mathcal{I}. \tag{66}$$

The first element of the state vector $e[k]$ is defined as

$$e_1[k] = \Delta f_1[k] - \hat{\Delta f}_1[k]$$

and based on the inequalities (65) and (66) we have

$$-3 \leq e_1[k] \leq 3 = 2\Delta f_{1,\max}, \quad \forall k \geq 0. \tag{67}$$

The inequality (67) creates a major problem, because it implies that it is both possible and admissible to have, for example, $\hat{\Delta f}_1[k] = 0$ and also $\Delta f_1[k] = 2 \geq 1.5$ for some $k \geq 0$. In other words, even if the inequality (67) is always satisfied, the inequalities (65), that concern the actual state variables, may be violated. Clearly, this problem can be generalized for the other state variables as well.

Since the constraints imposed on the estimation error do not guarantee the satisfaction of the inequalities (50)–(53), our next best alternative is to use the estimated state vector $\hat{x}_{net}$ alone. This is possible only if the estimator (40) is designed with a matrix $L = \mathbb{O}_{(n+1)N\times(n+1)N}$. In this case, the estimator dynamics (64) become

$$\hat{x}_{net}[k+1] = A_{net}\hat{x}_{net}[k] + D_{net}\Delta P_{L,net}[k], \quad \hat{x}_{net}[0] = 0. \tag{68}$$

We remark that the matrix A_{net} used in the design of the state estimators is defined by the Equations (32)–(35) for $n = 3$ and is always asymptotically stable. Therefore, the dynamics (68) are guaranteed to converge, even without the use of the gain matrix L. Careful observation of the Equation (68) reveals that it has the same form as the Equation (28), when $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. If we consider that the overall state vector x_{net} is always measurable, then it becomes clear that there is no need to involve the estimated vector $\hat{x}_{net}$ in the calculation of the robust invariant set. Instead, we can work directly with the actual state vector x_{net} and calculate the robust invariant set for the overall networked power system. As a matter of fact, the set-theoretic detector, described by the alarm signal (63), is the only way to quantify the admissible bounds of the state trajectories of the power network.

4.3. Hybrid Detection Concept

Although the set-theoretic detectors provide strict detection guarantees, even in the case of an intermittent attack that occurs at the same time with a power load change, they tend to react slowly when the attack signals obtain small values. On the other hand, the estimation-based detectors are inherently unable to detect intermittent attacks using the steady-state value of the estimation residue, but the transient overshoots of the residue can be helpful to hint the presence of an adversary almost immediately after the activation of the attack signal. Since an estimator does not offer strict detection guarantees, we can exploit the transient behavior of the residue to put the system into alert state, thus reacting faster to any observed abnormalities, and then wait to verify these abnormalities once the state vector exits the convex and compact polyhedral robust invariant set, whereupon the set-theoretic detector triggers the actual alarm signal.

As a matter of fact, we can combine the best feats of the two methods based in the following reasoning: First, we calculate the critical steady-state threshold of the full state-based estimator and then we use it to determine an upper bound r_b of the transient response overshoots. For the transient bound r_b we assume that

$$r_b = \gamma r_{crit} \tag{69}$$

for some arbitrary parameter $\gamma > 1$, that has to be chosen ad-hoc. We highlight that although we do not provide an explicit mathematical expression for the transient bound r_b, an appropriate value can always be extracted through simulations. It is also worth mentioning that the transient bound r_b does not qualify as a robust criterion and does not offer any specific detection assurances.

Finally, the hybrid attack detection mechanism can be introduced in terms of the following three-modal system operation concept:

$$\text{mode} = \begin{cases} \text{alarm condition,} & \text{if} \quad \rho(x_{net}[k]) = 1 \quad \text{or} \quad \lim_{k\to\infty} r[k] > r_{crit} \\ \text{alert state,} & \text{if} \quad r[k] > r_b \quad \text{and} \quad \rho(x_{net}[k]) = 0 \\ \text{normal operation,} & \text{otherwise} \end{cases}. \tag{70}$$

5. Simulation Studies

In this section, we investigate the security enhancing capabilities of the previously developed estimation-based and set-theoretic attack detectors, considering the load-frequency control loop of the test case two-area power system, subject to an intermittent data corruption cyber-attack on the

frequency measurements. The efficiency of the proposed hybrid detection scheme is also commented and verified in each attack scenario studied.

According to [27], an attacker can change the nominal regulation point of the electrical frequency of the power network, and thus cause serious system malfunctions, via a persistent coordinated attack scheme with a constant attack signal $a_i = a$ for all $i \in \mathcal{I}$ and $\sigma_i[k] = 1$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. Although persistent attacks of this kind are potent, they require a significant amount of resources for their implementation, such as free access to an abundance of communication channels and the ability to corrupt them. Furthermore, their effect on the response of the system renders them immediately detectable, without the need of any kind of detector, because every nonzero steady-state frequency deviation is immediately regarded as unacceptable.

In contrast, intermittent attacks, that affect only some control areas, are much more realistic, since they require less resources for their implementation. From an impact point of view, the intermittent attacks are equally potent, since they can create oscillations on the power exchanged between the connected areas. In this way, they stress the tie lines to their thermal limits, forcing them to transmit on average more power than they are scheduled to, and they also compromise the synchronized operation of the coupled generators. Finally, their effect on the response of the power system is only temporary, something that makes them hard to detect during the nominal operation of the grid and even harder in the presence of load changes.

The parameters of the two-area power network used in the simulations are provided in the Table 1. The formulas of $K_{p,i}$ and $T_{p,i}$ are given as

$$K_{p,i} = \frac{1}{D_i}, \quad T_{p,i} = \frac{2H_i P_{B,i}}{f^\circ D_i}, \quad \forall i \in \mathcal{I},$$

where $f^\circ = 50$ [Hz] is the nominal network frequency. The simulations start at $k = 0$, the initial condition is set to be $x_{net}[0] = 0$ and we have a global sampling frequency $f_s = 100$ [Hz]. The tie line is assumed to be lossless, the nominal exchanged power between the areas is scheduled to be $P^\circ_{tie,1} = -P^\circ_{tie,2} = 1000$ [MW] and the synchronization coefficients are $T_{12} = T_{21} = 175$ [MW/rad]. The bounds $\Delta P_{L,i,\max}$ are selected as small percentages of the power bases $P^\circ_{B,i}$ and the bounds $\Delta P_{tie,i,\max}$ are selected through simulations. Specifically, we observed that even in the presence of the maximum admissible step load changes $\Delta P_{L,i,\max}$, the deviations $\Delta P_{tie,i}$ remain always bounded by $\Delta P_{tie,i,\max} = 0.5|P^\circ_{tie,i}|$. The bounds $u_{i,\max}$ service the maximum admissible load changes and they also ensure the existence of a nonempty robust invariant set $\hat{\mathcal{A}}_{net,\infty}$. The design parameters of the set-theoretic detector that was used here are taken directly from [27].

Table 1. Parameter values for the two-area power system. *Source:* [29].

Parameter	Symbol	Area 1 Value	Area 2 Value	Units
Power Base	$P_{B,i}$	2000	1500	MW
Load Dependency Factor	D_i	16.66	10.5	MW/Hz
Speed Droop	R_i	1.2×10^{-3}	1.33×10^{-3}	Hz/MW
Generator Inertia Constant	H_i	5	4	s
Turbine Static Gain	$K_{T,i}$	1	1	MW/MW
Turbine Time Constant	$T_{T,i}$	0.3	0.25	s
Area Static Gain	$K_{p,i}$	0.06	0.095	Hz/MW
Area Time Constant	$T_{p,i}$	24	22.85	s
Controller Static Gain	$K_{I,i}$	0.5	0.45	1/s
Control Input Bound	$u_{i,\max}$	600	450	MW
Power Load Bound	$\Delta P_{L,i,\max}$	20	15	MW

The performance of both the estimation-based and the set-theoretic attack detectors is assessed in the case of an intermittent attack that affects only the first control area. The switching signal $\sigma_1[k]$ is given by the Equation (26), the switching bounds are given as $a_{1,\min} = 0.01$ [Hz] and $a_{1,\max} = 0.1$ [Hz],

whereas the tolerance is selected as $\delta = 10^{-3}$. For the second area we have an attack signal $\alpha_2 = 0$ and $\sigma_2[k] = 0$ for all $k \geq 0$. We remark that the value of $\alpha_{1,\min}$ is meaningful only if it is smaller than the frequency measurement error ($\sim 10^{-3}$).

The remainder of this section is split into three parts. In the first part, we calculate the steady-state and transient critical thresholds of the output-based and the full state-based estimators, verifying their values via simulations. In the second part, we perform a comparative study on the behavior of the output-based, the full state-based and the set-theoretic attack detectors, for indicative values of the attack signal α_1. It is shown that the two detectors complement each other and that we can achieve better results if we use the three-modal system operation proposed in the Equation (70). Finally, in the third part, we address certain limitations of the above detection mechanisms.

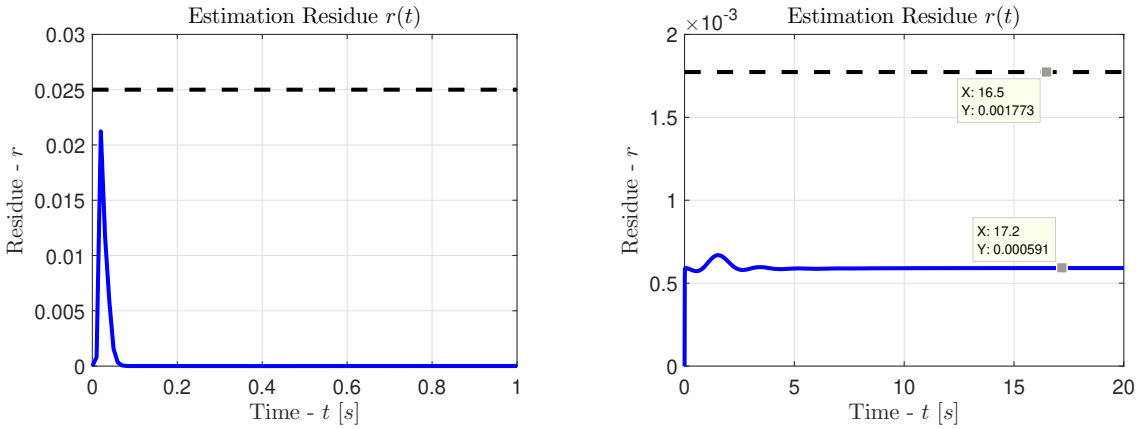
The polyhedral constraints as well as the optimization problems were handled with the MPT Toolbox 3.0 [35] and the YALMIP library [36].

5.1. Estimator Thresholds Verification

For the extraction of the steady-state thresholds, we have to consider that each control area is affected by its maximum admissible step load change. Since the bounds of all $\Delta P_{L,i}$ are symmetric for all $i \in \mathcal{I}$, the maximum admissible step load changes for the overall networked system can be described by any vertex of the polyhedral set $\mathcal{W}_{net}$. Consequently, we can select $\Delta P_{L,net,\max} = [20 \quad 15]^\top$ and then apply the disturbance $\Delta P_{L,net}[k] = \Delta P_{L,net,\max}$ for all $k \geq 0$. We highlight that for the extraction of the estimator thresholds, we must assume that the system evolves in the absence of an attacker, that is when $\sigma_i[k] = 0$ for all $k \geq 0$ and for all $i \in \mathcal{I}$.

For the output-based estimator we have explained that the steady-state threshold must always be selected as $r_{crit} = 0$. Hence, the transient threshold cannot be explicitly derived by the Equation (69) and needs to be decided arbitrarily. We select $r_b = 2.5 \times 10^{-2}$ and based on the Figure 2a the values of both thresholds are meaningful. For the full state-based estimator the steady-state threshold is given by the Equations (59)–(61) as $r_{crit} = 5.91 \times 10^{-4}$ and this value is verified by the Figure 2b. Based on the Equation (69) for $\gamma = 3$, we can obtain the transient bound as $r_b = 1.773 \times 10^{-3}$.

We highlight that for a two-area power system, and only then, we encounter the degenerate case $\Delta P_{tie,1} = -\Delta P_{tie,2}$. Therefore, the equation used to describe $\Delta P_{tie,2}$ in the network dynamics (28) and (29) can be neglected as redundant during the design of the estimators.



(**a**) Output-based estimator thresholds.

(**b**) Full state-based estimator thresholds.

Figure 2. Verification of the estimators steady-state and transient thresholds.

5.2. Case Study of an Intermittent Attack

The detection capabilities of the proposed detectors are assessed in the presence of unknown load changes. We assume that the power network is subject to the following step load changes:

$$\Delta P_{L,1}(t) = 15 \text{ [MW]}, \quad \forall t \geq 0 \text{ [s]},$$

$$\Delta P_{L,2}(t) = \begin{cases} 0 \text{ [MW]}, & \text{if} \quad 0 \leq t < 20 \text{ [s]} \\ -10 \text{ [MW]}, & \text{if} \quad t \geq 20 \text{ [s]} \end{cases}.$$

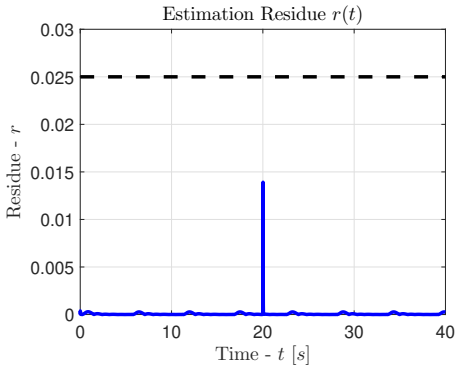
Let us observe the Figure 3, where we depict with solid blue lines (•) the response of the estimation residue and with dashed black lines (•) the transient threshold r_b. It is evident that an output estimator is unable to disclose an attack with $\alpha_1 = 2.2$, when it is driven by the switching signal of the Equation (26). In addition, if we consider the transient threshold r_b and the hybrid detection scheme (70), then it is obvious that the output estimator will never put the system in alert mode, since $r[k] \leq r_b$ for all $k \geq 0$. Furthermore, the intense spike observed in the Figure 3a occurs not due to the attack but due to the load change that happens at $t = 20$ [s]. On the other hand, the improved full state estimator is able to put the system into alert mode every time the attacker affects the automatic generation control unit and is always able to discern between an attack and a load change. Indeed, the normalization of the state variables with their maximum safety values smooths the spike at $t = 20$ [s], and creates overshoots in the response of the residue only when an actual attacker affects the networked system. Finally, the set-theoretic detector will never trigger an alarm, unless the state vector exits the set $\hat{\mathcal{A}}_{net,\infty}$ at $t = 33.6$ [s]. In this case, the hybrid scheme (70) allows us to react faster in comparison to the case where we would depend only on the use of a set-theoretic detector.

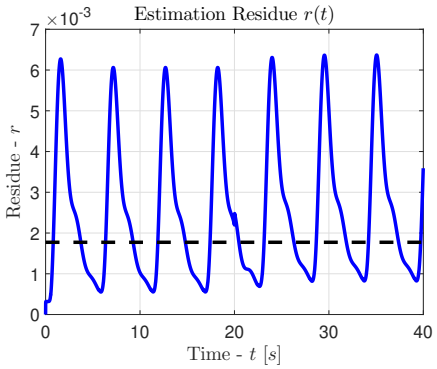
Let us now perform an analysis regarding the detection and the early detection of the attack, in terms of absolute time units. For $\alpha_1 = 2.2$, the Figure 3b reveals that the first activation of an alert occurs at $t = 0.81$ [s]. Considering that the set-theoretic detector will trigger an alarm at $t = 33.6$ [s], we infer that the hybrid detection scheme hints the abnormal system behavior $\Delta t = 33.6 - 0.81 = 32.79$ [s] faster than the set-theoretic detector, offering a significant improvement.

Accordingly, in the Figure 4, we consider the case of an attack with $\alpha_1 = 4.2$. Again, the output estimator is unable to detect an attack during the transient response and decides that the system operates normally. Once more, the spike observed in the Figure 4a at $t = 20$ [s] is caused by the load change and not by the attack. On the other hand, the improved full state estimator is again able to smooth the load change spike and can put the system into alert mode at the correct time instances. Lastly, for $\alpha_1 = 4.2$, we observe that the set-theoretic detector can trigger an alarm whenever the attacker affects the system, ensuring its disclosure even from the very first activation.

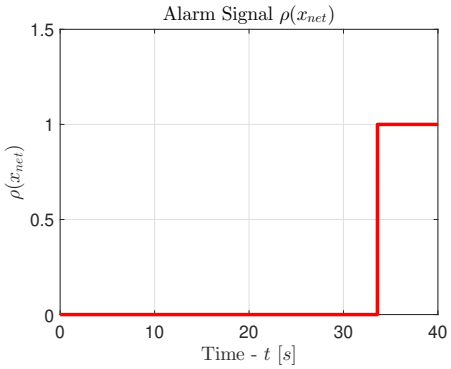
Next, we address again the matter of the detection and the early detection of the attack, in terms of absolute time units. For $\alpha_1 = 4.2$, the Figure 4b reveals that the first activation of an alert occurs at $t = 0.42$ [s] and occurs faster than the first activation of an alert when $\alpha_1 = 2.2$, because now we have an attack signal with a larger value. On the other hand, the Figure 4c reveals that the set-theoretic detector activates an alarm for the first time at $t = 0.38$ [s]. In this case, an alarm is triggered almost immediately after the activation of the attacker, therefore the hybrid detection scheme does not offer further advantages in the faster detection of the adversary.

In conclusion, we infer that the hybrid detection scheme (70) encapsulates successfully the best traits of the two detection methods. Whenever the set-theoretic detector triggers an alarm, we know with certainty that the state vector has exited the robust invariant set $\hat{\mathcal{A}}_{net,\infty}$, thus we can guarantee the existence of an attacker. Accordingly, whenever the residue of the full state estimator exceeds a transient threshold, while the alarm signal is inactive, the existence of an attacker is hinted, and the system is put in alert state. In this way, we can deploy countermeasures faster than we would if we relied only on the outcome of the set-theoretic detection mechanism.



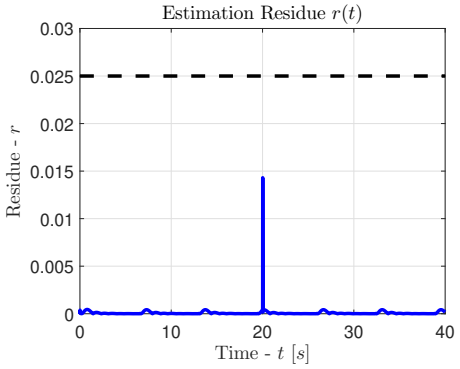


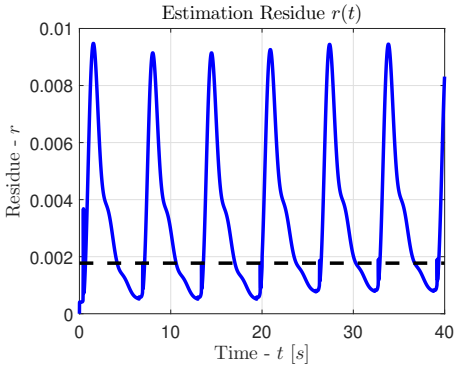
(**a**) Residue evolution of the output-based estimator. (**b**) Residue evolution of the full state-based estimator.



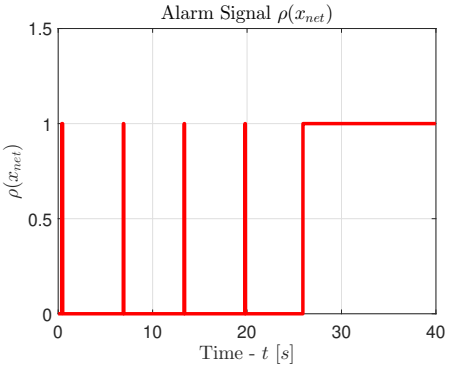
(**c**) Alarm signal evolution of the set-theoretic detector.

Figure 3. Estimation residue and alarm signal evolution for an intermittent attack with $\alpha_1 = 2.2$.





(**a**) Residue evolution of the output-based estimator. (**b**) Residue evolution of the full state-based estimator.

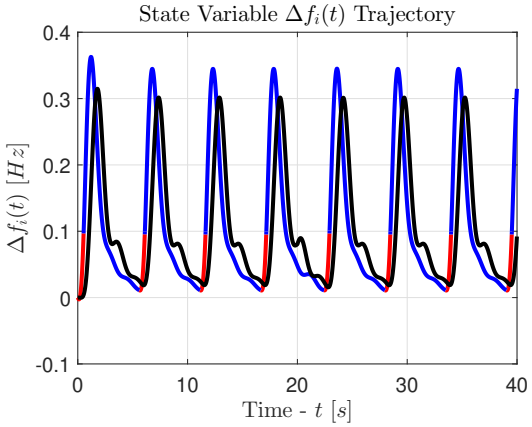


(**c**) Alarm signal evolution of the set-theoretic detector.

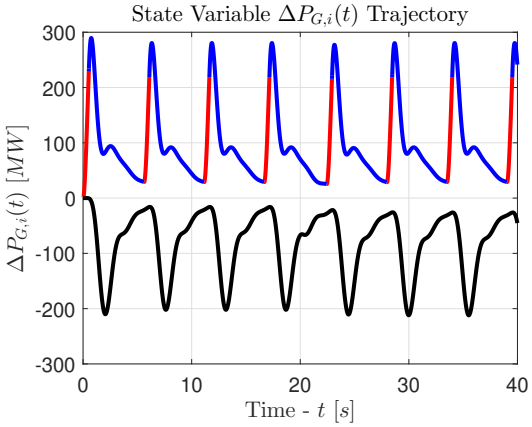
Figure 4. Estimation residue and alarm signal evolution for an intermittent attack with $\alpha_1 = 4.2$.

We close this part by providing a set of figures that depict the response of the state variables in the case of an intermittent attack with $\alpha_1 = 2.2$. These figures will be helpful in order to understand the differences between the two modeling approaches for $n = 3$ and $n = 4$, and will also clarify how the set-theoretic oriented modeling approach works for the detection of an intermittent attack pattern. The results obtained for an attack signal with $\alpha_1 = 2.2$ are similar to those obtained for an attack signal with $\alpha_1 = 4.2$, so the latter ones are omitted.

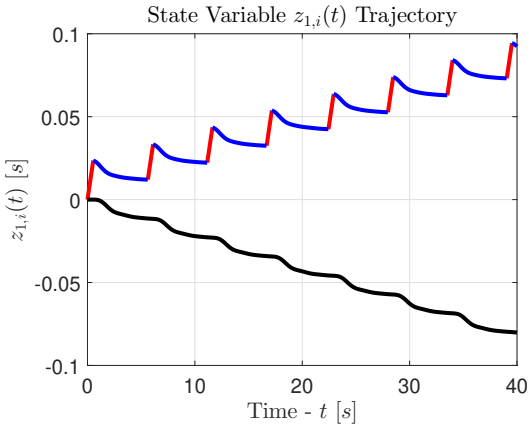
Let us observe the graphs in the Figure 5. The state variables associated with the first control area are printed in red (•) and blue (•) lines, whereas the state variables associated with the second control area are printed in black (•) lines. The red color indicates an active attacker, whereas the blue color indicates an inactive attacker. It is evident that the intermittent pattern forces the electrical frequency to oscillate, according to the hysteresis-based switching signal (26). The frequency oscillations remain within the bounds of the inequality (50). However, these oscillations cause fluctuations on the power which is exchanged between the two areas through the connecting tie line. As we can see, the state variables $\Delta P_{tie,i}$ have an oscillatory behavior and the average power that flows through the connecting tie line is increased. This leads to problems associated with the synchronization of the generators and with the thermal limits of the tie line. Finally, the state variables z_i obtained for $n = 3$ are always stable, while the state variables $z_{1,i}$ and $z_{2,i}$ obtained for $n = 4$ have an unstable behavior. Indeed, after every new activation of the attacker, the state variables $z_{1,i}$ and $z_{2,i}$ begin to diverge linearly towards infinity. This behavior is essential for a set-theoretic detector, since it ensures that at some point the state vector will exit the set $\hat{\mathcal{A}}_{net,\infty}$, ultimately triggering an alarm. This last remark is visible in the Figure 3c after $t = 33.6$ [s], when $\alpha_1 = 2.2$, and in the Figure 4c after $t = 25.92$ [s], when $\alpha_1 = 4.2$.



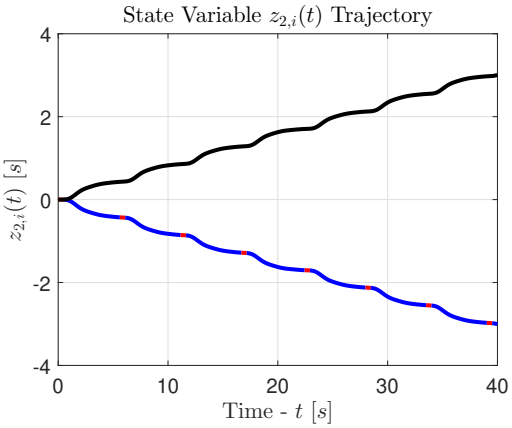
(**a**) Evolution of the electrical frequency deviations Δf_i.



(**b**) Evolution of the electrical power deviations $\Delta P_{G,i}$.

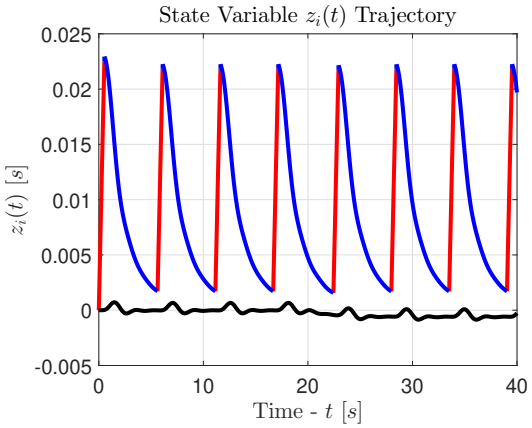


(**c**) Evolution of the accumulated time errors $z_{1,i}$.

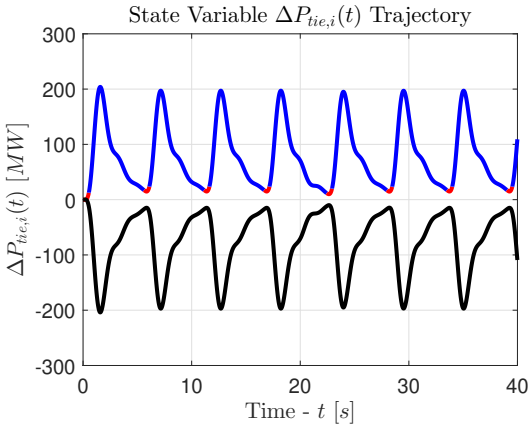


(**d**) Evolution of the accumulated time errors $z_{2,i}$.

Figure 5. *Cont.*



(**e**) Evolution of the accumulated time errors z_i.



(**f**) Evolution of the tie line power deviations $\Delta P_{tie,i}$.

Figure 5. State variable trajectories for an intermittent attack with $\alpha_1 = 2.2$.

Finally, it is worth mentioning that we cannot use the networked system dynamics (28)–(37) for $n = 4$ to develop both a set-theoretic detector and a full state estimator, because the Equation (60), which is necessary for the extraction of the steady-state estimation threshold r_{crit}, involves the matrix $(\mathbb{I}_{(n+1)N\times(n+1)N} - A_{net})^{-1}$. Indeed, the latter matrix does not exist when the matrix A_{net} is Lyapunov stable with unit eigenvalues, something that happens when we select the modeling approach that depends on the Equations (28)–(37) for $n = 4$ [27].

5.3. Limitations of the Proposed Detectors

The limitations of the proposed detection mechanisms are presented based on the occurrences of false positive or false negative activations of the alarm condition and the alert state. For the remainder of our analysis, the term false positive refers to an activation of the alarm condition or the alert state of the hybrid detection scheme (70), when the system is not affected by an attacker. Accordingly, the term false negative refers to the case where an attacker affects the system but neither the alarm condition nor the alert state are ever activated. We address the occurrences of false positives and false negatives considering the cases of the set-theoretic detectors and the estimation-based detectors separately.

First, we examine the false positives and the false negatives of the set-theoretic detectors. The set-theoretic detector triggers an alarm signal only when the state trajectory exits the robust invariant set $\hat{\mathcal{A}}_{net,\infty}$. The robust invariance property ensures that if the state vector $x_{net}[k]$ begins to evolve from an initial condition $x_{net}[0]$ that belongs to the set $\hat{\mathcal{A}}_{net,\infty}$, then the emanating state trajectory will remain within the set $\hat{\mathcal{A}}_{net,\infty}$ for all future time instances $k \geq 0$ and for all admissible disturbances sequences $\Delta P_{L,net}[k] \in \mathcal{W}_{net}$. In addition, the set $\hat{\mathcal{A}}_{net,\infty}$ has been calculated in the absence of an attacker. This implies that, it is impossible for the state trajectory to exit the set $\hat{\mathcal{A}}_{net,\infty}$ in the absence of an attacker, and therefore it is impossible for the set-theoretic detector to yield a false positive.

On the other hand, the stability analysis that was established in our previous work [27] and was briefly reviewed in the current article, indicates that the set-theoretic detector may be unable to disclose an attack, only when the adversary affects all the control areas of the network with the same persistent attack signals, that is when $\alpha_i = \alpha$ and $\sigma_i[k] = 1$ for all $k \geq 0$ and for all $i \in \mathcal{I}$. In this case, a false negative may occur, because the response of the system is proven to be asymptotically stable. Indeed, if the attack signals remain small, then the attained steady-state equilibrium may still belong to the set $\hat{\mathcal{A}}_{net,\infty}$ and respect the alarm constraints. For further details on this scenario, the reader is referred to [27]. However, if the attacker affects only some control areas of the network, but not all of them, as is the case in this work, then the integral variables $z_{1,i}$ and $z_{2,i}$ will always demonstrate an unstable behavior. In other words, it is only a matter of time until the state trajectory exits the convex and compact set $\hat{\mathcal{A}}_{net,\infty}$, whereupon an alarm will be triggered. In conclusion, it is impossible for the

set-theoretic detector to yield a false negative, unless $\alpha_i = \alpha$ and $\sigma_i[k] = 1$ for all $k \geq 0$ and for all $i \in \mathcal{I}$ and for relatively small values of the attack signals α_i.

We proceed with the analysis of the false positives and the false negatives of the estimation-based detectors. Out of the two estimation-based detectors, the output-based estimators perform poorly and unreliably in every scenario that we presented. They yield constantly false negatives and are unable to disclose an attacker. Therefore, we focus on the full state-based estimators. We have to address the false positives and the false negatives associated both with the alarm condition $\lim_{k\to\infty} r[k] > r_{crit}$ and the alert state condition $r[k] > r_b$.

First, we address the false positives and the false negatives of the alarm condition $\lim_{k\to\infty} r[k] > r_{crit}$. The full state-based estimator triggers an alarm signal only when the estimation residue reaches a steady-state value larger than the critical threshold r_{crit}. As long as the estimation residue oscillates, something that occurs in an intermittent attack, the alarm condition $\lim_{k\to\infty} r[k] > r_{crit}$ can never be used to activate an alarm, either correct or false. Persistent attacks can lead to false negatives, as shown in [27], but never to false positives, since r_{crit} is calculated for $\Delta P_{L,net,\max}$ and $\alpha_i = 0$ for all $i \in \mathcal{I}$.

Let us now address the false positives and the false negatives of the alert state condition $r[k] > r_b$. In the case of an intermittent attack, the full state-based estimator can help with the timely disclosure of the attacker by putting the system into alert mode, whenever $r[k] > r_b = \gamma r_{crit}$. Clearly, since the value of the design parameter γ has to be selected ad hoc, it is possible for this detector to fail to trigger an alert state, thus yielding a false negative. This scenario is presented in the following simulation, where the responses of the state variables are again similar to the ones given in the Figure 5 and, therefore, are omitted for brevity. We consider that the attack signal becomes $\alpha_1 = 0.2$ and that the system is now subject to a different disturbance sequence, which is described by the equations

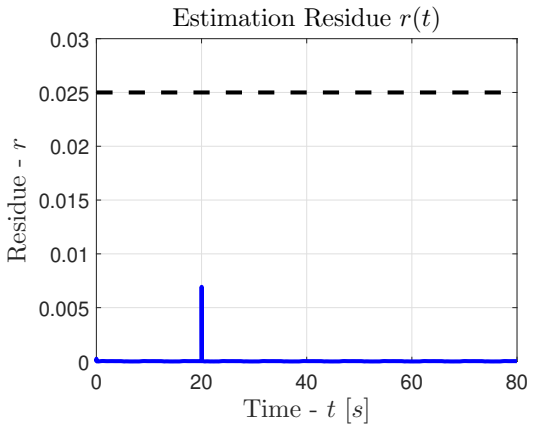
$$\Delta P_{L,1}(t) = 10 \text{ [MW]}, \quad \forall t \geq 0 \text{ [s]},$$

$$\Delta P_{L,2}(t) = \begin{cases} 0 \text{ [MW]}, & \text{if} \quad 0 \leq t < 20 \text{ [s]} \\ 5 \text{ [MW]}, & \text{if} \quad t \geq 20 \text{ [s]} \end{cases}.$$

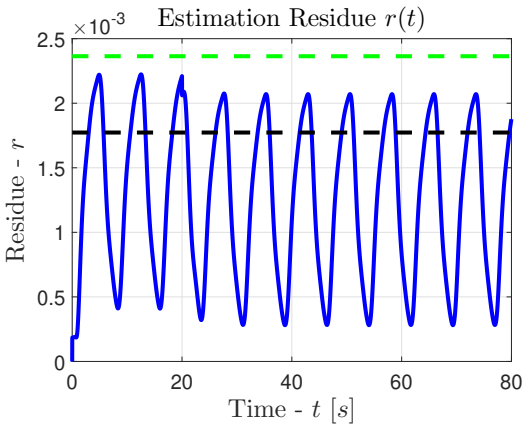
The output-based estimator and the set-theoretic detector remain the same but we consider two different transient thresholds for the full state-based estimator. Specifically, in the Figure 6b, we depict with black (•) dashed line the standard transient threshold $r_b = \gamma r_{crit}$, with $\gamma = 3$, whereas we use the green (•) dashed line to depict a different transient threshold with $\gamma = 4$. It is evident that for $\gamma = 3$ the full state-based estimator is able to put the system in alert state whenever the attacker actually affects it and the first activation of the alert state occurs at $t = 3.01$ [s]. However, if we select $\gamma = 4$, then the full state-based estimator yields constantly false negatives.

The Figure 6c also reveals that the set-theoretic detector is able to disclose the attack only when the state vector exits the robust invariant set $\hat{\mathcal{A}}_{net,\infty}$, something that happens at $t = 57.45$ [s]. Finally, we remark that the hybrid detection scheme for $\gamma = 3$ improves the behavior of the set-theoretic detector, since it is able to hint the existence of an attack $\Delta t = 57.45 - 3.01 = 54.44$ [s] faster than the set-theoretic detector alone. However, the hybrid detection scheme for $\gamma = 4$ is unable to trigger an alert state and can only activate an alarm at $t = 57.45$ [s] due to the set-theoretic detector.

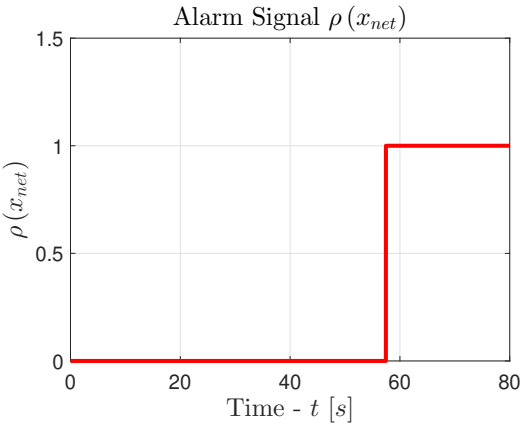
Lastly, we address the false positives regarding the condition $r[k] > r_b$. As we have already explained, since the parameter γ is selected ad hoc, the transient threshold r_b does not offer any detection guarantees. Consequently, it is possible for the full state-based estimator to yield a false positive, when an attacker is absent and the system is affected by a more elaborate disturbance sequence $\Delta P_{L,net}[k] \in \mathcal{W}_{net}$. However, typical disturbance sequences assume the form of step load changes, in which case a sufficiently large value of the parameter γ reduces the occurrence of false positives.



(**a**) Residue evolution of the output-based estimator.



(**b**) Residue evolution of the full state-based estimator.



(**c**) Alarm signal evolution of the set-theoretic detector.

Figure 6. Estimation residue and alarm signal evolution for an intermittent attack with $\alpha_1 = 0.2$.

To sum up, the set-theoretic detectors can never yield a false positive and they can yield a false negative only when $\alpha_i = \alpha$ and $\sigma_i[k] = 1$ for all $k \geq 0$ and for all $i \in \mathcal{I}$ and for small attack signals. Thus, they can never yield a false negative during intermittent attacks. The output-based detectors yield constantly false negatives and are not suitable for the disclosure of intermittent attacks. Finally, the full state-based estimators cannot rely on their alarm condition to disclose intermittent attacks but they can help inprove the security of the system through the alert condition $r[k] > r_b = \gamma r_{crit}$, which can be activated during the transients. The activation of this condition depends on the value of the ad hoc selected parameter γ and can either lead to false negatives, for small-valued attack signals paired with small-valued disturbance sequences and a poorly chosen value of γ, or to false positives, for elaborately constructed disturbance sequences and a poorly chosen value of γ. However, the latter case is rare, since usually we are mostly concerned with step load changes.

In conclusion, the above analysis along with the overall simulation results demonstrate that a hybrid detection scheme that combines a set-theoretic detector and a full state-based estimator is particularly robust, regarding the matter of false positives and false negatives. A false positive simply cannot occur, whereas a false negative may occur only under certain circumstances and can be avoided with the proper selection of the design parameters of the detectors.

6. Conlusions

In this article, we develop estimation-based and set-theoretic detectors as security-enhancing tools for the load-frequency control loop of a networked power system. The two detectors are linked in terms of a hybrid concept that combines the best feats of each approach. Comparative studies and

an overall assessment of the efficiency of the proposed compound scheme are performed in the case of a two-area power system, considering an intermittent attack, that occurs at the same time with power load disturbances. It is shown that a hybrid detection scheme, that implicates both methods, allows for the timely and precise disclosure of an intermittent adversary.

The proposed detectors can also be used for the detection of other types of attacks as long as these attacks alter directly the system dynamics. Examples include cyber-attacks that can cause parametric changes in the model of the system or physical attacks that can lead to infrastructure failures and, therefore, result in a different model of the system. If the attack alters the dynamics of the system, then clearly there is no guarantee that the emanating state trajectories will always remain inside the robust invariant set, which is used by the set-theoretic detector, and there is also no guarantee that the estimation residue will always respect the steady-state and transient thresholds. The reason is that both the robust invariant set and the estimator thresholds were calculated based on a different system model. This implies that any attack scenarios that alter the system dynamics are potentially disclosable with our methods. As a matter of fact, future researchers can test and assess the performance of the proposed detectors considering the above attack scenarios. However, we stress that the proposed detectors offer strict detection guarantees only in the case of the data corruption cyber-attacks which were addressed in this work.

Author Contributions: Conceptualization, E.K. and A.T.; Formal analysis, E.K. and A.T.; Funding acquisition, A.T. and L.D.; Investigation, E.K.; Methodology, E.K.; Project administration, A.T.; Software, E.K.; Supervision, A.T.; Validation, L.D.; Visualization, E.K.; Writing—original draft, E.K.; Writing—review & editing, A.T. and L.D. All authors have approved the publication of this paper.

Funding: L. Dritsas acknowledges financial support from the Special Account for Research of ASPETE through the funding program "Strengthening research of ASPETE faculty members".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mahmood, A.; Javaid, N.; Razzaq, S. A review of wireless communications for smart grid. *Renew. Sustain. Energy Rev.* **2015**, *41*, 248–260. [CrossRef]
2. Zhang, H.; Peng, M.; Guerrero, J.; Gao, X.; Liu, Y. Modelling and vulnerability analysis of cyber-physical power systems based on interdependent networks. *Energies* **2019**, *12*, 3439. [CrossRef]
3. Mo, Y.; Sinopoli, B. Secure control against replay attacks. In In Proceedings of the IEEE 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 30 September–2 October 2009; pp. 911–918. [CrossRef]
4. Amin, S.; Cárdenas, A.A.; Sastry, S.S. Safe and secure networked control systems under denial-of-service attacks. In *Hybrid Systems: Computation and Control*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 31–45.
5. Zhang, H.; Cheng, P.; Shi, L.; Chen, J. Optimal denial-of-service attack scheduling with energy constraint. *IEEE Trans. Autom. Control* **2015**, *60*, 3023–3028. [CrossRef]
6. Deka, D.; Baldick, R.; Vishwanath, S. Jamming aided generalized data attacks: Exposing vulnerabilities in secure estimation. In Proceedings of the IEEE 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 2556–2565. [CrossRef]
7. Gerard, B.; Rebaï, S.B.; Voos, H.; Darouach, M. Cyber security and vulnerability analysis of networked control system subject to false-data injection. In Proceedings of the IEEE 2018 American Control Conference (ACC), Milwaukee, WI, USA, 27–29 June 2018; pp. 992–997. [CrossRef]
8. Yang, Q.; Yang, J.; Yu, W.; An, D.; Zhang, N.; Zhao, W. On false data-injection attacks against power system state estimation: Modeling and countermeasures. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 717–729. [CrossRef]
9. Liu, Y.; Ning, P.; Reiter, M.K. False data injection attacks against state estimation in electric power grids. In Proceedings of the 16th ACM Conference on Computer and Communications Security, Chicago, IL, USA, 9–13 November 2009; ACM: New York, NY, USA, 2009, pp. 21–32. [CrossRef]
10. Kim, J.; Tong, L. On topology attack of a smart grid: Undetectable attacks and countermeasures. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 1294–1305. [CrossRef]

11. Kwon, C.; Liu, W.; Hwang, I. Security analysis for cyber-physical systems against stealthy deception attacks. In Proceedings of the IEEE 2013 American Control Conference (ACC), Washington, DC, USA, 17–19 June 2013; pp. 3344–3349. [CrossRef]
12. Hashemi, N.; Murguia, C.; Ruths, J. A comparison of stealthy sensor attacks on control systems. In Proceedings of the 2018 American Control Conference (ACC), Milwaukee, WI, USA, 27–29 June 2018; pp. 973–979. [CrossRef]
13. Ahmed, S.; Lee, Y.; Hyun, S.H.; Koo, I. Mitigating the impacts of covert cyber attacks in smart grids via reconstruction of measurement data utilizing deep denoising autoencoders. *Energies* **2019**, *12*, 3091. [CrossRef]
14. Sridhar, S.; Hahn, A.; Govindarasu, M. Cyber–physical system security for the electric power grid. *Proc. IEEE* **2012**, *100*, 210–224. [CrossRef]
15. Pasqualetti, F.; Dörfler, F.; Bullo, F. Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design. In Proceedings of the 50th IEEE Conference on Decision and Control (CDC) and European Control Conference (ECC), Orlando, FL, USA, 12–15 December 2011; pp. 2195–2201. [CrossRef]
16. Pasqualetti, F.; Dörfler, F.; Bullo, F. Attack detection and identification in cyber-physical systems. *IEEE Trans. Autom. Control* **2013**, *58*, 2715–2729. [CrossRef]
17. Teixeira, A.; Sandberg, H.; Johansson, K.H. Networked control systems under cyber attacks with applications to power networks. In Proceedings of the IEEE 2010 American Control Conference (ACC), Baltimore, MD, USA, 30 June–2 July 2010; pp. 3690–3696. [CrossRef]
18. Dan, G.; Sandberg, H. Stealth attacks and protection schemes for state estimators in power systems. In Proceedings of the IEEE 2010 1st IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 4–6 October 2010; pp. 214–219. [CrossRef]
19. Gallo, A.J.; Turan, M.S.; Nahata, P.; Boem, F.; Parisini, T.; Ferrari-Trecate, G. Distributed cyber-attack detection in the secondary control of DC microgrids. In Proceedings of the IEEE 2018 European Control Conference (ECC), Limassol, Cyprus, 12–15 June 2018; pp. 344–349. [CrossRef]
20. Manandhar, K.; Cao, X.; Hu, F.; Liu, Y. Detection of faults and attacks including false data injection attack in smart grid using Kalman filter. *IEEE Trans. Control Netw. Syst.* **2014**, *1*, 370–379. [CrossRef]
21. Esfahani, P.M.; Vrakopoulou, M.; Margellos, K.; Lygeros, J.; Andersson, G. Cyber attack in a two-area power system: Impact identification using reachability. In Proceedings of the IEEE 2010 American Control Conference (ACC), Baltimore, MD, USA, 30 June–2 July 2010; pp. 962–967. [CrossRef]
22. Esfahani, P.M.; Vrakopoulou, M.; Margellos, K.; Lygeros, J.; Andersson, G. 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 (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 5973–5978. [CrossRef]
23. Franzè, G.; Tedesco, F.; Casavola, A. A leader-follower architecture for load frequency control purposes against cyber attacks in power grids—Part I. In Proceedings of the 55th IEEE Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 5128–5133. [CrossRef]
24. Franzè, G.; Tedesco, F.; Casavola, A.; Garone, E. A leader-follower architecture for load frequency control purposes against cyber attacks in power grids—Part II. In Proceedings of the 55th IEEE Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 5134–5139. [CrossRef]
25. Teixeira, A.; Pérez, D.; Sandberg, H.; Johansson, K.H. Attack models and scenarios for networked control systems. In Proceedings of the 1st International Conference on High Confidence Networked Systems, Beijing, China, 17–18 April 2012; pp. 55–64. [CrossRef]
26. Kontouras, E.; Tzes, A.; Dritsas, L. Set-theoretic detection of bias injection cyber-attacks on networked power systems. In Proceedings of the IEEE 2018 American Control Conference (ACC), Milwaukee, WI, USA, 27–29 June 2018; pp. 165–170. [CrossRef]
27. Kontouras, E.; Tzes, A.; Dritsas, L. Set-theoretic detection of data corruption attacks on cyber physical power systems. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 872–886. [CrossRef]
28. Blanchini, F. Set invariance in control. *Automatica* **1999**, *35*, 1747–1767. [CrossRef]
29. Elgerd, O.I. *Electric Energy Systems Theory: An Introduction*; McGraw-Hill: New York, NY, USA, 1982.
30. Kundur, P.; Balu, N.J.; Lauby, M.G. *Power System Stability and Control*; McGraw-Hill: New York, NY, USA, 1994.

31. Suehiro, T.; Namerikawa, T. Decentralized control of smart grid by using overlapping information. In Proceedings of the IEEE SICE Annual Conference, Akita, Japan, 20–23 August 2012; pp. 125–130.
32. Liberzon, D. *Switching in systems and control, Systems & Control: Foundations & Applications*; Birkhäuser: Boston, MA, USA, 2003.
33. Fosha, C.E.; Elgerd, O.I. The megawatt-frequency control problem: A new approach via optimal control theory. *IEEE Trans. Power Apparatus Syst.* **1970**, *PAS-89*, 563–577. [CrossRef]
34. Kolmanovsky, I.; Gilbert, E.G. Theory and computation of disturbance invariant sets for discrete-time linear systems. *Math. Probl. Eng.* **1998**, *4*, 317–367. [CrossRef]
35. Herceg, M.; Kvasnica, M.; Jones, C.; Morari, M. Multi-parametric toolbox 3.0. In Proceedings of the 2013 European Control Conference (ECC), Zurich, Switzerland, 17–19 July 2013; pp. 502–510. [CrossRef]
36. Löfberg, J. YALMIP: A toolbox for modeling and optimization in MATLAB. In Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design, Taipei, Taiwan, 2–4 September 2004; pp. 284–289. [CrossRef]