Residual-Based Detection of Attacks in Cyber-Physical Inverter-Based Microgrids

Andres Intriago, *Student Member, IEEE*, Francesco Liberati, Nikos D. Hatziargyriou, *Life Fellow, IEEE*, Charalambos Konstantinou, *Senior Member, IEEE*

Abstract—This paper discusses the challenges faced by cyber-physical microgrids (MGs) due to the inclusion of information and communication technologies in their already complex, multi-layered systems. The work identifies a research gap in modeling and analyzing stealthy intermittent integrity attacks in MGs, which are designed to maximize damage and cancel secondary control objectives. To address this, the paper proposes a nonlinear residual-based observer approach to detect and mitigate such attacks. In order to ensure a stable operation of the MG, the formulation then incorporates stability constraints along with the detection observer. The proposed design is validated through case studies on a MG benchmark with four distributed generators, demonstrating its effectiveness in detecting attacks while satisfying network and stability constraints.

Index Terms—Cyber-physical microgrids, stealthy attacks, detection, nonlinear observer, stability.

I. INTRODUCTION

A. Background and Motivation

Electric power systems are increasingly using more distributed energy resources (DERs) [1]. These DERs are connected to the main grid through inverters, known as inverter-based resources (IBRs). IBRs can operate in grid-forming or grid-following mode. In grid-following mode, they cannot maintain system frequency and voltage when disconnected from the main grid. In contrast, in grid-forming mode, IBRs can maintain frequency and voltage within narrow limits, enabling an autonomous operation of the microgrids (MGs). Grid-forming converters can act as an ideal AC voltage source at the point of common coupling (PCC). They can minimize load interruptions during adverse events and provide black-start services during wide-area blackouts [2].

In MG hierarchical control [3], [4], primary control utilizes droop-based control and virtual impedance control to maintain MG frequency and voltage within acceptable limits, relying on millisecond cycle times. Secondary control compensates for residual deviations in frequency and voltage and operates with response times of milliseconds to seconds. Tertiary control coordinates the MG with the main grid, setting power output setpoints for each distributed generator (DG) based on electricity markets and tariffs. It operates with cycle times of minutes to hours. Automating primary control with local DG data is possible, but secondary and tertiary control require

Andres Intriago and Charalambos Konstantinou are with the Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia (e-mail: firstname.lastname@kaust.edu.sa).

Francesco Liberati is with the Department of Computer, Control and Management Engineering "Antonio Ruberti" (DIAG), University of Rome "La Sapienza", Via Ariosto, 25, 00185 Rome, Italy (e-mail: liberati@diag.uniroma1.it).

Nikos D. Hatziargyriou is with the School of Electrical and Computer Engineering, National Technical University of Athens, 15773 Athens, Greece.

communication infrastructure, adding cost and complexity to the cyber-physical system (CPS). Cyber-attacks can target MG communication controls, making robust and resilient control policies necessary to mitigate disruption of the MG operating state [5], [6].

B. Review of Previous Work

Over the past ten years, a significant number of research articles have been written on detecting MG attacks, and various innovative initiatives have been carried out to showcase the advantages of IBR MGs in trial runs. After analyzing the landscape in the literature, we have identified three pillars that should be employed when reviewing this prior research.

The first criterion involves the model utilized for the attack modeling due to the attack nature and based on which previous work can be broadly classified into two categories:

1.1) Work focusing on denial-of-service (DoS) [7]. Interference with data transmitted by sensors, actuators, and control systems can lead to DoS. Attackers may flood communication channels with random data or jamming them, preventing devices from transmitting and/or receiving information. This has been documented in several studies [8]–[10].

1.2) Work focusing on false data injection (deception/integrity) [11]. Malicious adversaries can perform deception attacks by accessing and replacing real measurements in a system with fake information to cause harm. Such attacks are challenging to identify as attackers can quickly access data, but data veracity is hard to confirm [12], [13]. Intermittent injection of false data is an overlooked problem that can significantly affect the stealthiness of an attack and the energy consumed by the attacker. A self-generated approach in [14] generates specific false data to achieve perfect stealthiness. Notably, recent studies have shown that deception attacks are more challenging to identify than DoS attacks [1].

DoS and intermittent attacks can compromise information integrity and result in data loss. Intermittent attacks involve piece-wise signals and optimal energy management achieved through scheduling DoS attack time instants [15]–[17]. Zero-dynamics, replay, jamming, and covert attack models are used to simulate stealthy and intermittent integrity attacks, becoming difficult to detect by anomaly detectors [18]–[21]. In replay attacks, adversaries record data from compromised sensors during a steady state and replay it while concealing the attack effects [22].

The second criterion involves the approach adopted for attack detection. Based on this criterion, previous work can be broadly classified into two categories, consisting of model-based and data-driven techniques:

2.1) Work focusing on model-based approaches [23]–[40], which rely on a mathematical model of a system. Estimation

methods are often employed to observe the state of the system and create analytical redundancy for detecting attacks. Common techniques include Kalman filtering methods, such as the Kalman filter (KF) [23]–[25] and the unscented KF (UKF) [26], as well as observer-based methods [27]–[38] and sliding mode observation [39], [40].

In terms of KF methods, in [24], an optimal attack strategy is developed for a linear CPS using linear quadratic Gaussian (LQG) and KF to control and estimate the state. A distributed KF approach is introduced in [25] to address CPSs with bandwidth constraints. A resilient state estimator using L_0-norm-based state observer is presented in [29]. Distributed observer-based approach is proposed in [30] for the mitigation of cyber-attacks on CPS sensors and actuators. An event-triggered and observer-based control frame for detecting DoS attacks in CPSs is developed in [31]. Finally, [40] introduces an attack isolation method based on sliding-mode technique for linear systems with unknown inputs.

Regarding observer-based detection techniques for MGs, Rana *et al.* [32] presented a KF-based observer to recover states corrupted by attacks or random noise in a MG. In [33], an unknown input observer within the secondary frequency controller is developed to maintain MG stability under cyber-attacks. An observer-based algorithm that uses real-time phasor measurements to detect and mitigate attacks is presented in [34]. Sahoo *et al.* [35] considered an adaptive and distributed nonlinear observer to detect, reconstruct, and mitigate false data injection attacks in MGs. Other studies focused on designing observers for Lipschitz nonlinear systems [36]–[38], or used sliding mode nonlinear observers [39].

2.2) Work focusing on model-free, data-driven detection approaches [41]–[49]. Data-driven methods use historical system data to detect attacks when a mathematical model is unavailable or contains parameter uncertainties. They commonly employ intelligent methods such as neural networks (NNs) [41], support vector machines (SVMs) [42], and naive Bayesian classifiers [43]. However, the accuracy of these methods depends on the quality of historical information, data noise, and various operating and attack conditions. In the following, we survey representative data-based works.

In the context of intelligent techniques, Li *et al.* proposed an NN-based approach that combines a physical model and a generative adversarial network to detect deviations in measurements [44]. In [45], a deep learning method using convolutional NNs is suggested to detect DoS in cellular networks caused by flooding, signaling, and silent calls. In [46], a method is proposed for constructing an attack detector in MGs by identifying the stable kernel representation in the absence of attacks. In [47], the authors introduced a Bayesian-based approach to attack detection that uses the hybrid Bernoulli random set method to jointly estimate states and detect attacks. In [48], a two-stage approach is presented for false data detection that reduces data dimensionality and detects malicious activities via an SVM classifier. In [49], the authors developed a multi-agent supervised attack detection that employs SVM with a decision tree for each agent, with final decisions made through consensus among all agents.

The third criterion involves the system considerations of the modeling approach and subsequently the realistic operating conditions. Based on this criterion, previous work can be broadly classified into two categories:

3.1) Work focusing on considering MG network models and corresponding operation constraints. Specifically, optimal power flow (OPF) methods have been widely applied for MG operation. In [50], a grid-connected MG OPF method is presented that optimizes energy storage while treating renewable energy sources (RES) and loads as fixed power injections. In [51], the authors propose a two-stage optimization method for energy management of islanded MGs involving a mixed integer programming unit commitment problem followed by an unbalanced three-phase OPF problem. Building on this work, [52] extends the first-stage unit commitment to a robust optimization problem to address uncertainties in RES. For distributed MGs, [53] introduces a method that relaxes power flow constraints to reform the non-convex OPF model. The authors then transform the centralized optimization into a distributed one using predictor-corrector proximal multipliers.

3.2) Work focusing on stability-constrained network modeling in MGs. Stability is a significant concern in MGs, in addition to steady-state security constraints. This is mainly due to the intermittent nature and fluctuating output of RES, and the reduction of system inertia by DGs, which are interfaced by inverters, leading to lower resistance to disturbances in IBR-interfaced MGs. However, few OPF models in previous works have integrated stability constraints. A single machine equivalent method is used to obtain a bus equivalent rotor angular trajectory as a stability constraint in a trajectory-based transient stability-constrained OPF [54]. Conventional power systems have studied voltage stability-constrained OPFs [55], [56]. In MGs, a nonlinear optimization-based voltage stability-constrained OPF is created in [57] to enhance multi-MG voltage stability. Furthermore, a small-signal stability-constrained OPF problem has been studied extensively in traditional power systems [58], and recently in MGs [59].

C. Paper Motivation and Contributions

The challenges towards the security of cyber-physical MGs have increased with the inclusion of more resources and services to the already complex, multi-layered MGs with information and communication technologies. Based on the review of previous work, we have identified a significant research gap that drives the motivation behind this paper: no previous work has modelled and analyzed stealthy intermittent integrity attacks, adopting an observer-based approach for detection and mitigation, and employing within MG modeling both network and stability operation constraints. In the context of filling this knowledge gap, the following contributions are achieved:

(1) We analyze and formulate the state space dynamics of a generic DG in a MG system and consider integrity attacks performed in an intermittent mode, i.e., intermittent integrity attacks. In contrast to the approach used for conventional continuous integrity attacks, intermittent integrity attacks follow a two-step process that involves creating a covert attack model and scheduling the activation and pause times for the attack. The attacker goal is to maximize potential damage in order to

cancel secondary control objectives in removing deviations in both MG global frequency and local voltage.

(2) We design a nonlinear observer to detect such stealthy attacks, and to mitigate their impact by addressing the corrupted signals with their corresponding estimation from the observer. Rather than designing the observer based on nonlinear change of coordinates and the use of "canonical forms" (or "normal forms", see, e.g., [27]), the proposed design is based on simpler Luenberger observers, properly extended to nonlinear systems. Finding a globally defined change of coordinates to put the system into a normal form is a particularly challenging task for the system at study since it is a multiple-output system, with more outputs than inputs, and with a high number of state variables (as explained, e.g., in [27, Chapter 4]). Our work instead is a contribution in the lines of the papers presented in [36], [37], and more recently, in [38], on Luenberger-like observers for nonlinear systems. Differently from the above works, in the present paper, a nonlinear gain is considered in the design of the observer, which is a novelty that allows for greater flexibility in the proof of the convergence of the estimate, achieved via Lyapunov arguments. Also, a nonlinear gain allows us to better tailor the design of the observer to the specific dynamics of MG, reducing the number of conservative boundings in the proof (Theorem 2).

(3) We present the OPF problem for the considered MG and then identify the worst-case cyber-attack in the formulation of a bi-level optimization problem. The adversary adopts intermittent integrity attacks while satisfying the OPF of the system operation. Operators, to ensure a stable operation of the MG, react by incorporating both stability constraints as well as the design of the residual-based detection observer. These conditions ensure not only the stable and optimal operation of the MG but also guarantee the detection of cyber-attacks.

(4) Case studies on a MG benchmark serve a twofold purpose: *a)* validating that the proposed design successfully encapsulates the stealthy nature of attack design in terms of impact assessment, and *b)* demonstrating that the proposed observer-based method achieves reasonable detection accuracy with respect to other observer-based approaches while satisfying the network and stability constraints of the system.

D. Paper Organization

The remainder of this paper is structured as follows. Section II presents the modeling of an islanded MG. Section III presents the attack model formulation, with a discussion on the scheduling interval and stealthiness of the intermittent integrity attack. Section IV presents the formulation of the proposed observer-based mitigation method as well as the attack detectability analysis. Section V presents the stability-constrained mitigation formulation. Section VI presents simulation results, and Section VII concludes the paper.

E. Notation

I_n and 0_n denote identity and zero matrices of dimension $n \times n$, respectively. n_v denotes the dimension of the generic vector v. $[A, B]$ denotes the horizontal concatenation of matrices A and B. $\mathrm{blockdiag}(A_1, ..., A_n)$ denotes the block diagonal matrix having as diagonal blocks the matrices
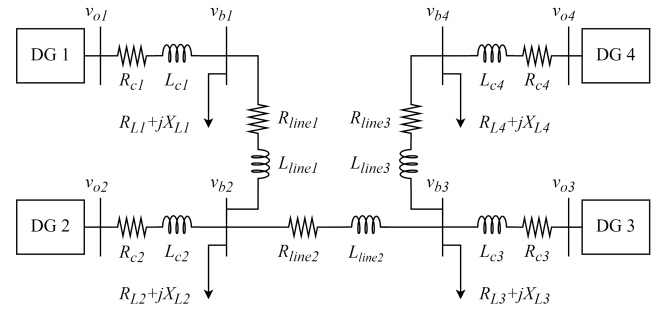

Fig. 1: MG test system considered in this study [3].

$A_1, ..., A_n$. $A_{i,j}$ denotes the entry at row i and column j of matrix A. For a linear map $A : \mathcal{X} \longrightarrow \mathcal{Y}$, we define $\ker(A) \triangleq \{x \in \mathcal{X} \mid Ax = 0\}$ and $\mathrm{Im}(A) \triangleq \{Ax \mid x \in \mathcal{X}\}$. Vector e_i denotes the i_{th} vector of the canonical base. $\mathrm{convh}(x, y)$ denotes the convex hull of the two vectors x and y. $\|v\|$ denotes the Euclidean norm (l^2) of vector v, i.e., $\|v\| \triangleq \sqrt{\sum_i v_i^2}$. $|a|$ denotes the absolute value of number a, or the cardinality of a set, depending on the context.

II. CYBER-PHYSICAL STRUCTURE OF ISLANDED MICROGRID

In this part, the mathematical model of the DG-based MG system, also instrumental for the design of the proposed nonlinear observer, is presented. The single-line diagram in Fig. 1 shows the islanded MG considered in this paper.

A. Cyber Model of Islanded Microgrid

An islanded or autonomous MG can be considered as a CPS with DGs, loads, and a communication network for monitoring and control. A directed graph (digraph) is used to describe the communication network of a multi-agent cooperative system. The digraph is modelled as $\mathcal{G}_r = (\mathrm{V}_\mathcal{G}, \mathrm{E}_\mathcal{G}, \mathrm{A}_\mathcal{G})$, where $\mathrm{V}_\mathcal{G} = \{\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_N\}$ is a nonempty finite collection of N nodes, $\mathrm{E}_\mathcal{G} \subseteq \{\mathrm{V}_\mathcal{G} \times \mathrm{V}_\mathcal{G}\}$ is a set of edges, and $\mathrm{A}_\mathcal{G} = [\mathrm{a}_{ij}] \in \mathbb{R}^{N \times N}$ is the associated adjacency matrix. In the adopted model, the DGs are the nodes of the digraph, whereas the communication links are the edges of the communication network. For this study, it is assumed that the DGs communicate with each other through the digraph presented in the case study of [3], and the term $\mathrm{A}_\mathcal{G}$ remains constant since the digraph is assumed to be invariant in time.

B. Physical Model of Islanded Microgrid

The AC microgrid has $\mathcal{N} = \{1, 2, \ldots, n\}$ set of buses on the power distribution network, $\mathcal{L} = \{1, 2, \ldots, l\} \subseteq \mathcal{N} \times \mathcal{N}$ set of distribution lines, and $\mathcal{G} = \{1, 2, \ldots, n^\mathrm{g}\}$ set of inverters at every DG. Each DG is considered to incorporate a primary and a secondary controller. The droop control for the active (P_i) and the reactive power (Q_i) are described in [3].

The synchronization goals for the secondary voltage and frequency controllers are to set the control input V_{ni} such that the d-axis component of the voltage after the LC filter, v_{odi}, reaches the reference value for the secondary controller, v_{ref}, and to set the control input ω_{ni} to ensure that the frequency at the i_{th} inverter, ω_i, reaches the reference frequency for the secondary controller, ω_{ref}. The voltage magnitude of
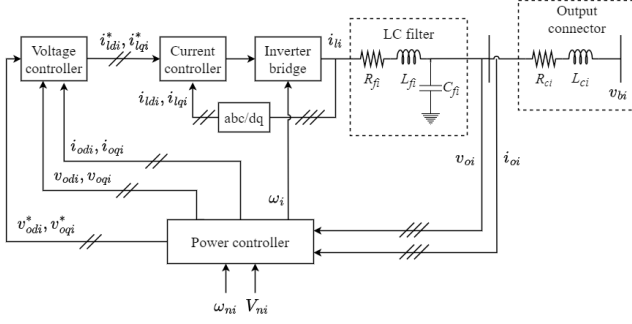
Fig. 2: Voltage-controlled voltage source inverter (VCVSI) DG.

the i_{th} DG is $\|v_o\| = \sqrt{v_{odi}^2 + v_{oqi}^2}$. The voltage-controlled voltage source inverter (VCVSI) DG is shown in Fig. 2. The functionalities of different blocks are explained briefly below.

$\underline{\delta_i \text{ calculator:}}$ The angle of the i_{th} inverter-based DGs with respect to a common reference, typically the inverter at bus 1 (i.e., $\omega_{com} = \omega_1$) is δ_i, and satisfies $\dot{\delta}_i = \omega_i - \omega_{com}$. The purpose of this angle is to transform the periodic and balanced three-phase signal abc to a dq reference frame, while ensuring that v_{oqi} reaches zero.

$\underline{\text{Power controller:}}$ This block includes the droop controller and a low-pass filter to extract the fundamental components of P_i and Q_i. The outputs of this block are v_{odi}^*, v_{oqi}^*, and ω_i [3].

$\underline{\text{Voltage and current controller:}}$ The dynamics of the voltage and current controllers can be found in [3]. In both controllers, auxiliary state variables, ϕ_{di}, ϕ_{qi} and γ_{di}, γ_{qi}, respectively, are introduced to formulate the complete state-space model, where $\dot{\phi}_{di} = v_{odi}^* - v_{odi}$, $\dot{\phi}_{qi} = v_{oqi}^* - v_{oqi}$, $\dot{\gamma}_{di} = i_{ldi}^* - i_{ldi}$, and $\dot{\gamma}_{qi} = i_{lqi}^* - i_{lqi}$.

$\underline{\text{LC filter and output connector:}}$ The d-axis and q-axis equations for the LC filter and output connector are provided in [3], and are considered in the state-space model.

As for the distributed frequency and voltage controller design, the details can be found in [4]. Overall, the objective of frequency control for VCVSIs is to achieve synchronization with the nominal frequency. In addition, it should distribute the output P_i of VCVSIs based on their P_i ratings. By taking the derivative of the frequency droop characteristic ($\omega_i = \omega_{ni} - m_{Pi}P_i$), the rate of change of the primary control frequency reference ω_{ni} is expressed as $\dot{\omega}_{ni} = \dot{\omega}_i + m_{Pi}\dot{P}_i = v_{fi}$, where v_{fi} represents the auxiliary control input defined in (30) of [4]. $\dot{\omega}_{ni}$ is utilized to calculate the control input ω_{ni} from v_{fi}, which involves integration of v_{fi} over time. Similarly, the distributed voltage control mechanism selects appropriate control inputs V_{ni} in the reactive droop ($|v_o|^* = V_{ni} - n_{Qi}Q_i$) to ensure synchronization of the voltage magnitudes $|v_o|$ of VCVSIs with the reference voltage v_{ref}. Additionally, it distributes the Q_i of VCVSIs based on their Q_i ratings. Synchronization of the voltage magnitudes $|v_o|$ is equivalent to synchronization of the direct term of output voltages v_{odi}. By differentiating the voltage droop characteristic, the rate of change of V_{ni} is expressed as $\dot{V}_{ni} = \dot{v}_{odi} + n_{Qi}\dot{Q}_i = v_{vi}$, where v_{vi} denotes the auxiliary control input defined in (48) of [4]. $\dot{V}_{ni}$ is utilized to calculate the control input V_{ni} from v_{vi} by integration over time. In our work, ω_{ni} and V_{ni} will be considered as the outputs of the secondary controller, to

follow the same structure as previous works, with m_{Pi} and n_{Qi} being the active and reactive power droop coefficients that are selected based on the P_i and Q_i ratings of each VCVSI.

C. Full State Space Model

The state space dynamics of the generic i_{th} DG over time t are obtained by gathering and manipulating Eqs. (1) - (20) of [3]. The complete model can be written in matrix form as:

$$W : \begin{cases} \dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}) + B\boldsymbol{u} \\ \boldsymbol{y} = C\boldsymbol{x} \end{cases} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^{n_x}$ is the state vector, $\boldsymbol{u} \in \mathbb{R}^{n_u}$ is the input vector generated by the power controller. The vector function $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{n_x}$ captures the nonlinearity of the system and $\boldsymbol{y} \in \mathbb{R}^{n_y}$ is the vector of sensor measurements at the DG. The state vector $\boldsymbol{x}$ with $n_x = 15$ in (1) is defined as:

$$\begin{aligned} \boldsymbol{x} = &[P_i, Q_i, \phi_{di}, \phi_{qi}, \gamma_{di}, \gamma_{qi}, i_{ldi}, i_{lqi}, v_{odi}, v_{oqi}, \\ &i_{odi}, i_{oqi}, \delta_i, \omega_{ni}, V_{ni}]^\top. \end{aligned} \tag{2}$$

The input vector $\boldsymbol{u}$, with $n_u = 9$, in (1) where u_2 and u_3 describes the voltage at the i_{th} bus DG measured after the output connector (Fig. 2) in terms of d and q axes components, respectively. It also includes in u_5 and u_8 the setpoints for the secondary controller ω_{ref} and v_{ref}, while the remaining elements u_4, u_6, u_7, and u_9 represent the data collected from the set of neighbouring/adjacent DGs in the MG $j \neq i$, i.e., ω_j, P_j, v_{odj}, and Q_j, respectively, where c_{fi} and c_{vi} are control gains, and g_i the pinning gain. It is defined as:

$$\begin{aligned} \boldsymbol{u} = &[\omega_{com}, v_{bdi}, v_{bqi}, c_{fi} \sum_{j \in N_{j \neq i}} a_{ij}\omega_j, c_{fi}g_i\omega_{ref}, \\ &c_{fi} \sum_{j \in N_{j \neq i}} a_{ij}m_{Pj}P_j, c_{vi} \sum_{j \in N_{j \neq i}} a_{ij}v_{odj}, \\ &c_{vi}g_i v_{ref}, c_{vi}g_i v_{ref}, c_{vi} \sum_{j \in N_{j \neq i}} a_{ij}n_{Qj}Q_j]^\top. \end{aligned} \tag{3}$$

Remark 1 (Measured Variables). *The DG measured variables are assumed to be the generated active and reactive powers, i.e., the state variables x_1 and x_2 (as mentioned, e.g., in [3, see after Fig. 1]), the inverter bridge current i_l, the output voltage v_{oi}, and the output current i_{oi}, and their direct and quadrature components, i.e., respectively, the state variables x_7, x_8, x_9, x_{10}, x_{11}, and x_{12} (see, e.g., [3] and [35]); notice in fact from Fig. 2 that the measurement of those variables are available and fed-back to the current, voltage, and power controllers. The secondary control inputs ω_{ni} and V_{ni} are known, since they are the inputs to the power controller (they are computed according to, respectively, (20) and (46) in [3]). Finally, the DG operating frequency, ω_i is also measured (since it has to be communicated to the neighbour DGs for distributed secondary control purpose - see, e.g., [4, eq. (30)]). In conclusion, the state variables assumed measurable are: x_1, x_2, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{14}, and x_{15}.*

The fifteen differential equations in (1), and the resulting matrices/vectors A, B, C, and $\boldsymbol{f}(\boldsymbol{x})$ are detailed in the following subsection II-D. The model of (1) is used for the design of the observer in Section (IV).

D. Detailed DG State Space Model

The state vector is defined in (2). The complete state space model of the generic i_{th} DG is given in (1) in matrix form. In scalar form, the fifteen differential state equations can be obtained by manipulating (1) – (20) of [3] as follows:

$$\dot{x}_1 = -\omega_{ci}x_1 + \omega_{ci}x_9x_{11} + \omega_{ci}x_{10}x_{12} \tag{4}$$

$$\dot{x}_2 = -\omega_{ci}x_2 - \omega_{ci}x_9x_{12} + \omega_{ci}x_{10}x_{11} \tag{5}$$

$$\dot{x}_3 = -n_{Qi}x_2 - x_9 + x_{15} \tag{6}$$

$$\dot{x}_4 = -x_{10} \tag{7}$$

$$\dot{x}_5 = -K_{PVi}n_{Qi}x_2 + K_{IVi}x_3 - x_7 - K_{PVi}x_9 + \\ -\omega_b C_{fi}x_{10} + F_i x_{11} + K_{PVi}x_{15} \tag{8}$$

$$\dot{x}_6 = K_{IVi}x_4 - x_8 + \omega_b C_{fi}x_9 - K_{PVi}x_{10} + F_i x_{12} \tag{9}$$

$$\begin{aligned}\dot{x}_7 =& -\frac{K_{PCi}K_{PVi}}{L_{fi}}n_{Qi}x_2 + \frac{K_{PCi}K_{IVi}}{L_{fi}}x_3 + \\ &+\frac{K_{ICi}}{L_{fi}}x_5 - \Big(\frac{R_{fi}}{L_{fi}} + \frac{K_{PCi}}{L_{fi}}\Big)x_7 - \omega_b x_8 + \\ &-\Big(\frac{1}{L_{fi}} + \frac{K_{PCi}K_{PVi}}{L_{fi}}\Big)x_9 - \omega_b \frac{K_{PCi}}{L_{fi}}C_{fi}x_{10} + \\ &+\frac{K_{PC_i}}{L_{fi}}F_i x_{11} + \frac{K_{PCi}K_{PVi}}{L_{fi}}x_{15} + \\ &-m_{P_i}x_1 x_8 + x_8 x_{14}\end{aligned} \tag{10}$$

$$\begin{aligned}\dot{x}_8 =& \frac{K_{PCi}}{L_{fi}}K_{IVi}x_4 + \frac{K_{ICi}}{L_{fi}}x_6 + \omega_b x_7 + \\ &-\Big(\frac{R_{fi}}{L_{fi}} + \frac{K_{PCi}}{L_{fi}}\Big)x_8 + \omega_b \frac{K_{PCi}}{L_{fi}}C_{fi}x_9 + \\ &-\Big(\frac{K_{PCi}K_{PVi}}{L_{fi}} + \frac{1}{L_{fi}}\Big)x_{10} + \frac{K_{PCi}}{L_{fi}}F_i x_{12} + \\ &+m_{P_i}x_1 x_7 - x_7 x_{14}\end{aligned} \tag{11}$$

$$\dot{x}_9 = \frac{1}{C_{fi}}x_7 - \frac{1}{C_{fi}}x_{11} - m_{P_i}x_1 x_{10} + x_{10}x_{14} \tag{12}$$

$$\dot{x}_{10} = \frac{1}{C_{fi}}x_8 - \frac{1}{C_{fi}}x_{12} + m_{P_i}x_1 x_9 - x_9 x_{14} \tag{13}$$

$$\dot{x}_{11} = \frac{1}{L_{ci}}x_9 - \frac{R_{ci}}{L_{ci}}x_{11} - \frac{1}{L_{ci}}v_{bdi} - m_{P_i}x_1 x_{12} + x_{12}x_{14} \tag{14}$$

$$\dot{x}_{12} = \frac{1}{L_{ci}}x_{10} - \frac{R_{ci}}{L_{ci}}x_{12} - \frac{1}{L_{ci}}v_{bqi} + m_{P_i}x_1 x_{11} - x_{11}x_{14} \tag{15}$$

$$\dot{x}_{13} = -m_{P_i}x_1 + x_{14} - \omega_{com} \tag{16}$$

$$\begin{aligned}\dot{x}_{14} =& \Big[m_{P_i}c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij} + g_i\Big) - c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij}m_{P_i}\Big)\Big]x_1 + \\ &-c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij} + g_i\Big)x_{14} + \\ &+c_{fi}\sum_{j\in N_i}\mathrm{a}_{ij}\omega_j + c_{fi}g_i\omega_{ref} + \\ &+c_{fi}\sum_{j\in N_i}\mathrm{a}_{ij}m_{P_j}P_j\end{aligned} \tag{17}$$

$$\begin{aligned}\dot{x}_{15} =& -c_{vi}n_{Qi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij}\Big)x_2 - c_{vi}\Big(\sum_j \mathrm{a}_{ij} + g_i\Big)x_9 + \\ &+c_{vi}\sum_j \mathrm{a}_{ij}v_{odj} + c_{vi}g_i v_{ref} + \\ &+c_{vi}\sum_j \mathrm{a}_{ij}n_{Q_j}Q_j\end{aligned} \tag{18}$$

In the above equations, other parameters involve the nominal angular frequency of the MG, ω_b, the cut-off frequency of the low-pass filter in the power controller, ω_{ci}, gains of voltage and frequency control, c_{vi} and c_{fi}, the current compensator F_i, and impedance characteristics of the LC filter and the output connector. The equations are compactly written in (1) in matrix form, in which A, $\boldsymbol{f}(\boldsymbol{x})$, B, and C can be inferred from inspection of the equations and are as follows.

The elements of matrix A are: $A_{1,1} = A_{2,2} = -w_{ci}$, $A_{13,1} = -m_{P_i}$, $A_{3,2} = -n_{Qi}$, $A_{5,2} = -K_{PVi}n_{Qi}$, $A_{5,3} = A_{6,4} = K_{IVi}$, $A_{7,5} = A_{8,6} = \frac{K_{ICi}}{L_{fi}}$, $A_{5,7} = A_{6,8} = A_{3,9} = A_{4,10} = -A_{13,14} = -A_{3,15} = -1$, $A_{8,7} = -A_{7,8} = \omega_b$, $A_{9,7} = A_{10,8} = -A_{9,11} = -A_{10,12} = \frac{1}{C_{fi}}$, $A_{5,9} = A_{6,10} = -A_{5,15} = -K_{PVi}$, $A_{6,9} = \omega_b C_{fi}$, $A_{11,9} = A_{12,10} = \frac{1}{L_{ci}}$, $A_{5,11} = A_{6,12} = F_i$, $A_{7,11} = A_{8,12} = \frac{K_{PCi}}{L_{fi}}F_i$, $A_{11,11} = A_{12,12} = -\frac{R_{ci}}{L_{ci}}$, $A_{14,1} = \Big[m_{P_i}c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij} + g_i\Big) - c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij}m_{P_i}\Big)\Big]$, $A_{7,7} = A_{8,8} = -(\frac{R_{fi}}{L_{fi}} + \frac{K_{PCi}}{L_{fi}})$, $A_{15,2} = -c_{vi}n_{Q_i}\Big(\sum_{j\in N_i}\mathrm{a}_{ij}\Big)$, $A_{7,9} = A_{8,10} = -(\frac{1}{L_{fi}} + \frac{K_{PCi}K_{PVi}}{L_{fi}})$, $A_{15,9} = -c_{vi}\Big(\sum_j \mathrm{a}_{ij} + g_i\Big)$, $A_{7,2} = \frac{K_{PCi}K_{PVi}}{L_{fi}}n_{Qi}$, $A_{7,3} = \frac{K_{PCi}K_{IVi}}{L_{fi}}$, $A_{14,14} = -c_{fi}\Big(\sum_{j\in N_i}\mathrm{a}_{ij} + g_i\Big)$, $A_{7,15} = \frac{K_{PCi}K_{PVi}}{L_{fi}}$, $A_{7,10} = A_{8,9} = -\omega_b\frac{K_{PCi}}{L_{fi}}C_{fi}$, and all the other elements are equal to zero.

The elements of matrix B are: $B_{11,2} = B_{12,3} = -\frac{1}{L_{ci}}$, $B_{13,1} = -1$, $B_{14,4} = B_{14,5} = B_{14,6} = B_{15,7} = B_{15,8} = B_{15,9} = 1$, and all the other elements are equal to zero.

In view of Remark 1, matrix C is as follows:

$$C = \text{blockdiag}(I_2, [0_{6,4}, I_6], [0_{2,1}, I_2]) \tag{19}$$

The elements of $\boldsymbol{f}(\boldsymbol{x})$ are: $f_1 = \omega_{ci}x_9 x_{11} + \omega_{ci}x_{10}x_{12}$, $f_2 = -\omega_{ci}x_9 x_{12} + \omega_{ci}x_{10}x_{11}$, $f_7 = -m_{P_i}x_1 x_8 + x_8 x_{14}$, $f_8 = m_{P_i}x_1 x_7 - x_7 x_{14}$, $f_9 = -m_{P_i}x_1 x_{10} + x_{10}x_{14}$, $f_{10} = m_{P_i}x_1 x_9 - x_9 x_{14}$, $f_{11} = -m_{P_i}x_1 x_{12} + x_{12}x_{14}$, $f_{12} = m_{P_i}x_1 x_{11} - x_{11}x_{14}$, where the matrix $\boldsymbol{M}^{\mathrm{p}}$ is defined in Appendix, and where each m_{P_i} indicates the droop constant of the i_{th} inverter, and all the other elements are zero.

III. ATTACK MODEL FORMULATION

The communication network connecting the participating DGs forming the MG can potentially expose the cyber-physical MG system to cyber-attacks [46]. In literature, several procedures rely on the adversary's system knowledge and the skills to corrupt measurements from remote terminal units (RTUs). For example, in [12], the consequences of false data on RTU measurements are examined at the distributed transient stability control schemes of a MG. Cyber-physical MGs include also smart inverters for renewable-based DERs grid-connection. Recent work has demonstrated that such firmware electronic devices can be tampered leading to system disruption and instabilities [6].

In this paper, the threat model considers an adversary aiming to corrupt the operation of grid converters. The attack entry point could be direct, i.e., at the DG, or indirect, e.g., by modifying the firmware of the inverters. First, the attacker intends to undermine the communication links of the MG by infiltrating compromised data that can harm substantionally
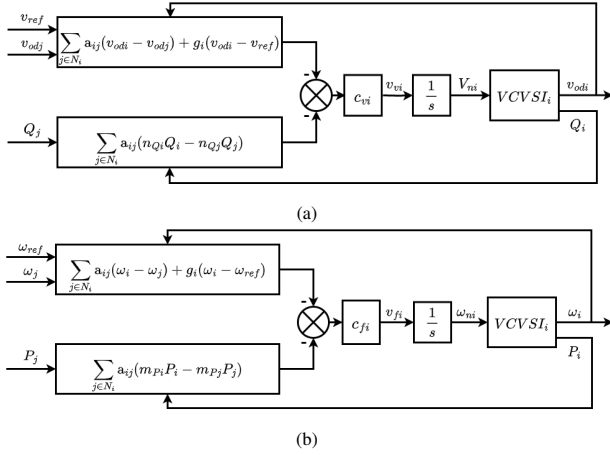
Fig. 3: Distributed voltage (a) and frequency (b) controllers.

the voltage and frequency of all the inverters. The act of introducing malicious data to the inputs of each DG will affect the dynamics of the grid and hence, create instabilities in the MG. *The aim of the attack is to maximize the effect of the disruption at the input control signals of the MG secondary controller, shown in Fig. 3, and thus the states of the MG, while remaining stealthy.* Specifically, the attacker would aim to modify control actions $act_i \in \mathcal{A}_i, \forall i \in \mathcal{G}$, of inverter-based DG control under normal conditions to actions $act_i' \in \mathcal{A}'i, act_i' \neq act_i, \forall i \in \mathcal{G}$, that would harm DGs and/or MG/distribution and transmission systems. Following this modification, the production and consumption at every node are changed. Using this mechanism, the attacker will aim to coordinate actions of compromised signals in the MG, i.e., $\left\{act_{it}' : \prod_{i \in \mathcal{G}} \mathcal{S}_i \rightarrow \prod_{i \in \mathcal{G}} \mathcal{A}'_i\right\}$, defined over a discrete time horizon ($\mathcal{T} = [1, 2, \ldots, t, \ldots, N_T]$), i.e., $\forall t \in \mathcal{T}$, to maximize potential damage in order to cancel secondary control objective in removing deviations in both MG global frequency and local voltage, with the action space $\mathcal{A}$ representing control actions (e.g., secondary control signal disruption, etc.) that influence state space $\mathcal{S}$ and regulate interactions of DGs with the MG system (e.g., withdraw/inject power from/to the system or idle). The attack model assumes that the attacker possesses some, but not necessarily exact, knowledge of the system design (e.g., topology, parameters of lines, buses, generators, etc), which can be obtained from publicly available data sets (e.g., ISO New England [60] and WECC-240 [61]).

Throughout this work, stealthiness is considered with respect to an anomaly detector $\mathcal{D}$, characterized by the residual, and a threshold η for the identification of an anomaly. Current detection approaches within the communication of local controllers (LCs) (primary) and MG centralized controller (MGCC) (secondary) involve using residual-based methods [46]. The residual is often determined from the difference between measurements y and their estimated values $\hat{y}$. As an example, the χ^2-distribution with $n - m$ degrees of freedom and confidence interval can determine the threshold η equal to $\sigma\sqrt{\chi^2_{n-m,\alpha}}$. In the event that $\|y - \hat{y}\|_2 > \eta$, an alarm will be raised.

Proposition 1. *Let us consider the attacked system of Eq.* (1) *becomes as follows based on the aforementioned attack model*

at which an attack starts at an unknown time $t = T_0$ *(i.e., and hence $a(t) = 0$ for $0 \leq t < T_0$) targeting a subset of the input vector $u(t)$ from the neighboring DGs.*

$$W_a : \begin{cases} \dot{x} = Ax + f(x) + Bu + B\Gamma_u a_u \\ y = Cx \end{cases} \quad (20)$$

where $\Gamma_u \in \mathbb{B}^{n_u \times |K_u|}$, $\mathbb{B} \triangleq \{0, 1\}$, is the binary incidence matrix mapping the attack signal along with the mapping of the disturbance vector to the respective channels. The attack signal, incorporating also the disturbance, is $a_u = \left[a_{u,1}(t), \ldots, a_{u,|K_u|}(t)\right]^T \in \mathbb{R}^{|K_u|}$, where $K_u \subseteq \{1, \ldots, n_u\}$ represents the disruption resources available to the attacker able to corrupt the neighboring inputs of the system. For each $i \in \{1, \ldots, |K_u|\}$, $a_{u,i}(t) = 0$ for $t \in \mathbb{R}_+$ if no attack occurs on the i-th transmission channel of $\mathcal{A}_{a_{u_i}}$. In the specific model of our MG, $n_u = 9$ and $|K_u| = 2$.

Taking into account Proposition 1 and the threat model described so far, we consider that the attacker has partial model knowledge (A, B, Γ, C), available disruption resources (K), and overall some limited but yet sufficient knowledge about the internal system signals as well as a restrict understanding of the network topology in order to create an adverse disturbance. The goal is to generate a stealthy intermittent integrity attack in which first the attack activation time is identified, and then, the integrity model of the attack is constructed.

A. Scheduling Intervals for Intermittent Integrity Attacks

The time $\forall t \in \mathcal{T}$ in which the attack signal will be activated or not is assumed to be determined by the adversary at the time instants $t_1, \ldots, t_{Na} \in \mathcal{T}$, where $N_a \in \mathbb{Z}^+$. Let us consider a specific time slot k in which the signal is active for a time interval τ_k, i.e., $0 < \tau_k \leq t_{k+1} - t_k$. Thus, the attack is inactive when $t \geq t_k + \tau_k$. Consequently, the time interval where the attack signal will be activated (Θ_k^{ac}) and deactivated (Θ_k^{de}) for the k-th attack slot is the following:

$$\begin{cases} \Theta_k^{ac}, & t_k \leq t < t_k + \tau_k \\ \Theta_k^{de}, & t_k + \tau_k \leq t < \infty \end{cases} \quad (21)$$

The time interval where the k-th attack is implemented can be defined as $\Theta_k = \Theta_k^{ac} \cup \Theta_k^{de}$, and an auxiliary time interval where the attack slot is conducted can be formulated as:

$$\begin{cases} \Theta_k, & t_k \leq t < t_{k+1}, \forall k \in \{1, \ldots, N_a - 1\} \\ \Theta_{N_a}, & t_{N_a} \leq t < \infty \end{cases} \quad (22)$$

B. Attack Model for Stealthy Intermittent Integrity Attacks

Given the attack interval where the signal is activated and deactivated Θ_k, the complete attack model for the k-th slot can be described as:

$$\dot{\zeta}_k(t) = (A + B\Gamma_u Q_k)\zeta_k(t) + B\Gamma_u L_a l(t) \quad (23a)$$

$$\zeta_k(t_k) = -\Delta z_k \quad (23b)$$

$$a_k(t) = \begin{cases} Q_k \zeta_k(t) + L_a l(t), & \forall t \in \Theta_k^{ac} \\ 0, & \forall t \in \Theta_k^{de} \end{cases} \quad (24)$$

$$a(t) = \sum_{i=1}^{k} a_i(t), \quad \forall t \in \Theta_k \quad (25)$$

where the matrix $Q_k \in \mathbb{R}^{(n_u+n_y) \times n_x}$ is generated using Theorem 1. The dimensions of the matrix L_a depend on the orthonormal basis for the inverse map of $\text{Im}(V_a)$ in $B\Gamma_u$, and can be computed with Theorem 1. The vector $l(t)$ can be any signal with proper dimensions. $\Delta z_k \in \mathbb{R}^{n_x}$ shows the difference between the real values of the system state and the states known by the adversary. The initial conditions for Δz_k are usually bounded, where the upper and lower bounds are not specific requirements that the adversary needs to compute to execute the disturbance in the system. Accordingly, the initial condition from (23b), Δz_k, is not zero. This strategy provides realism to the model since the attacker is not required to know the exact value of each state. Furthermore, we consider time slots when the attack signal is deactivated, $t \geq t_k + \tau_k$, because we want to guarantee continuity in the output signal $y(t)$. The convergence of $a(t)$ is also guaranteed if the attack model is built as (21), while the design of $a_k(t)$ ensures that the generated attacks can pass through some statistical anomaly detectors $\mathcal{D}$ [46], [62]. It is clear that the stealthiness of $a(t)$ is based on Δz_k, which is required to be sufficiently small not to be perceived. Δz_k is valid under Assumption 1.

Assumption 1. *There exists two scalars $c_1 > 0$ and $c_2 > 0$ such that the initial condition, described in (23b), is bounded:*

$$c_1 \leq \|\Delta z_k\| \leq c_2, \forall k \in (1, \dots, N_a) \quad (26)$$

where c_1 and c_2 are sufficiently small and not required to be known by the adversary.

Remark 2. *The inferior bound for $\|\Delta z_k\|$ shows that the attacker is not supposed to know the exact value of the states vector because the measurements are usually affected by the noise that comes to the system. The superior bound, on the other hand, is chosen in such a way that it will not be distinguished by the detector $\mathcal{D}$.*

In [17], the input signal $u(t)$ and Δz_k are taken into consideration to evaluate the stealthiness of the attacks. These two parameters are important to establish the convergence of the output $y(t)$ when the disruption starts. Following the structure in [16], we split the system W in (1) during the time interval Θ_k where the attack holds:

$$W_1 : \begin{cases} \dot{x}_1(t) = Ax_1(t), \ \forall t \in \Theta_k \\ y_1(t) = Cx_1(t), \ \forall t \in \Theta_k \end{cases} \quad (27)$$

$$W_2 : \begin{cases} \dot{x}_2(t) = Ax_2(t) + f(t,x_2) + Bu(t), \forall t \in \Theta_k \\ y_2(t) = Cx_2(t), \forall t \in \Theta_k \end{cases} \quad (28)$$

where W_1 and W_2 correspond to the case where the system is working under nominal conditions. The initial conditions are $x_1(t_k) = -\Delta z_k$ and $x_2(t_k) = x(t_k)$. The way the split is done satisfies the condition of $x(t) = x_1(t) + x_2(t)$. In the same way, $y(t) = y_1(t) + y_2(t)$. Following the same technique, we can split our system W_a in (20), when the attacker injects a malicious signal $a(t)$ into the communication of the DGs:

$$W_{1_a} : \begin{cases} \dot{x}_{1_a}(t) = Ax_{1_a}(t) + B\Gamma_u a(t) \\ y_{1_a}(t) = Cx_{1_a}(t) + D'a(t) \end{cases} \quad (29)$$

$$W_{2_a} : \begin{cases} \dot{x}_{2_a}(t) = Ax_{2_a}(t) + f(t,x) - f(t,x_n) + Bu(t) \\ y_{2_a}(t) = Cx_{2_a}(t) \end{cases} \quad (30)$$

where x_n represents the state vector in nominal conditions and $D' = [0_{n_y \times |K_u|}, \Gamma_y]$, where $\Gamma_y \in \mathbb{B}^{n_y \times |K_y|}$. $a(t) = \beta_i(t-T_o)a_u(t)$, where $\beta_i(t-T_o)$ describes the attack function dynamics and is given by:

$$\beta_i(t - T_o) \equiv \begin{cases} 0, & \forall t \in \Theta_k^{de} \\ 1 - e^{-b_i(t-T_o)}, & \forall t \in \Theta_k^{ac}, \end{cases} \quad (31)$$

with parameter $b_i \in \mathbb{R}$ representing the attack evolution rate. The initial conditions are $x_1(t_k) = -\Delta z_k$ and $x_2(t_k) = x(t_k) - \Delta z_k$. The splitting only works if the summation of W_{a_1} and W_{a_2} yields W_a. The computation of the matrix parameter Q_k and the generation of several subspaces are discussed below.

In order to calculate the largest controlled invariant subspace of W_{1_a} under an attack signal $a(t)$, the nonlinear function $f(t,x)$ of W_{2_a} in (30) needs to satisfy:

$$f(t,x) = f(t, x_n + x_{2_a}) \quad (32)$$

Using (32) and the mean value theorem in [16], [28], the difference $\Delta f = f(t,x) - f(t, x_n + x_{2_a})$ can be described as $\Delta f = F(t,\rho)x_{1_a}(t)$, where $\rho = \phi(x, x_n + x_{2_a}) \triangleq [\rho_1, \dots, \rho_{n_x}] \in \mathbb{R}^{n_x \times n_x}$ with $\rho_i \in \text{convh}(x, x_n + x_{2_a})$ for $i = 1, \dots, n_x$, and $F(t,\rho)$ as:

$$F(t,\rho) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(t,\rho_1) & \cdots & \frac{\partial f_1}{\partial x_{n_x}}(t,\rho_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_x}}{\partial x_1}(t,\rho_{n_x}) & \cdots & \frac{\partial f_{n_x}}{\partial x_{n_x}}(t,\rho_{n_x}) \end{bmatrix}$$

The theory behind the decoupling of the largest controlled invariant subspace and $F(t,\rho)$ can be found in [16].

Theorem 1 (Stealthiness). *Let us consider there is a weakly unobservable subspace generated by the incremental system W_{a_1} in (29), where this subspace can be represented as $V(W_1^a)$, the unobservable subspace created by the pair (C, A), denoted by $H \subset \mathbb{R}^{n_x}$, and the largest controlled invariant subspace of W_1^a in H, denoted by $V(H)$.*

$$V_a = V(W_{a_1}) \cap V(H) \quad (33)$$

Then, there is a matrix Q_k, such that:

$$(A + B\Gamma_u Q_k)V_a \subset V_a \quad (34)$$

$$\Delta z_k \in V_a, \forall k \in (1, \dots, N_a) \quad (35)$$

Finally, there is a matrix L_a that represents the attack generator gain as follows:

$$\text{Im}(L_a) = \ker(D') \cap (B\Gamma_u)^{-1} V_a \quad (36)$$

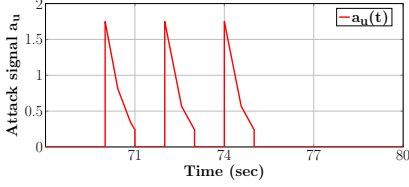*Proof.* The proof for Theorem 1 can be found in [16], [17]. ∎

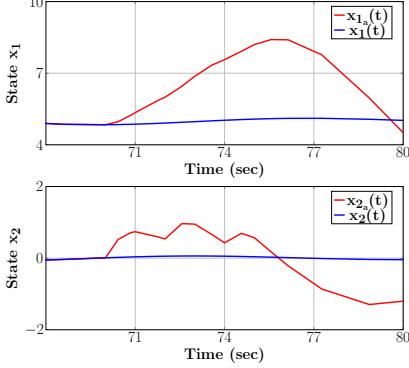Fig. 4: Time responses of the attack signal $a_u(t)$.

Fig. 5: Time responses of the state vector $x(t)$ under nominal $x_n(t)$ and attack $x_{n_a}(t)$ scenarios.

C. Numerical Simulation Example

This section provides an illustration of a numerical simulation using a linear time-invariant state space representation of (20). The matrices of the system are specified as follows:

$$A = \begin{bmatrix} 0 & 1 \\ -0.2 & -0.1 \end{bmatrix}, \ B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

where the pair (A, C) is observable.

In order to demonstrate the nature of the stealthy intermittent integrity attack, we consider an adversary able to corrupt all the input-related disruption resources, i.e., $\Gamma_u = 1$. We also assume that the attack activating time instants t_k are given first as follows: $t_1 = 70s$, $t_2 = 72s$, $t_3 = 74s$, with the same interval time being used when the attack signal is inactive, i.e., $\tau_k = 1s$, $\forall k \in \{1, 2, 3\}$. Following Theorem 1 and utilizing the geometric approach toolbox [63], we can obtain the weakly unobservable subspace V_a and the matrix Q_k as follows:

$$V_a = \begin{bmatrix} -0.4472 \\ 0.8944 \end{bmatrix}, \ Q_k = \begin{bmatrix} 0.8 & -1.6 \\ 0.4 & -0.8 \end{bmatrix}$$

The initial conditions Δ_{z_k} to generate the stealthy attack while satisfying (35) are chosen as:

$$\Delta_{z_1} = \begin{bmatrix} -0.2236 \\ 0.4472 \end{bmatrix}, \ \Delta_{z_2} = \begin{bmatrix} -0.1118 \\ 0.2236 \end{bmatrix}, \ \Delta_{z_3} = \begin{bmatrix} -0.0559 \\ 0.1118 \end{bmatrix}$$

The effects of the attack signal on the system are shown in Figs. 4 and 5. The attack signal a_u introduced into the system input is intermittent and activated during specific time intervals. By comparing the system states at nominal and during attack scenarios, we can observe that, in this numerical example, the stealthy attack compromises the stability of the state vector. It can be observed from Fig. 5 that the system state $x(t)$ and its variation do not exhibit any abrupt changes during the attack pausing time instants. Detecting this attack can be challenging, especially when measurement noises and disturbances are present in the system.
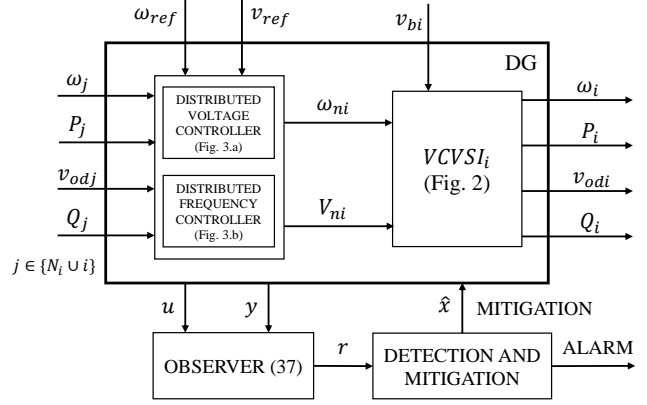

Fig. 6: Proposed observer-based attack detection and mitigation architecture for the generic DG.

IV. RESIDUAL-BASED DETECTION OBSERVER

In this section, a framework is proposed to detect and mitigate the attacks previously introduced. The proposed detection architecture is illustrated in Fig. 6. An observer is deployed at each DG. On the left of the figure, the measurements transmitted from the connected DGs are represented. The detection method is residual-based: it compares the value of the measured variables, which is possibly affected by an attack, with the estimation of the value that the variable would have in the absence of the attack, as obtained through the use of the proposed observer. In nominal conditions, the two values coincide. When an attack occurs, the error between the two increases, making it possible to detect the presence of the attack. The observer's inputs include the measured variables, y, and the inputs of the DG, u, possibly subject to attacks.

Inspired by the works in [36], [37], which address also observers for systems having a structure like (1), this work proposes to use the following Luenberger-like observer:

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + f(\hat{x}, y) + Bu + L(\hat{x})(y - \hat{y}) \\ \hat{y} = C\hat{x}, \end{cases} \tag{37}$$

where u and y are, respectively, the inputs and the outputs of the DG, which are both input signals for the observer. Later on, in IV-A we will present the nomenclature to distinguish the attack case from the normal operation. Notice that the observer is essentially a copy of the system, plus a correction term. Notice also that f and L in the observer may depend in general on the output of the system and on the state of the observer. Different observers arise depending on the choice of f and L. Differently from the above-mentioned works [36], [37], in this paper we consider a nonlinear gain $L(\hat{x})$ in the correction term in (37), which will allow reducing the number of boundings done in the convergence proof for the observer, leading to less conservative results. It is proved in the following that, with a particular choice of $L(\hat{x})$, in the absence of attacks on the state and the input measurements, the state of the observer, $\hat{x}$, tracks the state of the system, x. This is used then to reveal the presence of attacks, by analyzing the residual vector $r \triangleq y - \hat{y}$.

Theorem 2 (Convergence of the observer estimate). *There exists a choice of $L(\hat{x})$ in (37) such that the state $\hat{x}$ of the observer (37) globally asymptotically tracks the state x of system (1) in absence of attacks.*

Proof. The result is proved by showing that there exists a choice of $L(\hat{x})$ such that the dynamics of the estimation error $\boldsymbol{\xi} \triangleq \boldsymbol{x} - \hat{\boldsymbol{x}}$ is globally asymptotically stable (i.e., $\boldsymbol{\xi}$ converges to $\mathbf{0}$, for any initial condition).

The derivative of $\boldsymbol{\xi}$ is:

$$\dot{\boldsymbol{\xi}} = \dot{\boldsymbol{x}} - \dot{\hat{\boldsymbol{x}}} = A\boldsymbol{\xi} + \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}}) - L(\hat{\boldsymbol{x}})C\boldsymbol{\xi}. \quad (38)$$

Without loss of generality, one can write the nonlinear gain term $L(\hat{x})$ as $L(\hat{x}) \triangleq L' + L''(\hat{x})$, so that (38) becomes

$$\dot{\boldsymbol{\xi}} = (A - L'C)\boldsymbol{\xi} + \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}}) - L''(\hat{\boldsymbol{x}})C\boldsymbol{\xi}. \quad (39)$$

The stability of the dynamics of $\boldsymbol{\xi}$ can proved via classical Lyapunov arguments (see, e.g., [64] [Theorem 4.2]). The proof is constructive, i.e., a particular choice for L' and L'' will be derived, which results in a stable observer. Take the positive definite and radially unbounded function $V(\boldsymbol{\xi}) = \frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}$. Its time derivative is:

$$\dot{V} = \boldsymbol{\xi}^T\dot{\boldsymbol{\xi}} = \boldsymbol{\xi}^T(A - L'C)\boldsymbol{\xi} + \boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) - \boldsymbol{\xi}^T L''(\hat{\boldsymbol{x}})C\boldsymbol{\xi}. \quad (40)$$

The result is proved if it can be shown that there exists a choice of L' and L'' for which $\dot{V}$ is negative definite.

First, it can be verified that, after removing (16), the couple (A, C) is detectable, hence the term $\boldsymbol{\xi}^T(A - L'C)\boldsymbol{\xi}$ can be made negative-definite by a proper selection of L' (i.e., any which makes $(A - L'C)$ Hurwitz). This part is addressed later in the proof.

Focusing on the term $N(\boldsymbol{\xi}, \boldsymbol{x}, \hat{\boldsymbol{x}}) \triangleq \boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) - \boldsymbol{\xi}^T L''C(\hat{\boldsymbol{x}})\boldsymbol{\xi}$, the strategy of the proof will be first to select L'' so that the resulting terms in $-\boldsymbol{\xi}^T L''C(\hat{\boldsymbol{x}})\boldsymbol{\xi}$ cancel as many terms as possible of $\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}}))$, and then to bound the contribution to $\dot{V}$ of any residual term in $N(\boldsymbol{\xi}, \boldsymbol{x}, \hat{\boldsymbol{x}})$ with a proper choice of L'. First of all, the non-zero components of $\boldsymbol{f}$ are (see Section II-C):

- $f_1(x) = \omega_{ci}x_9 x_{11} + \omega_{ci}x_{10}x_{12}$
- $f_2(x) = -\omega_{ci}x_9 x_{12} + \omega_{ci}x_{10}x_{11}$
- $f_7(x) = -m_{P_i}x_1 x_8 + x_8 x_{14}$
- $f_8(x) = m_{P_i}x_1 x_7 - x_7 x_{14}$
- $f_9(x) = -m_{P_i}x_1 x_{10} + x_{10}x_{14}$
- $f_{10}(x) = m_{P_i}x_1 x_9 - x_9 x_{14}$
- $f_{11}(x) = -m_{P_i}x_1 x_{12} + x_{12}x_{14}$
- $f_{12}(x) = m_{P_i}x_1 x_{11} - x_{11}x_{14}$

By noticing that $x_i x_j - \hat{x}_i \hat{x}_j = \xi_i \xi_j + \xi_i \hat{x}_j + \hat{x}_i \xi_j = \xi_i x_j + \hat{x}_i \xi_j$, the term $\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}}))$ in (40) can be written, after simplifications, as:

$$\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) =$$
$$\omega_{ci}\xi_1\xi_9 x_{11} + \omega_{ci}\xi_1\hat{x}_9\xi_{11} + \omega_{ci}\xi_1\xi_{10}x_{12} + \omega_{ci}\xi_1\hat{x}_{10}\xi_{12} +$$
$$-\omega_{ci}\xi_2\xi_9 x_{12} - \omega_{ci}\xi_2\hat{x}_9\xi_{12} + \omega_{ci}\xi_2\xi_{10}x_{11} + \omega_{ci}\xi_2\hat{x}_{10}\xi_{11}$$
$$-m_{P_i}\xi_7\xi_1 x_8 + \xi_7\hat{x}_8\xi_{14} +$$
$$+m_{P_i}\xi_8\xi_1 x_7 - \xi_8\hat{x}_7\xi_{14} +$$
$$-m_{P_i}\xi_9\xi_1 x_{10} + \xi_9\hat{x}_{10}\xi_{14} +$$
$$+m_{P_i}\xi_{10}\xi_1 x_9 - \xi_{10}\hat{x}_9\xi_{14} +$$
$$-m_{P_i}\xi_{11}\xi_1 x_{12} + \xi_{11}\hat{x}_{12}\xi_{14} +$$
$$+m_{P_i}\xi_{12}\xi_1 x_{11} - \xi_{12}\hat{x}_{11}\xi_{14}. \quad (41)$$

The choice of L'' to cancel as many terms in (41) as possible can now be done in two steps. First, a proper selection of elements in L'' allows to cancel all the entries in (41) that explicitly depend on $\hat{\boldsymbol{x}}$ (i.e., $\omega_{ci}\xi_1\hat{x}_9\xi_{11}$, $\omega_{ci}\xi_1\hat{x}_{10}\xi_{12}$, etc.). In

fact, first of all, given the structure (19) of C it is easy to see that the generic i_{th} row of $L''C$ is given by:

$$[L''_{i1}, L''_{i2}, 0, 0, 0, 0, L''_{i3}, L''_{i4}, L''_{i5}, L''_{i6}, L''_{i7}, L''_{i8}, 0, L''_{i9}, L''_{i10}] \quad (42)$$

Hence, the generic term $c\xi_i\hat{x}_j\xi_k$ in (41) (with c a constant and $i, j, k \in \{1, ..., 15\}$), can be cancelled by choosing the (i, k) entry of $L''C$ as $[L''C]_{i,k} = -c\hat{x}_j$. Hence, one can take: $L''_{1,7} = -\omega_{ci}\hat{x}_9$, $L''_{1,8} = -\omega_{ci}\hat{x}_{10}$, $L''_{2,8} = \omega_{ci}\hat{x}_9$, $L''_{2,7} = -\omega_{ci}\hat{x}_{10}$, $L''_{7,9} = -\hat{x}_8$, $L''_{8,9} = \hat{x}_7$, $L''_{9,9} = -\hat{x}_{10}$, $L''_{10,9} = \hat{x}_9$, $L''_{11,9} = -\hat{x}_{12}$, $L''_{12,9} = \hat{x}_{11}$.

Then, by a proper selection of the remaining L'' entries, it is also possible to cancel the six terms in (41) which depend on parameter m_{P_i}. This is done by noticing that they can be divided into three couples with a similar structure: $-m_{P_i}\xi_7\xi_1 x_8 + m_{P_i}\xi_8\xi_1 x_7$, $-m_{P_i}\xi_9\xi_1 x_{10} + m_{P_i}\xi_{10}\xi_1 x_9$, and $-m_{P_i}\xi_{11}\xi_1 x_{12} + m_{P_i}\xi_{12}\xi_1 x_{11}$. The first couple can be eliminated by selecting $L''_{1,3} = m_{P_i}\hat{x}_8$ and $L''_{1,4} = -m_{P_i}\hat{x}_7$, the second couple by selecting $L''_{1,5} = m_{P_i}\hat{x}_{10}$ and $L''_{1,16} = -m_{P_i}\hat{x}_9$, and the third one by selecting $L''_{11,1} = m_{P_i}\hat{x}_{12}$ and $L''_{12,1} = -m_{P_i}\hat{x}_{11}$.

With the above choices, the term $\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) + \boldsymbol{\xi}^T L''(\hat{\boldsymbol{x}})C\boldsymbol{\xi}$ in (40) reduces to:

$$\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) + \boldsymbol{\xi}^T L''(\hat{\boldsymbol{x}})C\boldsymbol{\xi} =$$
$$\omega_{ci}\xi_1\xi_9 x_{11} + \omega_{ci}\xi_1\xi_{10}x_{12} - \omega_{ci}\xi_2\xi_9 x_{12} + \omega_{ci}\xi_2\xi_{10}x_{11} \quad (43)$$

which consists of the sum of terms of the kind $cx_i\xi_j\xi_k$, with c being a coefficient. The generic state component can be upper bounded as $x_i \leq \bar{x}_i$, since the state space is contained in a compact set. From the Young's inequality, it is $c\bar{x}_i\xi_j\xi_k \leq |c||\bar{x}_i|(\xi_j^2 + \xi_k^2)$. Hence, with simple calculations, (43) can be upper-bounded as:

$$\boldsymbol{\xi}^T(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\hat{\boldsymbol{x}})) + \boldsymbol{\xi}^T L''(\hat{\boldsymbol{x}})C\boldsymbol{\xi} \leq$$
$$\leq \omega_{ci}(\bar{x}_{11} + \bar{x}_{12})(\xi_1^2 + \xi_2^2 + \xi_9^2 + \xi_{10}^2) = \boldsymbol{\xi}^T D\boldsymbol{\xi}, \quad (44)$$

with D a diagonal, positive semi-definite matrix, such that, $D_{i,i} = \omega_{ci}(\bar{x}_{11} + \bar{x}_{12})$ for $i = 1, 2, 9, 10$, and zero otherwise. In conclusion, by plugging (44) in (40), $\dot{V}$ can be bounded as:

$$\dot{V} \leq \boldsymbol{\xi}^T(A + D - L'Cs)\boldsymbol{\xi}. \quad (45)$$

All that is left is to select L' in (45) so that matrix $(A + D - L'C)$ is negative definite. Since $(A+D, C)$ is detectable (after removing (16),), this can be done by selecting (via a standard eigenvalue assignment problem) any L' such that $(A + D - L'C)$ is Hurwitz. The eigenvalues of $(A + D - L'C)$ can be chosen to control the speed of convergence to zero of the estimation error.

In conclusion, it has been proven that the observer (37) with $L(\hat{x}) \triangleq L' + L''(\hat{x})$, and any L', $L''(\hat{x})$ chosen as:

- $L''_{1,7} = -\omega_{ci}\hat{x}_9$, $L''_{1,8} = -\omega_{ci}\hat{x}_{10}$, $L''_{2,8} = \omega_{ci}\hat{x}_9$, $L''_{2,7} = -\omega_{ci}\hat{x}_{10}$, $L''_{7,9} = -\hat{x}_8$, $L''_{8,9} = \hat{x}_7$, $L''_{9,9} = -\hat{x}_{10}$, $L''_{10,9} = \hat{x}_9$, $L''_{11,9} = -\hat{x}_{12}$, $L''_{12,9} = \hat{x}_{11}$, and all the other entries of $L''(\hat{x})$ equal to zero,
- L' such that $(A + D - L'C)$ is Hurwitz, with D a diagonal matrix with entries $D_{i,i} = \omega_{ci}(\bar{x}_{11} + \bar{x}_{12})$ for $i = 1, 2, 9, 10$, and zero otherwise,

globally asymptotically reconstructs the state x. ∎

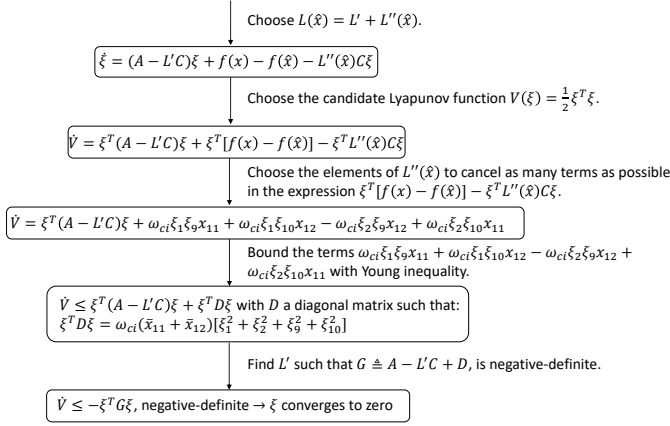The key steps in the proof are summarized in Fig. 7.

Fig. 7: Main steps of the proof of Theorem 2.

The detection and mitigation strategy towards resilient MG operation is based on the analysis of the residual vector $\boldsymbol{r} \triangleq \boldsymbol{y} - \hat{\boldsymbol{y}}$, which represents the difference between the measurements, $\boldsymbol{y}$, affected by the attacks, and the estimated measurements, built on the basis of the estimated state, $\hat{\boldsymbol{x}}$, which tracks $\boldsymbol{x}$ when there are no attacks. A threshold-based detection scheme is adopted, in which an alarm is raised when $\|\boldsymbol{r}\|$ exceeds a pre-specified threshold η. When this occurs, the impacted state variables are replaced with their estimates from the observer. The same detection and mitigation strategy can be used on any other signal that is a function of the state variables. Finally, the following remark presents an alternative observer based on output injection [27].

Remark 3 (Output Injection Observer [27]). *A first simpler and alternative observer design based on "output injection" can be made by noticing that all the state variables appearing in $\boldsymbol{f}$ (see Section II-C) are actually measurable for our particular system (see Remark 1), and thus available to the observer from the output of the system. Hence, by following the simple design suggested, e.g., in [27, Section 3.1.2], we can take in (37) $\boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{x})$, and $L(\hat{\boldsymbol{x}}, \boldsymbol{y}) = L$ any constant matrix which makes $(A - LC)$ Hurwitz. As proved in [27], this simple design choice results in an observer which correctly estimates the DG state in normal operation (as it can be seen also from a simple adaptation of the proof of Theorem 2).*

A. Attack Detectability Analysis

In this part, we present the attack detectability analysis for the two observers introduced previously and then we analyze the conditions for the aforementioned detection scheme. We start by analyzing the impact that the attack vector has on the dynamics of the states and on the residual variable. Let $x_n(t)$ denote the state of the system at time t in normal conditions (i.e., when there is no attack according to (1)), $x(t)$ denote the state of the system under attack (according to (20)), and $B\Gamma_u = B_a$. We want to characterize the deviation in the state caused by the attack, i.e., the dynamics of the incremental state variables $x_a(t) \triangleq x(t) - x_n(t)$:

$$\dot{x}_a = Ax_a + f(x) - f(x_n) + B_a a_u \quad (46)$$

Moving now to the analysis of the impact on the residual, let $r_n(t)$ denote the residual at time t, in normal conditions,

and denote with $r(t)$ the residual when the system is under attack. Define the incremental residual, i.e., the impact of the attack on the residual, as $r_a(t) \triangleq r(t) - r_n(t) = C(x(t) - \hat{x}(t)) - C(x_n(t) - \hat{x}_n(t)) = C(\xi(t) - \xi_n(t)) \triangleq C\xi_a$, where ξ and ξ_n are, respectively, the estimation error in the attack and in the normal scenario (i.e., the deviation between the state of the system and the observer estimate).

Consider first the output injection observer of Remark 3. In this case, simple calculations show that the increment to the detector state caused by the attack is governed by the following:

$$\dot{\hat{x}}_a = (A - LC)\hat{x}_a + f(x) - f(x_n) + LC(x - x_n). \quad (47)$$

Also, the dynamics of $\xi_n(t)$ is:

$$\dot{\xi}_n \triangleq \dot{x}_n - \dot{\hat{x}}_n = (A - LC)\xi_n, \quad (48)$$

and similarly the dynamics of $\xi(t)$ is:

$$\dot{\xi} \triangleq \dot{x} - \dot{\hat{x}} = (A - LC)\xi + B_a a_u \quad (49)$$

From the above calculations, we can compute the dynamics of variable r_a, which characterizes the deviation that the attack causes to the residual, fundamental to evaluate the detectability of the attacks.

$$\dot{r}_a = C(\dot{\xi} - \dot{\xi}_n) \triangleq C\dot{\xi}_a = C(A - LC)\xi_a + CB_a a_u \quad (50)$$

Remark 4 (Undetectability condition for observer in Remark 3). *From (50) it is seen that the attack is perfectly undetectable if $a_u \in \ker(CB_a)$, since then the residual does not depend on the attack. In particular, given the structure of (19), it follows that any attack targeting only the state variables which are not measured is undetectable under the observer of Remark 3. Hence, the use of this observer is not recommended in practice. Furthermore, even attacks such that $CB_a \neq 0$ could be easily made undetectable. As a matter of fact, dynamics (50) are linear, and the solutions to the equation $\dot{\xi}_a = (A - LC)\xi_a + B_a a_u$ are also solutions to (50). They are given by:*

$$\xi_a(t) = e^{(A-LC)t}\xi_a(0) + \int_0^t e^{[A-LC](t-\tau)}B_a a_u(\tau)d\tau. \quad (51)$$

Considering now that $\xi_a(0) = 0$ (at the initial time of the attack the effect on the state is zero), and that $\dot{r}_a = C(\dot{\xi} - \dot{\xi}_n) \triangleq C\dot{\xi}_a$, from (51) we find that if:

$$\left\| C \int_0^t e^{[A-LC](t-\tau)}B_a a_u(\tau)d\tau \right\| \leq \tau \quad (52)$$

then, the attack is not detected.

Finally, if instead the proposed observer is used (Theorem 2), the dynamics of the variation of the residual due to the attack are given by:

$$\begin{aligned} \dot{r}_a := C\dot{\xi}_a = C\{A\xi_a + [f(x) - f(x_n)] + \\ - [f(\hat{x}) - f(\hat{x}_n)] - L(\hat{x})C\xi_a + B_a a_u\}. \end{aligned} \quad (53)$$

Notice that (53) is fundamentally different from (50), since they are nonlinear and, beyond ξ_a, they include also x, x_n, $\hat{x}$, and $\hat{x}_n$. Thus, even when the attack satisfies $CB_a a_u = 0$ [11], the residual is affected by the impact of the attack on x.

1) Conditions for Attack Detection: The residual scheme is designed and analyzed based on the derivation of a suitable observer which globally asymptotically tracks the state x of the system in the absence of attacks. An important related question is determining the class of attacks that can be detected. This part focuses on deriving the conditions for the aforementioned detection scheme. The analysis provides a theoretical result that characterizes quantitatively and implicitly the class of attacks detectable by the proposed scheme.

From (20), and incorporating the attack function dynamics according (30) and (31), the state space model of each inverter can be described as:

$$\tilde{W}_a : \begin{cases} \dot{x}(t) = k(x,u) + f(x(t)) + \zeta(t) \\ + B_a \beta_i(t - To) a_u(t) \\ y(t) = Cx(t) \end{cases} \quad (54)$$

where $k(x,u)$ represents the nominal function dynamics of each DG considering both the state and input vactors, and $\zeta(t) = Ew(t)$ describes the known disturbances of the system according to (1) and (20). From (37), the state of the observer can be re-written as:

$$\dot{\hat{x}} = k(\hat{x},u) + f(\hat{x}) + \Phi(\hat{x},x) \quad (55)$$

where $\Phi(\hat{x},x)$ incorporates the nonlinear gain L($\hat{x}$) term from the Luenberger-like observer.

Following similar approaches in literature in which they use a filtering scheme to address unknown disturbances in nonlinear systems [18], by filtering the output signal $y(t)$ of each DG, we can compute the filtered output $z(t)$:

$$z(t) = H(s)[y(t)] = sH_p(s)[x(t)] \quad (56)$$

where $H(s)$ and H_p are asymptotically stable and, hence, BIBO stable. The conditions for selecting such filters can be found in [18]. Now, by using $s[x(t)] = \dot{x}(t) + x(0)$, we obtain:

$$\begin{aligned} z(t) &= H_p(s)\left[\dot{x}(t)\right] + H_p(s)\left[x(0)\delta(t)\right] \\ &= H_p(s)\left[k(x,u) + f(x) + B\Gamma_u a(t)\right] \\ &+ H_p(s)[\zeta(t)] + h_p(t)x(0) \end{aligned} \quad (57)$$

The initial conditions $x(0) = \hat{x}(0)$ are considered to be known and the term $h_p(t)$ describes a exponential decay [18], converging to zero, eventually. Filtering the output of the observer, $\hat{z}$ can be described as:

$$\begin{aligned} \hat{z}(t) &= H(s)[\hat{x}(t)] \\ &= H_p(s)[k(\hat{x},u) + f(\hat{x}) + \Phi\hat{x}, x] + h_p(t)\hat{x}(0) \end{aligned} \quad (58)$$

The residual error $r(t)$ to detect the attack in each DG can be calculated from (56) and (58), and defined as:

$$r(t) = z(t) - \hat{z}(t) \quad (59)$$

The logic for attack detection is defined as:

$$\begin{cases} |r(t)| \leq \eta \rightarrow No\ attack \\ |r(t)| > \eta \rightarrow Attack\ detected \end{cases} \quad (60)$$

Prior to the attack ($t < T_o$), the residual signal can be written using (56) and (57) as:

$$r(t) = H_p(s)[\chi(t)] + H_p(s)[\zeta(t)] = H_p(s)[\chi(t)] + \zeta(t) \quad (61)$$

where the term $\chi(t)$ incorporates the characteristics of the system and the observer, can be defined as:

$$\chi(t) = [k(x,u) - k(\hat{x},u)] + [f(x) - f(\hat{x})] - \Phi(\hat{x},x) \quad (62)$$

Now, by taking bounds on the residual error, we obtain:

$$\begin{aligned} r(t) &= |H_p(s)[\chi(t)] + H_p(s)[\zeta(t)]| \\ &\leq |H_p(s)[\chi(t)]| + |H_p(s)[\zeta(t)]| \\ &= \left|\int_0^t h_p(t-\tau)\chi(\tau)\,d\tau\right| + \left|\int_0^t h_p(t-\tau)\zeta(\tau)\,d\tau\right| \\ &\leq \int_0^t |h_p(t-\tau)||\chi(\tau)|\,d\tau + \int_0^t |h_p(t-\tau)||\zeta(\tau)|\,d\tau \\ &\leq \int_0^t |h_p(t-\tau)|\chi(\tau)d\tau + \int_0^t |h_p(t-\tau)|\zeta(\tau)d\tau \end{aligned} \quad (63)$$

A suitable threshold $\eta(t)$ is given as:

$$\eta(t) = \int_0^t \bar{h}_p(t-\tau)\bar{\chi}(\tau)d\tau + \int_0^t \bar{h}_p(t-\tau)\bar{\zeta}(\tau)d\tau \quad (64)$$

Finally, using the approach in [18], the threshold can be implemented as:

$$\eta(t) = \bar{H}_p(s)\bar{\chi}(t) + \bar{H}_p(s)\bar{\zeta}(t) = \bar{H}_p(s)[\bar{\chi}(t) + \bar{\zeta}(t)] \quad (65)$$

Theorem 3 (Attack Detectability). *Consider the nonlinear system described in* (54)*, with the residual-based observer detection scheme in* (55) *with the conditions* (56)–(59) *and* (65)*. An attack introduced at $t = T_o$ is detectable if the attack function $a_u(t)$ satisfies the following inequality:*

$$\left|\int_{T_o}^t h_p(t-\tau)(1 - e^{-b_i(t-T_o)})a_u(t)d\tau\right| > 2\eta(t) \quad (66)$$

Proof. In the presence of an attack at $t = T_o$, (61) becomes:

$$r(t) = H_p(s)[\chi(t) + \beta_i(t - To)a_u(t)] + H_p(s)[\zeta(t)] \quad (67)$$

By using the triangle inequality, for $t > T_o$, $r(t)$ is computed:

$$\begin{aligned} r(t) &\geq -|H_p(s)[\chi(t)]| - |H_p(s)[\zeta(t)]| \\ &+ |H_p(s)[\beta_i(t - To)a_u(t)]| \\ &\geq -\int_0^t |h_p(t-\tau)||\chi(\tau)|\,d\tau - \int_0^t |h_p(t-\tau)||\zeta(\tau)| \\ &+ |H_p(s)[\beta_i(t - To)a_u(t)]| \\ &\geq -\int_0^t \bar{h}_p(t-\tau)\bar{\chi}(\tau)d\tau - \int_0^t \bar{h}_p(t-\tau)\bar{\zeta}(\tau)d\tau \\ &+ |H_p(s)[\beta_i(t - To)a_u(t)]| \\ &\geq -\eta(t) + |H_p(s)[\beta_i(t - To)a_u(t)]| \end{aligned} \quad (68)$$

The final attack detectability condition, can be computed as:

$$|H_p(s)[\beta_i(t - To)a_u(t)]| > 2\eta(t) \quad (69)$$

∎

V. BI-LEVEL STABILITY-CONSTRAINED ATTACK-MITIGATION FORMULATION

In this section, we first present the OPF problem for AC MGs and then identify the worst-case cyber-attack in a formulation of a bi-level optimization problem. The attacker follows the threat model of Section III while satisfying the OPF of the system operation. Operators, in order to ensure a stable operation of the MG, react by incorporating both stability constraints as well as the design of the detection observer. These conditions ensure not only the MG stable and optimal operation but also guarantee the detection of attacks. Fig. 8 presents the concept of the formulation.
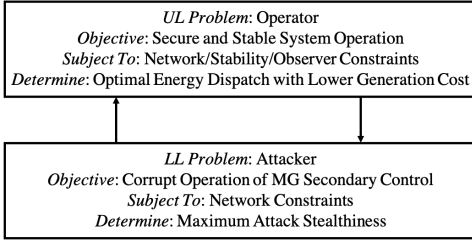
Fig. 8: Bi-level formulation that considers attacker and operator objectives.

A. Optimal Power Flow in AC Microgrids

The injected active and reactive power at every inverter-based DG are $p^g \in \mathbb{R}^{n^g \times 1}$ and $q^g \in \mathbb{R}^{n^g \times 1}$. The total load demand is characterized by $d \in \mathbb{C}^{n^d \times 1}$ in the form of constant complex impedances $d = [R_1 + X_1, \cdots, R_{n^d} + X_{n^d}]$. The MG admittance matrix is $Y \in \mathbb{C}^{n \times n}$ and the load and inverter incidence matrices are given by $\hat{D} \in \{0, 1\}^{n^d \times n}$ and $G \in \{0, 1\}^{n^g \times n}$, respectively. Thus, the from and to admittance matrices are represented as $\overrightarrow{Y}, \overleftarrow{Y} \in \mathbb{C}^{n \times n}$, and their respective branch-incidence matrices as $\overrightarrow{L}, \overleftarrow{L} \in \{0, 1\}^{l \times n}$. Finally, v_o is the terminal voltages at inverter/DG output terminals, and $v = [v^g, v^b] \in \mathbb{C}^{n \times 1}$ represents all the voltages, where $v^g \in \mathbb{C}^{n^g \times 1}$ is the vector of bus voltages at point of coupling, and $v^b \in \mathbb{C}^{n - n^g \times 1}$ is the vector of all remaining MG buses. The OPF problem is then formulated as:

$$\min \ h(p^g) = \min_{P_i} \sum_{i=1}^{n^g} C_i(P_i) \tag{70}$$

$$s.t.: \ G^\top (p^g + jq^g) = \hat{D}^\top d + \text{diag}\{vv^* Y^*\} \tag{71}$$

$$p^{\min} \le p^g \le p^{\max} \tag{72}$$

$$q^{\min} \le q^g \le q^{\max} \tag{73}$$

$$\text{diag}\left\{\overrightarrow{\overleftarrow{L}} vv^* \overrightarrow{\overleftarrow{Y}}^*\right\} \le l^{\max} \tag{74}$$

$$(v^{\min})^2 \le |v|^2 \le (v^{\max})^2 \tag{75}$$

Here, in order to minimize the total cost of DG output while satisfying system-wide requirements/constraints, we need to schedule each generator's active power output. The generation cost for the i_{th} generator is typically approximated by quadratic cost functions[1] $C_i(P_i) = \alpha_i P_i^2 + \beta_i P_i + \gamma_i$, where $\alpha_i, \beta_i,$ and γ_i are the cost parameters. The system constraints include the enforcement of the nodal power balance in Eq. (71), the bounding conditions of active and reactive power of the individual DG outputs in Eqs. (72) and (73), respectively, the line flows in either direction using Eq. (74), and the voltage magnitude limits within $[v^{\min}, v^{\max}]$ according to Eq. (75).

---

[1]Quadratic cost functions are widely used in OPF problems even for inverter-based systems [65]. The problem formulation, however, can be solved for any convex formulation of the cost function.

B. Lower Level Problem: Identification of Worst-Case Attack

The aim of the attack is to maximize the effect of the disruption at the output signals of the MG secondary controller while remaining stealthy as explained in Section III. The formalization of the worst-case attack model is achieved by co-optimizing it with the MG operations as given in (70)–(75) and adopting a common cybersecurity practice, which is to conservatively assume a strong (omniscient) and stealthy attacker with full knowledge of the network requirements including all nodal and line parameters, operating limits, and priorities. These assumptions lead to assessing the worst-case impact of the attack. The attacker with a generic objective, denoted as $O^A(act_{it}')$, can be modeled as in (76), where (77) presents the MG OPF model conditioned by the modified actions act_{it}' that satisfy the stealthiness properties of (78).

$$\max O^A(\cap_{\forall i \in \mathcal{G}} act_{it}') \tag{76}$$

$$s.t.: \ \{(70) - (75) \text{ with } act_{it} = act_{it}', \forall i \in \mathcal{G}, t \in \mathcal{T}\} \tag{77}$$

$$act_{it}' \ s.t.: \ \{(23) - (26) \text{ and } (33) - (35)\} \tag{78}$$

C. Upper Level Problem: Stability-Constrained Operation

In this part, the MG entity operates at the top of the hierarchy and aims to enforce the optimal operating setpoints to the inverters while satisfying OPF conditions and ensuring the small-signal stability of the inverter-based MG, since OPF cannot guarantee MG stability [59], let alone the existence of a stealthy attack, and assuming the existence of a residual-based observer able to detect and mitigate malicious signal manipulations according (76)–(78). To establish the stability restriction, the MG is initially represented as a collection of differential-algebraic equations:

$$\dot{x} = \tilde{f}(x, z) \tag{79}$$

$$0 = \tilde{g}(x, z) \tag{80}$$

In the above equations, $\tilde{f}$ and $\tilde{g}$ denote the nonlinear differential and algebraic equation vectors of a MG, respectively. x and z are the vectors of state and algebraic variables with sizes n_x and n_z, respectively. A third-order inverter model, as stated in [66], is employed to minimize the computational burden. Fig. 2 depicts voltage and current controllers that are equipped with an LC filter, which possess a significantly higher closed-loop bandwidth compared to the power controller module. It can be assumed that these control loops achieve quasi-steady-state quickly. Therefore, the vector of differential equations, denoted by $\tilde{f}$, includes state variables represented by $x = [p^\top, q^\top, \delta^\top]^\top \in \mathbb{R}^{n_x \times 1}$.

$$\dot{x}^\top = \begin{bmatrix} -\omega_c \cdot p + \omega_c \cdot \text{Re}\{v_o \cdot (i_o)^*\} \\ -\omega_c \cdot q + \omega_c \cdot \text{Im}\{v_o \cdot (i_o)^*\} \\ (\omega - \omega_{com})\omega_b \end{bmatrix} \tag{81}$$

where ω_c, p and $q \in \mathbb{R}^{n^g \times 1}$, $v_o \in \mathbb{C}^{n^g \times 1} = v_{od} + jv_{oq}$, and $i^o \in \mathbb{C}^{n^g \times 1} = i_{od} + ji_{oq}$. The operating inverter frequency ω is obtained using: $\omega = \omega_b - m_P \cdot p + m_P \cdot p_{opf}$, where p_{opf} is the active power setpoint provided by the OPF.

Table I: Parameters of the test MG.

DGs 1 & 2		DGs 3 & 4	
Parameter	Value	Parameter	Value
m_P	9.4×10^{-5}	m_P	12.5×10^{-5}
n_Q	1.3×10^{-3}	n_Q	1.5×10^{-3}
R_c	$0.03\ \Omega$	R_c	$0.03\ \Omega$
L_c	0.35 mH	L_c	0.35 mH
R_f	$0.1\ \Omega$	R_f	$0.1\ \Omega$
L_f	1.35 mH	L_f	1.35 mH
C_f	$50\ \mu$F	C_f	$50\ \mu$F
K_{PV}	0.1	K_{PV}	0.05
K_{IV}	420	K_{IV}	390
K_{PC}	15	K_{PC}	10.5
K_{IC}	20000	K_{IC}	16000
ω_b	314.16 rad/s	ω_b	314.16 rad/s
F	0.75	F	0.75
ω_c	31.41 Hz	ω_c	31.41 Hz
c_f	30	c_f	30
c_v	30	c_v	30

Table II: Parameters of MG lines and loads.

Lines 1, 3		Line 2	
R_{line}	$0.23\ \Omega$	R_{line}	$0.35\ \Omega$
L_{line}	$318\ \mu$H	L_{line}	$1847\ \mu$H

Load 1		Load 2		Load 3		Load 4	
R	$30\ \Omega$	R	$20\ \Omega$	R	$25\ \Omega$	R	$25\ \Omega$
X	$15\ \Omega$	X	$10\ \Omega$	X	$10\ \Omega$	X	$15\ \Omega$

The vector of algebraic equations, denoted by $\tilde{g}$, includes the algebraic variables represented by $z = [(i_{\text{od}})^\top, (i_{\text{oq}})^\top]^\top \mathbb{R}^{n_z \times 1}$.

$$z^\top = \left[\begin{array}{c} \text{Re}\left\{ \check{Y}\left(v_o - i_o \cdot z_c\right)\right\} \\ \text{Im}\left\{ \check{Y}\left(v_o - i_o \cdot z_c\right)\right\} \end{array} \right] \quad (82)$$

where $\check{Y}$ is the Kron-reduced admittance matrix of the system [66] and z_c is the line impedance connecting an inverter to the MG. The inverter terminal's voltage is given as:

$$v_o = (v_{\text{opf}} + n_Q \cdot q_{\text{opf}} - n_Q \cdot q) \cdot (\cos\delta + \text{j}\sin\delta) \quad (83)$$

where n_Q is $q - v$ droop constant, and q_{opf} and v_{opf} are the OPF resulted values for the reactive power and voltage setpoints, respectively.

Proposition 2. *Let us consider a MG described by (79)–(80). The state matrix, denoted by $\hat{A} \in \mathbb{R}^{n_x \times n_x}$, is defined as $\hat{A} = A(i_{\text{od}}, i_{\text{oq}}, \delta, q) - B(\delta, q)(D)^{-1}C(\delta, q)$, where $A(i_{\text{od}}, i_{\text{oq}}, \delta, q) = \frac{\partial \tilde{f}}{\partial x}$, $B(\delta, q) = \frac{\partial \tilde{f}}{\partial z}$, $C(\delta, q) = \frac{\partial \tilde{g}}{\partial x}$, and $D = \frac{\partial \tilde{g}}{\partial z}$. The expressions for these matrices are given in the Appendix by (89) – (92), along with the definition of the matrix M^p in (88). Furthermore, $\check{G}$ and $\check{B}$ represent the real and imaginary parts, respectively, of the Kron-reduced admittance matrix $\check{Y}$. According to [67], the small-signal stability of the MG can be ensured – incorporating the Lyapunov stability, with a minimum decay rate (i.e., damping ratio) of η, if there exists a symmetric positive definite matrix M that satisfies:*

$$\hat{A}^\top M + M\hat{A} \preceq -2\eta M. \quad (84)$$

The proof for Proposition 1 can be found in [59].
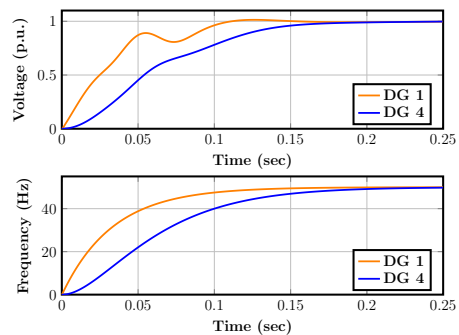

Fig. 9: Attack-free scenario.

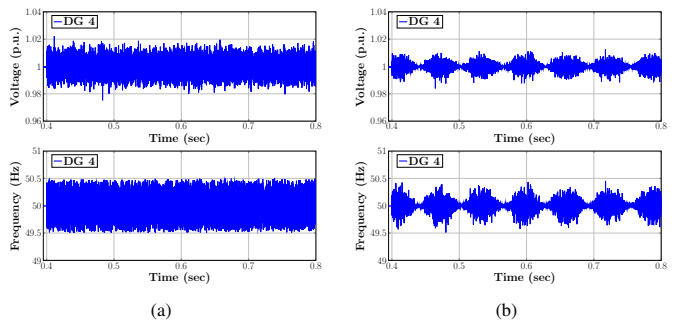(a)   (b)

Fig. 10: Effects (at the output of DG 4) of an arbitrary attack vector following for both voltage and frequency: (a) a uniform distribution with $\mu[-0.01, 0.01]$, and (b) a normal distribution with $\mathcal{N}(0, 0.00001)$ multiplied with a sine carrier of reference values (1 p.u., 50 Hz).

Additional constraints are formulated to connect the inverter's internal variables with the OPF variables:

$$v^{\text{g}} = (v_{\text{ref}} - n_Q \cdot q) \cdot (\cos\delta + \text{j}\sin\delta) - i_o \cdot z_c \quad (85)$$

$$i_{\text{o}} = \check{Y}v^{\text{g}} \quad (86)$$

$$p + \text{i}q = \text{diag}\left\{i_o\left(i_o\right)^* [z_c]\right\} + p^g + \text{i}q^g \quad (87)$$

The OPF represented by (70)-(75), along with the additional constraints in (84)-(87), constitutes a small-signal stability-constrained OPF for an inverter-dominant MG. In combination with the utilization of the designed observer (37), with $L(\hat{x}) \triangleq L' + L''(\hat{x})$, and any L', $L''(\hat{x})$ chosen as described in the proof of Theorem 2 and able to globally asymptotically reconstruct the state x, ensures that the operator can counter stealthy attacks while ensuring a sufficient stability margin during optimal generation. It should be noted that the constraints in (71)-(75) are quadratic functions of the bus voltage v, while the state matrix $\hat{A}$ in (84) and the constraints (85) and (87) are nonlinear functions of i_o, q, and δ. These bilinear matrix inequalities (BMIs) and nonlinearities make the problem nonconvex which can be solved using computationally tractable semidefinite programming (SDP) or parabolic relaxation techniques [68], or even sequentially through objective penalization [59].

VI. EXPERIMENTAL AND SIMULATION RESULTS

In this section, the simulations results are presented demonstrating the impact of attacks discussed in Section III and the effectiveness of the proposed detection and mitigation strategy of Section IV. The MG adopted for testing is shown in Fig. 1, and the relevant MG modeling parameters are detailed in Table I and Table II. Simulations have been performed with MATLAB 2020.b, on a Macbook Pro 2021, with 16 GB of RAM and a 512 GB hard disk drive.
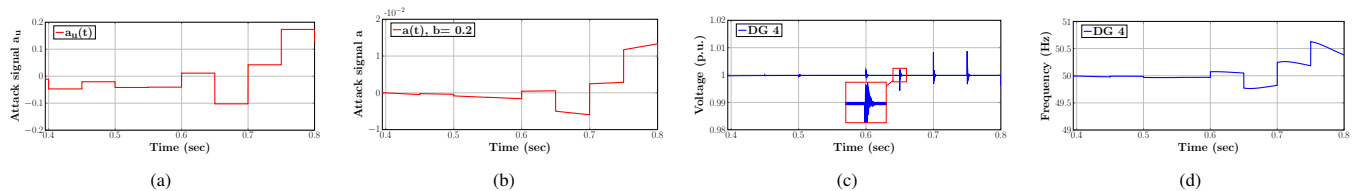
(a)   (b)   (c)   (d)

Fig. 11: Demonstration and effects of a stealthy attack at the outputs of DG 4 with $b_i = 0.2$ as the attack evolution rate.
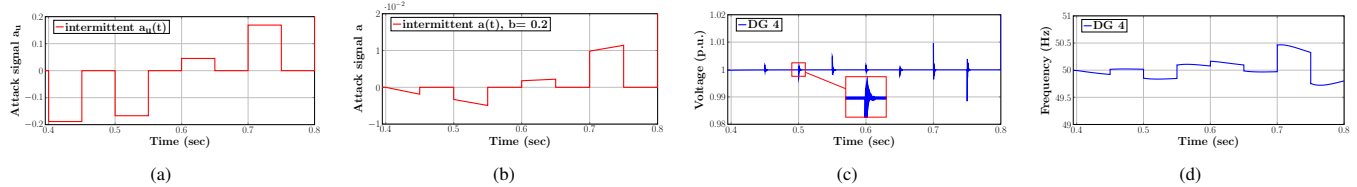

(a)   (b)   (c)   (d)

Fig. 12: Demonstration and effects of a stealthy and intermittent attack at the outputs of DG 4 with $b_i = 0.2$ as the attack evolution rate.
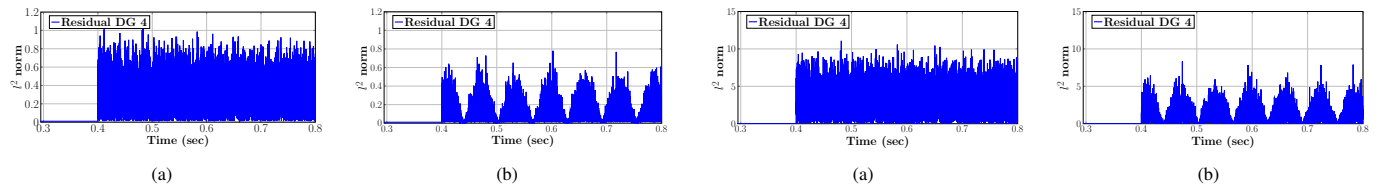

(a)   (b)

Fig. 13: l^2 norm of the proposed residual under an arbitrary attack vector following for both voltage and frequency the attack configuration demonstrated in Fig. 10.
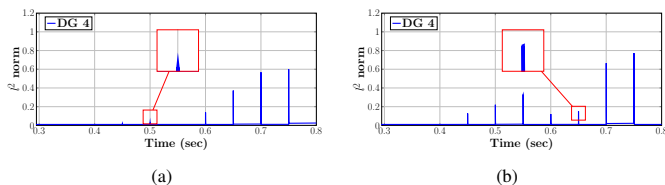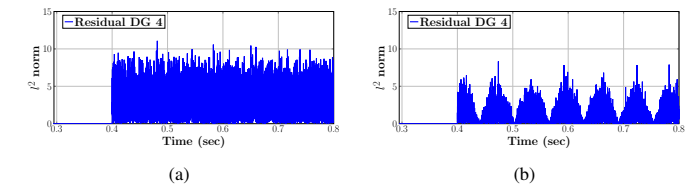

(a)   (b)

Fig. 15: l^2 norm of the residual acquired via the output injection observer (Remark 3) under an arbitrary attack vector following for both voltage and frequency the attack configuration demonstrated in Fig. 10.
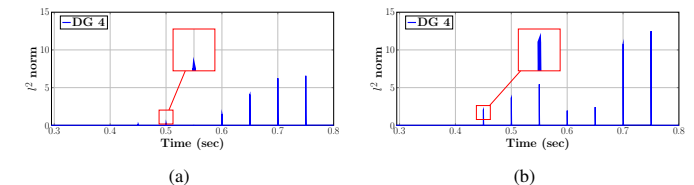

(a)   (b)

Fig. 14: l^2 norm of the proposed residual under: (a) a stealthy only attack, and ((b)) a stealthy and intermittent attack, with $b_i = 0.2$.

(a)   (b)

Fig. 16: l^2 norm of the residual acquired via the output injection observer (Remark 3) under: (a) a stealthy only attack, and ((b)) a stealthy and intermittent attack, with $b_i = 0.2$ as the attack evolution rate.

A. Effectiveness of Attack Models

Fig. 9 displays the voltage and frequency values in the attack-free scenario. The reference values are set to 1 per unit (p.u.) and 50 Hz, respectively. The figure shows that the secondary controller operates properly, i.e., voltage and frequency values reach the reference values. The magnitudes of voltage and frequency in the DGs present instabilities after attacks are introduced into the system. This effect is produced by tampering the input references ω_j and v_{odj} provided to the secondary controller. The attack duration is assumed to be bounded and last from $0.4s$ to $0.8s$. This action leads to distorting the active and reactive power values of the DGs inside the MG. The plots presented next are only for DG 4 for brevity. Similar results can be obtained for other DGs.

Figs. 10, 11, 12 present the simulation results under three different attack models: *(a)* an arbitrary attack vector following a uniform and normal-sinusoidal distribution (Fig. 10), *(b)* an attack following the stealthy but not the intermittent character of the attack modeling of Section III (Fig. 11), and *(c)* an attack following the complete proposed model of Section III (Fig. 12). In Figs. 11 and 12, we selected empirically within the attack function dynamics, $\beta_i(t - T_o)$ in (31), as an attack evolution rate $b_i = 0.2$. Increasing substantially the evolution rate of the attack will cause the attack to lead to much higher residual errors. On the other hand, lower values of b_i will have minimal effect on the output voltage and frequency of

the secondary controller, i.e., will minimize the attack model objectives. Overall, the attack vectors intend to affect the system's voltage and frequency stability while tampering the pre-defined setpoints of the MG. It is clear that the constraints of the scheduling interval as well as the stealthiness of the attack slightly reduce the effect on the secondary control output (especially on the frequency range).

B. Residual-based Observer Comparison

The importance of observers relies on computing an error, described as the residual variable r, presented in Theorem 2. This error indicates that the MG is experimenting an anomalous behavior. The norm of the error vector, e.g., the Euclidean l^2 norm (i.e., $\|r\|$), can be computed to measure the error. The error acts as an alarm for this work's residual-based detection approach. This paper focuses on effectively detecting attacks, and leverages a residual-threshold based approach. Similar to presented literature, where corrupted system states could be recovered with a selected detection threshold, in our experiments we are able to correctly identify corrupted values with a similar empirically-selected maximum error threshold. The l^2 response of the residual is presented, without any mitigation strategy, for three set of results: *(a)* the proposed residual-based observer (Theorem 2) under an arbitrary attack vector of effects presented in Fig. 10 (Fig. 13), and *(b)* the proposed residual-based observer (Theorem
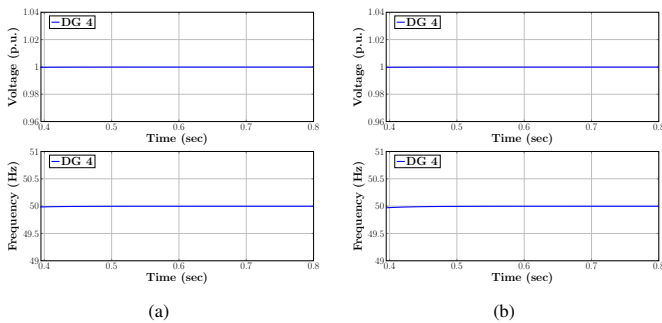
Fig. 17: Detection and mitigation strategy applied at DG 4 under an arbitrary attack vector following for both voltage and frequency the attack configuration demonstrated in Fig. 10.
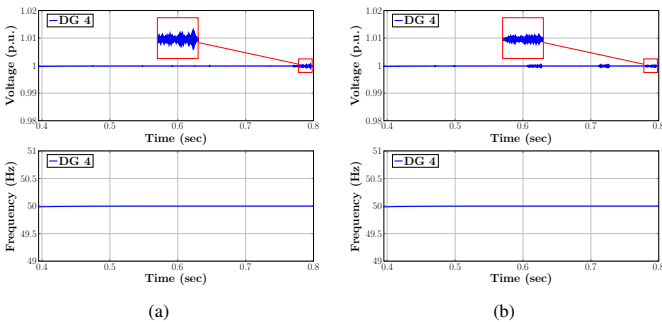

Fig. 18: Detection and mitigation strategy applied at DG 4 under: (a) a stealthy only attack, and (b) a stealthy and intermittent attack, with $b_i = 0.2$ as the attack evolution rate.

2) under the proposed stealthy and intermittent attack model of effects presented in Fig. 12 (Fig. 14), *(c)* an alternative residual-based observer from literature (Remark 3) under an arbitrary attack vector of effects presented in Fig. 10 (Fig. 15), and *(d)* an alternative residual-based observer from literature (Remark 3) under the proposed stealthy and intermittent attack model of effects presented in Fig. 12 (Fig. 16). In all cases, the initial state of the observer is set equal to the measured state at the initial time (for the measurable state variables, while it is set to zero for all the other ones). From Figs. 13–16, it can be observed that the DGs behavior captured by the residuals is sensitive to attack perturbations, especially when the error diverges more critically. In addition, the l^2 norm of the residual error of the proposed observer compared with the output injection observer (Remark 3 - see Section IV-A) indicates that the proposed one is more accurately estimating the system's state.

C. Mitigation Strategy

In this subsection, the mitigation strategy is applied to deal with the attack scenarios. A threshold value is used as an alarm in case the DG inverter is experiencing abnormal operating values for frequency and voltage; in the time instants when one or more components of the residual go beyond the respective threshold τ_i, the DG controllers will use the estimated state values from the observer, instead of the measured values that triggered the residual-based detection condition. Consequently, protection of the MG is effectively guaranteed.

Here, we present three sets of mitigation results following the proposed observer of our work: *(a)* under the effect of the arbitrary attacks (Fig. 17), *(b)* under the proposed stealthy and intermittent attack (Fig. 18), and *(c)* under the proposed stealthy and intermittent attack while complying to

the stability-constrained network modeling and overall formulation structure of Section V (Fig. 19). According to IEEE 1547.4-2011 guide [69], the MG control strategy necessitates that the MG's voltage and frequency deviations meet the standards acceptable to all parties involved. The MG's ability to maintain a narrow frequency range will determine its effectiveness in load following. This is clearly demonstrated by the results in Figs. 17–19. Specifically, after the residual-based observer mitigation is implemented in Figs. 17–18, the effect on both the voltage and frequency for all of the presented scenarios can be considered negligible. The results confirm that the mitigation strategy effectively counters the impact of the attacks, even in the stealthy and intermittent attack case. Furthermore, Fig. 19 illustrates the eigenvalues of the MG obtained under four scenarios: the attack-free case, the stealthy attack scenario, the stealthy and intermittent attack scenario, and after the stability-constrained mitigation formulation is employed while the stealthy and intermittent attack is active. We observe that four of the eigenvalues remain stationary, however, the remaining eleven, in the attack cases, they are located closer to the imaginary axis which can potentially cause instabilities in the MG in response to load perturbations or other disruptive events. In contrast, when stability and network constraints are employed during the stealthy and intermittent attack, the eigenvalues move further away from the imaginary axis, providing a better stability margin.

VII. CONCLUSIONS

This paper presents a detection and mitigation strategy for the operation of islanded MGs under disruptive and stealthy attacks. It depicts the consequences of such attacks at the output of the secondary controller. The work further develops a theoretical formulation of a detection strategy based on the analysis of the residual error between the actual state of the DGs, and the estimated one, as provided by a nonlinear observer. Finally, the paper considers both network and stability constraints to ensure secure MG operation, while it provides simulation-based experiments to demonstrate the effectiveness of the proposed mitigation strategy. The results demonstrate that unexpected disturbances can affect the reference of each DG provided by the secondary controller. However, an accurate mitigation strategy can significantly increase the effort required by the adversary to affect MG operation.

APPENDIX

The matrix $\boldsymbol{M}^{\mathrm{P}}$ is defined as:

$$\boldsymbol{M}^{\mathrm{P}} = \begin{bmatrix} -m_{P_1} & m_{P_2} & 0 & \cdots & 0 \\ -m_{P_1} & 0 & m_{P_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -m_{P_1} & 0 & 0 & \cdots & m_{P_{ng}} \end{bmatrix} \quad (88)$$

where each m_{P_i} indicates the droop constant of the i_{th} inverter, $\forall i \in \mathcal{G}$.

Definition 1 (Convexity). *A set* $\mathrm{P} \subseteq \mathbb{R}^d$ *is convex iff* $\sum_{i=1}^n \lambda_i p_i \in \mathrm{P}$, *for all* $n \in \mathbb{N}$, $p_1, \ldots, p_n \in P$, *and* $\lambda_1, \ldots, \lambda_n \geqslant 0$ *with* $\sum_{i=1}^n \lambda_i = 1$

Definition 2 (Convex Hull). *The convex hull* $\mathrm{convh}(\mathrm{P})$ *of a set* $\mathrm{P} \subseteq \mathbb{R}^d$ *is the intersection of all convex supersets of* P.
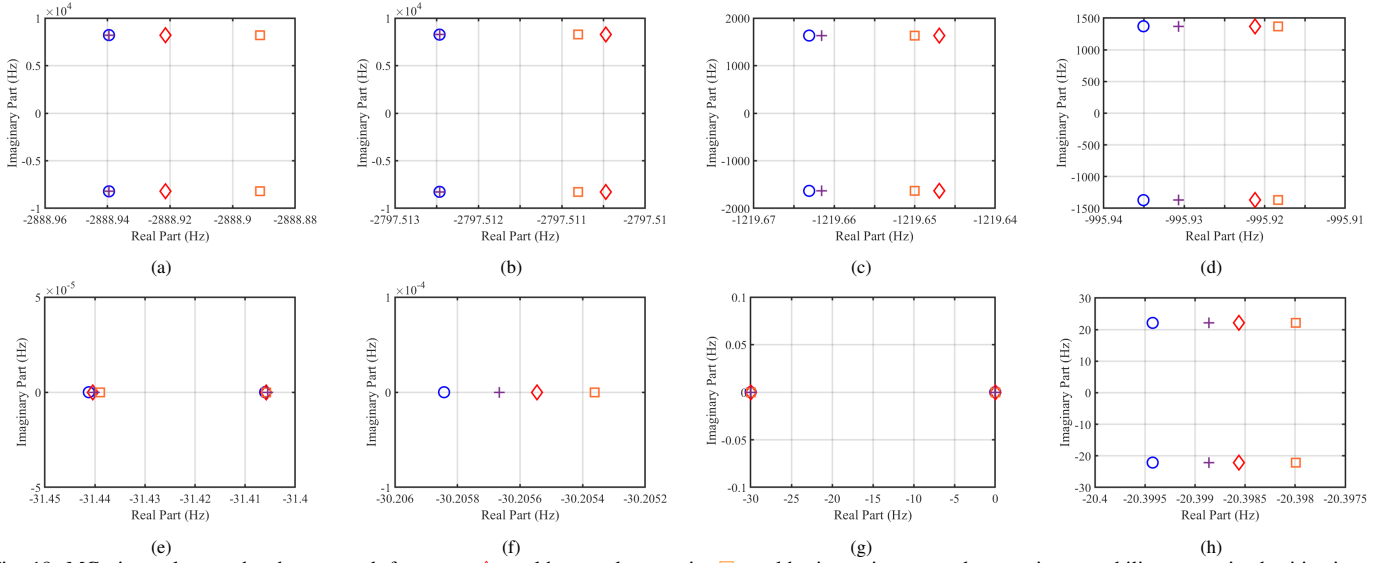
Fig. 19: MG eigenvalues under the: ○ attack-free case, ◇ stealthy attack scenario, □ stealthy-intermittent attack scenario, + stability-constrained mitigation formulation is employed while the stealthy and intermittent attack is active.

$$\frac{\partial \tilde{\mathbf{f}}}{\partial \boldsymbol{x}} = \begin{bmatrix} -[\boldsymbol{\omega}_b] & -[\boldsymbol{\omega}_b \cdot \boldsymbol{n}_Q \cdot (\boldsymbol{i}_{\mathrm{oq}} \cdot \sin\boldsymbol{\delta} + \boldsymbol{i}_{\mathrm{od}} \cdot \cos\boldsymbol{\delta})] & -[\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot (\boldsymbol{i}_{\mathrm{od}} \cdot \sin\boldsymbol{\delta} - \boldsymbol{i}_{\mathrm{oq}} \cdot \cos\boldsymbol{\delta})] \\ 0^{n^{\mathrm{g}}} & [\boldsymbol{\omega}_b \cdot \boldsymbol{n}_Q \cdot (\boldsymbol{i}_{\mathrm{oq}} \cdot \cos\boldsymbol{\delta} - \boldsymbol{i}_{\mathrm{od}} \cdot \sin\boldsymbol{\delta}) - \boldsymbol{\omega}_b] & [\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot (\boldsymbol{i}_{\mathrm{od}} \cdot \cos\boldsymbol{\delta} + \boldsymbol{i}_{\mathrm{oq}} \cdot \sin\boldsymbol{\delta})] \\ \boldsymbol{M}^{\mathrm{P}} & 0^{n^{\mathrm{g}}-1} & 0^{n^{\mathrm{g}}-1} \end{bmatrix} \tag{89}$$

$$\frac{\partial \tilde{\mathbf{f}}}{\partial \boldsymbol{z}} = \begin{bmatrix} [\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \cos\delta] & [\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \sin\delta] \\ [\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \sin\delta] & -[\boldsymbol{\omega}_b \cdot (\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \cos\delta] \\ 0^{n^{\mathrm{g}-1}} & 0^{n^{\mathrm{g}-1}} \end{bmatrix} \tag{90}$$

$$\frac{\partial \tilde{\mathbf{g}}}{\partial \boldsymbol{x}} = \begin{bmatrix} 0^{n^{\mathrm{g}}} & \left(\check{\boldsymbol{G}} \cdot [\boldsymbol{n}_Q \cdot \cos\boldsymbol{\delta}]^{n^{\mathrm{g}}} - \check{\boldsymbol{B}} \cdot [\boldsymbol{n}_Q \cdot \sin\boldsymbol{\delta}]^{n^{\mathrm{g}}}\right) & \check{\boldsymbol{G}} \cdot [(\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \sin\boldsymbol{\delta}]^{n^{\mathrm{g}}} + \check{\boldsymbol{B}} [(\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \cos\boldsymbol{\delta}]^{n^{\mathrm{g}}} \\ 0^{n^{\mathrm{g}}} & \left(\check{\boldsymbol{G}} \cdot [\boldsymbol{n}_Q \cdot \sin\boldsymbol{\delta}]^{n^{\mathrm{g}}} + \check{\boldsymbol{B}} \cdot [\boldsymbol{n}_Q \cdot \cos\boldsymbol{\delta}]^{n^{\mathrm{g}}}\right) & \check{\boldsymbol{B}} \cdot [(\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \sin\boldsymbol{\delta}]^{n^{\mathrm{g}}} - \check{\boldsymbol{G}} [(\boldsymbol{v}_{\mathrm{ref}} - \boldsymbol{n}_Q \cdot \boldsymbol{q}) \cdot \cos\boldsymbol{\delta}]^{n^{\mathrm{g}}} \end{bmatrix} \tag{91}$$

$$\frac{\partial \tilde{\mathbf{g}}}{\partial \boldsymbol{z}} = \begin{bmatrix} I^{n^{\mathrm{g}}} + \check{\boldsymbol{G}} \cdot [\boldsymbol{r_c}]^{n^{\mathrm{g}}} - \check{\boldsymbol{B}} \cdot [\boldsymbol{x_c}]^{\mathrm{g}} & -\check{\boldsymbol{G}} \cdot [\boldsymbol{x_c}]^{n^{\mathrm{g}}} - \check{\boldsymbol{B}} \cdot [\boldsymbol{r_c}]^{n^{\mathrm{g}}} \\ \check{\boldsymbol{B}} \cdot [\boldsymbol{r_c}]^{n^{\mathrm{g}}} + \check{\boldsymbol{G}} \cdot [\boldsymbol{x_c}]^{n^{\mathrm{g}}} & I^{n^{\mathrm{g}}} + \check{\boldsymbol{G}} \cdot [\boldsymbol{r_c}]^{n^{\mathrm{g}}} - \check{\boldsymbol{B}} \cdot [\boldsymbol{x_c}]^{n^{\mathrm{g}}} \end{bmatrix} \tag{92}$$

REFERENCES

[1] I. Zografopoulos *et al.*, "Cyber-physical energy systems security: Threat modeling, risk assessment, resources, metrics, and case studies," *IEEE Access*, vol. 9, pp. 29 775–29 818, 2021.

[2] A. Jain, J. N. Sakamuri, and N. A. Cutululis, "Grid-forming control strategies for black start by offshore wind power plants," *Wind Energy Science*, vol. 5, no. 4, pp. 1297–1313, 2020.

[3] A. Bidram *et al.*, "Distributed cooperative secondary control of microgrids using feedback linearization," *IEEE Transactions on Power Electronics*, vol. 28, no. 3, pp. 3462–3470, 2013.

[4] A. Bidram, A. Davoudi, and F. L. Lewis, "A multiobjective distributed control framework for islanded ac microgrids," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1785–1798, 2014.

[5] A. Vasilakis *et al.*, "The evolution of research in microgrids control," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 331–343, 2020.

[6] A. P. Kuruvila *et al.*, "Hardware-assisted detection of firmware attacks in inverter-based cyberphysical microgrids," *International Journal of Electrical Power & Energy Systems*, vol. 132, p. 107150, 2021.

[7] M. Long, C.-H. Wu, and J. Hung, "Denial of service attacks on network-based control systems: impact and mitigation," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 2, pp. 85–96, 2005.

[8] S. Shahkar and K. Khorasani, "A resilient control against time-delay switch and denial of service cyber attacks on load frequency control of distributed power systems," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 2020, pp. 718–725.

[9] H. Yuan, Y. Xia, and H. Yang, "Resilient state estimation of cyber-physical system with multichannel transmission under dos attack," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 11, pp. 6926–6937, 2021.

[10] S. Hu *et al.*, "Observer-based event-triggered control for networked linear systems subject to denial-of-service attacks," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1952–1964, 2020.

[11] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using pca approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015.

[12] A. Farraj, E. Hammad, and D. Kundur, "On the impact of cyber attacks on data integrity in storage-based transient stability control," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3322–3333, 2017.

[13] A. Bamigbade, Y. Dvorkin, and R. Karri, "Cyberattack on phase-locked loops in inverter-based energy resources," *IEEE Transactions on Smart Grid*, 2023.

[14] T.-Y. Zhang and D. Ye, "False data injection attacks with complete stealthiness in cyber–physical systems: A self-generated approach," *Automatica*, vol. 120, p. 109117, 2020.

[15] H. Zhang *et al.*, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.

[16] K. Zhang *et al.*, "Stealthy integrity attacks for a class of nonlinear cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 12, pp. 6723–6730, 2022.

[17] ——, "Passive attack detection for a class of stealthy intermittent integrity attacks," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 4, pp. 898–915, 2023.

[18] C. Keliris, M. Polycarpou, and T. Parisini, "A distributed fault detection filtering approach for a class of interconnected continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 58, p. 2032–2047, 08 2013.

[19] G. Park *et al.*, "Stealthy adversaries against uncertain cyber-physical systems: Threat of robust zero-dynamics attack," *IEEE Transactions on*

Automatic Control, vol. 64, no. 12, pp. 4907–4919, 2019.

[20] A. Hoehn and P. Zhang, "Detection of replay attacks in cyber-physical systems," in *2016 American Control Conference (ACC)*, 2016, pp. 290–295.

[21] Y. Li *et al.*, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2831–2836, 2015.

[22] M. Zhu and S. Martínez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 804–808, 2014.

[23] Y. Wu and J. Dong, "Cyber-physical attacks against state estimators based on a finite frequency approach," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 2, pp. 864–874, 2021.

[24] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1157–1168, 2017.

[25] B. Chen *et al.*, "Secure fusion estimation for bandwidth constrained cyber-physical systems under replay attacks," *IEEE transactions on cybernetics*, vol. 48, no. 6, pp. 1862–1876, 2017.

[26] J. Zhao, L. Mili, and A. Gómez-Expósito, "Constrained robust unscented kalman filter for generalized dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3637–3646, 2019.

[27] P. Bernard, *Observer design for nonlinear systems*. Springer, 2019, vol. 479.

[28] A. Zemouche, M. Boutayeb, and G. Bara, "Observer design for nonlinear systems: An approach based on the differential mean value theorem." in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec 2005, pp. 6353–6358.

[29] M. Pajic *et al.*, "Robustness of attack-resilient state estimators," in *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2014, pp. 163–174.

[30] R. Moghadam and H. Modares, "Resilient autonomous control of distributed multiagent systems in contested environments," *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 3957–3967, 2018.

[31] A.-Y. Lu and G.-H. Yang, "Observer-based control for cyber-physical systems under denial-of-service with a decentralized event-triggered scheme," *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 4886–4895, 2019.

[32] M. M. Rana, L. Li, and S. W. Su, "Cyber attack protection and control of microgrids," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 2, pp. 602–609, 2018.

[33] A. O. Aluko, D. G. Dorrell, and E. E. Ojo, "Observer-based detection and mitigation scheme for isolated microgrid under false data injection attack," in *2021 IEEE Southern Power Electronics Conference (SPEC)*, 2021, pp. 1–6.

[34] X. Luo *et al.*, "Observer-based cyber attack detection and isolation in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 127–138, 2018.

[35] A. Cecilia *et al.*, "On addressing the security and stability issues due to false data injection attacks in dc microgrids—an adaptive observer approach," *IEEE Transactions on Power Electronics*, vol. 37, no. 3, pp. 2801–2814, 2022.

[36] Y. Wang, R. Rajamani, and D. M. Bevly, "Observer design for parameter varying differentiable nonlinear systems, with application to slip angle estimation," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1940–1945, 2016.

[37] R. Rajamani, "Observers for lipschitz nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 3, pp. 397–401, 1998.

[38] A. Zemouche and R. Rajamani, "Observer design for non-globally lipschitz nonlinear systems using hilbert projection theorem," *IEEE Control Systems Letters*, vol. 6, pp. 2581–2586, 2022.

[39] A. Cecilia *et al.*, "Detection and mitigation of false data in cooperative dc microgrids with unknown constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 8, pp. 9565–9577, 2021.

[40] M. Corradini and A. Cristofaro, "A sliding-mode scheme for monitoring malicious attacks in cyber-physical systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2702–2707, 2017.

[41] A. Al-Abassi *et al.*, "An ensemble deep learning-based cyber-attack detection in industrial control system," *IEEE Access*, vol. 8, pp. 83 965–83 973, 2020.

[42] S. Huda *et al.*, "Defending unknown attacks on cyber-physical systems by semi-supervised approach and available unlabeled data," *Information Sciences*, vol. 379, pp. 211–228, 2017.

[43] H. D. Nguyen and Q. Cheng, "An efficient feature selection method for distributed cyber attack detection and classification," in *2011 45th Annual Conference on Information Sciences and Systems*. IEEE, 2011, pp. 1–6.

[44] Y. Li, Y. Wang, and S. Hu, "Online generative adversary network based measurement recovery in false data injection attacks: A cyber-physical approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2031–2043, 2019.

[45] B. Hussain *et al.*, "Deep learning-based ddos-attack detection for cyber–physical system over 5g network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 860–870, 2020.

[46] I. Zografopoulos and C. Konstantinou, "Detection of malicious attacks in autonomous cyber-physical inverter-based microgrids," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 5815–5826, 2022.

[47] N. Forti *et al.*, "Distributed joint attack detection and secure state estimation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 96–110, 2017.

[48] S. Waghmare, F. Kazi, and N. Singh, "Data driven approach to attack detection in a cyber-physical smart grid system," in *2017 Indian Control Conference (ICC)*. IEEE, 2017, pp. 271–276.

[49] P. Wang and M. Govindarasu, "Multi-agent based attack-resilient system integrity protection for smart grid," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3447–3456, 2020.

[50] Y. Levron, J. M. Guerrero, and Y. Beck, "Optimal power flow in microgrids with energy storage," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3226–3234, 2013.

[51] D. E. Olivares, C. A. Cañizares, and M. Kazerani, "A centralized energy management system for isolated microgrids," *IEEE Transactions on smart grid*, vol. 5, no. 4, pp. 1864–1875, 2014.

[52] J. D. Lara, D. E. Olivares, and C. A. Canizares, "Robust energy management of isolated microgrids," *IEEE Systems Journal*, vol. 13, no. 1, pp. 680–691, 2018.

[53] W. Shi *et al.*, "Distributed optimal energy management in microgrids," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1137–1146, 2014.

[54] A. Pizano-Martianez, C. R. Fuerte-Esquivel, and D. Ruiz-Vega, "Global transient stability-constrained optimal power flow using an omib reference trajectory," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 392–403, 2010.

[55] B. Cui and X. A. Sun, "A new voltage stability-constrained optimal power-flow model: Sufficient condition, socp representation, and relaxation," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5092–5102, 2018.

[56] C. Wang *et al.*, "Sdp-based optimal power flow with steady-state voltage stability constraints," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4637–4647, 2019.

[57] M. Nojavan and H. Seyedi, "Voltage stability constrained opf in multi-micro-grid considering demand response programs," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5221–5228, 2020.

[58] R. Zarate-Minano, F. Milano, and A. J. Conejo, "An opf methodology to ensure small-signal stability," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1050–1061, 2011.

[59] D. Pullaguram *et al.*, "Small-signal stability-constrained optimal power flow for inverter dominant autonomous microgrids," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 7, pp. 7318–7328, 2021.

[60] D. Krishnamurthy, W. Li, and L. Tesfatsion, "An 8-zone test system based on iso new england data: Development and application," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 234–246, January 2016.

[61] J. E. Price and J. Goodin, "Reduced network modeling of wecc as a market design prototype," in *2011 IEEE Power and Energy Society General Meeting*, July 2011, pp. 1–6.

[62] M. Basseville *et al.*, *Detection of abrupt changes: theory and application*. prentice Hall Englewood Cliffs, 1993, vol. 104.

[63] "The Geometric Approach Toolbox," [Online]. Available: http://www3.deis.unibo.it/Staff/FullProf/GiovanniMarro/geometric.htm, 2023.

[64] H. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.

[65] X. Wu, C. Shen, and R. Iravani, "A distributed, cooperative frequency and voltage control for microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2764–2776, 2016.

[66] L. Luo and S. V. Dhople, "Spatiotemporal model reduction of inverter-based islanded microgrids," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 823–832, 2014.

[67] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994.

[68] M. Kheirandishfard, F. Zohrizadeh, and R. Madani, "Convex relaxation of bilinear matrix inequalities part i: Theoretical results," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 67–74.

[69] IEEE, "IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," *IEEE Std 1547.4-2011*, pp. 1–54, 2011.

Andres Intriago (Student Member, IEEE) received the B.E. degree in electrical and computer engineering from Escuela Superior Politécnica del Litoral (ESPOL), Ecuador, in 2022. He is currently pursuing the M.S./Ph.D. degree with the Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. His research interests include cyber-physical systems, power electronics, and industrial control systems cybersecurity with application to power systems.

Francesco Liberati received the PhD in Systems Engineering in 2015, from the University of Rome "La Sapienza", Rome, Italy. He has worked in the field of automatic control and research project management at the eCampus University, Novedrate (CO), Italy, at the "Consortium for the Research in Automation and Telecommunication (CRAT)", Rome, Italy, and at the "Innovation and Networks Executive Agency (INEA)" of the European Commission, Brussels, Belgium. He is currently a tenure track researcher at the University of Rome "La Sapienza". His research interests include automatic control problems in the sectors of smart grids, industry, and cyber-physical systems. He is Associate Editor of Advanced Control for Applications, Wiley.

Nikos D. Hatziargyriou (Life Fellow, IEEE) is with the National Technical University of Athens (NTUA), since 1984, Professor in Power Systems, since 1995, and Professor Emeritus, since 2022. He is Part-time Professor at the University of Vaasa, Finland. He has over 10 years of industrial experience as Chairman and the CEO of the Hellenic Distribution Network Operator (HEDNO) and as Executive Vice-Chair and Deputy CEO of the Public Power Corporation (PPC), responsible for the Transmission and Distribution Divisions. He has participated in more than 60 R&D projects funded by the EU Commission, electric utilities and industry for fundamental research and practical applications. He has authored or coauthored more than 300 journal publications and 600 conference proceedings papers. He was the Chair and Vice-Chair of ETIP-SNET. He is past EiC of the IEEE Transactions on Power Systems and currently EiC-at-Large for IEEE PES Transactions. He is included in the 2016, 2017 and 2019 Thomson Reuters lists of top 1% most cited researchers. He is the 2020 Globe Energy Prize laureate, recipient of the 2017 IEEE/PES Prabha S. Kundur Power System Dynamics and Control Award and recipient of the 2023 IEEE Herman Halperin Electric Transmission and Distribution Award.

Charalambos Konstantinou (SM'20) received the M.Eng. degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 2012, and the Ph.D. degree in electrical engineering from New York University (NYU), NY, USA, in 2018. He is currently an Assistant Professor of Electrical and Computer Engineering (ECE) and an Affiliate Professor of Computer Science (CS) with the Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. He is the Principal Investigator of the Secure Next Generation Resilient Systems Laboratory (sentry.kaust.edu.sa) and a member of the Resilient Computing and Cybersecurity Center (RC3), KAUST. His research interests include critical infrastructures security and resilience with special focus on smart grid technologies, renewable energy integration, and real-time simulation. Before joining KAUST, he was an Assistant Professor with the Center for Advanced Power Systems (CAPS), Florida State University (FSU). He is the Chair of the IEEE Task Force on Resilient and Secure Large-Scale Energy Internet Systems and the Co-Chair of the IEEE Task Force on Cyber-Physical Interdependence for Power System Operation and Control. He is a member of ACM and an ACM Distinguished Speaker (2021–2024). Prof. Konstantinou is an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.