Electric Power Systems Research 260 (2026) 113148

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Cyber attack detection in renewable-rich IBR-dominated microgrids: A natural watermarking approach

Imasha Balahewa, Lars Bjorndal, Chris Mi, Tong Huang*

Department of Electrical and Computer Engineering, San Diego State University, 5500 Campanile Dr, San Diego, 92182, CA, USA

ARTICLE INFO

Keywords:
Microgrids
Cyber security
Natural watermarking
Controller hardware-in-the-loop (C-HIL) simulations
Inverter-based resources (IBRs)

ABSTRACT

This paper proposes a natural watermarking approach for detecting cyber attacks in a renewable-rich microgrid (MG). The approach leverages the inherent variability of renewable energy generation to watermark the measurements feeding inverter-based resources (IBRs), enabling the IBRs to locally detect attacks. It detects cyber attacks by checking the statistics of the natural watermark exhibited in IBR measurements. It is physically interpretable, computationally efficient, and scalable for large-scale MGs. Its effectiveness is validated through pure-software simulations and real-time controller hardware-in-the-loop experiments in a MG test system.

1. Introduction

Microgrids are emerging as a cornerstone of future power systems, thanks to their key role in the transition towards smart grids. They support increased integration of renewables into the grids, while improving the flexibility, reliability, and resilience of the system [1]. Microgrids are cyber-physical systems that integrate various distributed energy resources (DERs) such as solar photovoltaic (PV) arrays, wind turbines, and diesel generators, as well as energy storage systems such as batteries. Additionally, microgrids (MGs) incorporate communication and control infrastructure to enable real-time monitoring, coordinated control, and stable autonomous operation of the microgrid. The MGs further improve the overall system reliability, resilience, and flexibility while reducing operational costs by sharing the available generation resources. These features make the MGs a driving force in shaping the future of smart grids. Despite the numerous advantages, MGs are potentially attractive targets for cyber-intruders who exploit the microgrids' vulnerabilities to launch destructive attacks, compromising the reliability and resilience of the MGs [2]. The cyber vulnerabilities of the MGs arise from the tight coupling between their physical and cyber layers. The cyber layer, comprising control and communication networks, is vulnerable to intrusions. Hence, the MGs are also vulnerable to cyber attacks, similar to other industrial control systems.

Different cyber attack detection methods have been proposed in the literature to address cyber threats associated with electric energy systems. These methods can be classified into two main categories, namely, data-driven methods and model-based methods [3]. The data-driven methods involve offline training of a machine learning (ML) model using real-world historical or real-time data to learn statistical or temporal patterns in the data. During online implementation, actual data from cyber-physical systems are compared against the learned patterns to detect cyber attacks [4]. There are several conventional ML techniques used in literature, and these can be classified into methods based on supervised learning, unsupervised learning, and semi-supervised learning [5]. Supervised learning methods use labeled data for the learning process of the models. Support vector machine (SVM) [6], K-nearest neighbor (KNN) [7], decision tree (DT) [8], long-short term memory (LSTM) network [9], artificial neural network (ANN) [10], and convolutional neural network (CNN) [11] are the supervised learning methods that are widely used for the detection of false data injection attacks (FDIAs) in the literature. Unsupervised learning methods learn the patterns from unlabeled data which requires a large amount of training. This approach has the ability to detect new attack scenarios which is an added advantage. K-means clustering (KMC) [12], autoencoder (AE) [13], and isolation forest (IF) [14] are some classical unsupervised learning methods used in literature. Semi-supervised learning uses both labeled and unlabeled data for model training and is a widely used method for attack detection. Semi-supervised adversarial autoencoder (SSAA) [15], generative-adversarial based semi-supervised (GBSS) learning framework [16] and robust semi-supervised prototypical network (RSSPN) [17] are some semi-supervised learning methods used to detect FDIAs. The above three categories do not require a system model for attack detection, which is an advantage. However, they require an extensive amount of data for training purposes, and the reliability of attack detection outcomes heavily depends on the quality of the training data.

* Corresponding author.
E-mail addresses: ibalahewa4996@sdsu.edu (I. Balahewa), lbjorndal@sdsu.edu (L. Bjorndal), cmi@sdsu.edu (C. Mi), thuang7@sdsu.edu (T. Huang).

https://doi.org/10.1016/j.epsr.2026.113148
Received 17 October 2025; Received in revised form 23 March 2026; Accepted 13 April 2026
Available online 19 May 2026

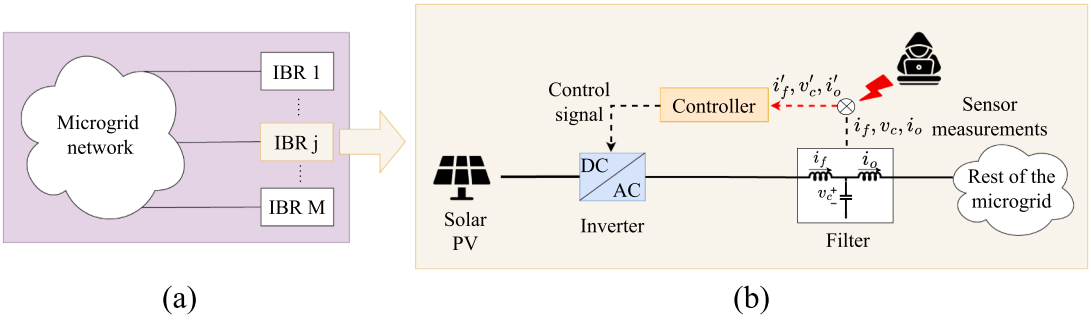


Fig. 1. Cyber and physical architecture of a microgrid.

Model-based methods require physical models of the system. Some model-based methods use first principles to predict measurements and compare these predictions with the actual measurements to compute discrepancies. These discrepancies are evaluated against a predefined threshold corresponding to the no-attack condition to detect attacks effectively [3,18,19]. Some model-based methods are based on state estimation [20–22]. These methods have a lower computational burden, require less data, and are also physically interpretable when compared to data-driven methods for attack detection. However, despite all the advantages, they depend on an accurate model of the system, which can be challenging to develop in complex and large-scale systems. Under the category of model-based methods, the dynamic watermarking approach has shown potential for detecting false data injection attacks in IBR systems [23–28]. In this approach, a small noise (called a "watermark" signal) is superimposed on the control command of the inverter to detect cyber attacks in IBR systems. This externally injected watermark signal is solely for attack detection, but it may compromise the IBR control performance.

In this paper, we propose a natural watermarking approach that overcomes the limitations of the original version of the dynamic watermarking approach. This is achieved by leveraging renewable fluctuations to watermark IBR measurements, thereby identifying measurements that are under cyber attacks. The novelty of the paper can be summarized as follows: 1) Compared to the original dynamic watermarking in [23–27,29], which injects external noise, this approach uses the inherent variability of renewable energy resources for attack detection, thereby avoiding performance degradation and implementation complexity. 2) Compared with the machine learning-based detection methods in [5,6,9,10], the proposed method is physically interpretable and does not require large, high quality training datasets. 3) The performance of the proposed method is validated through both pure-software simulations and controller-hardware-in-the-loop (C-HIL) simulations. This paper extends the conference paper, Anonymous [30], in the following aspects: 1) It tests the proposed natural watermarking approach through simulating multiple networked IBRs with switching dynamics. 2) It validates the approach by performing real-time C-HIL experiments using Texas Instruments C2000 microcontroller interfaced with Typhoon HIL 506 real-time simulator. 3) It includes detailed comparison studies between the conventional dynamic watermarking approach and the proposed natural watermarking approach.

The rest of the paper is structured as follows. Section 2 describes the dynamics and cyber vulnerabilities of MGs; Section 3 describes the use of natural watermarking approach for attack detection in MGs; Section 4 tests the effectiveness of the proposed approach in a microgrid using both pure-software simulations and real-time C-HIL simulations; and Section 5 summarizes the contributions of the paper.

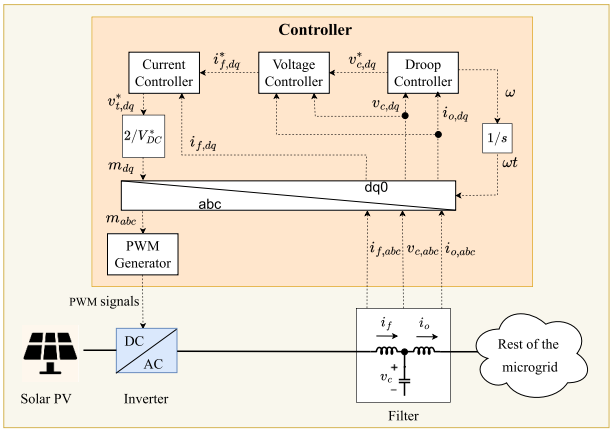


Fig. 2. An IBR controller [31].

2. Cyber vulnerability of IBR-dominated microgrids

2.1. Architecture of microgrids

Fig. 1a shows the physical architecture of a microgrid comprising M IBRs. The M IBRs are interconnected through a microgrid network. Fig. 1b presents the architecture of the jth IBR. The physical layer of the IBR includes a solar PV panel, an inverter, and an LCL low-pass filter. The cyber layer includes sensors and a controller. The controller computes the pulse width modulation (PWM) signals based on the three-phase currents, $\mathbf{i}_f$ and $\mathbf{i}_o$, and voltage $\mathbf{v}_c$. Fig. 2 presents one example of the IBR control architecture, which includes a droop controller, voltage controller, current controller, abc-dq0 transformation block, and a PWM generator.

As shown in Fig. 2, the sensors feeding the controller measure the three-phase currents and voltages at the filter which includes the inverter output current, $\mathbf{i}_f$, filter output current, $\mathbf{i}_o$, and the capacitor voltage, $\mathbf{v}_c$. Then these measurements are sent through an abc-dq0 transformation block, which converts them into the dq0 frame denoted by $\mathbf{i}_{f,dq}$, $\mathbf{i}_{o,dq}$, and $\mathbf{v}_{c,dq}$ using the reference angle, ωt, generated by the droop controller. The droop controller is fed with $\mathbf{i}_{o,dq}$ and $\mathbf{v}_{c,dq}$, and it calculates the frequency, ω, and the reference voltages, $\mathbf{v}^*_{c,dq}$, which are the set points of the voltage controller. The voltage controller then calculates the current set points of the current controller, $\mathbf{i}^*_{f,dq}$, using the reference and actual voltages, and actual currents denoted by $\mathbf{v}^*_{c,dq}$, $\mathbf{v}_{c,dq}$, and $\mathbf{i}_{o,dq}$, respectively. The current controller uses the current set points provided

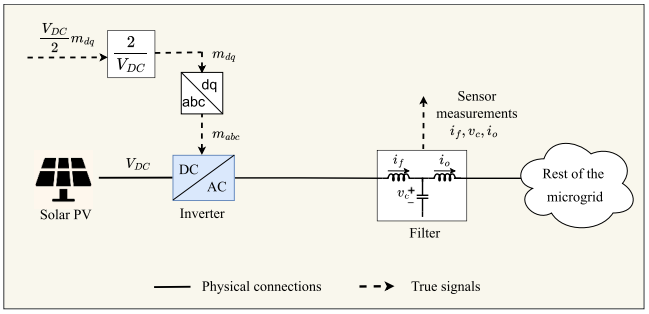


Fig. 3. Open-loop dynamics.

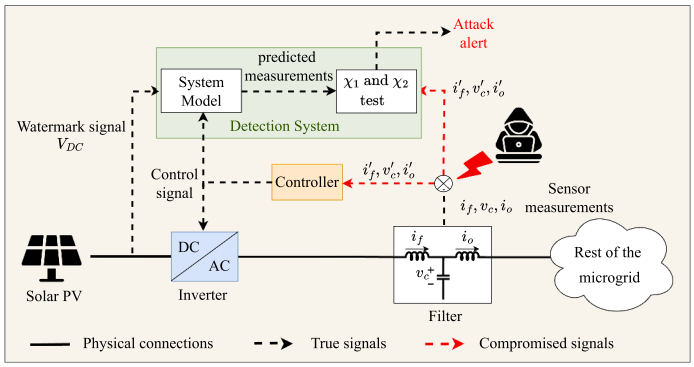


Fig. 4. Implementation of the natural watermarking approach.

by the voltage controller along with the actual currents $\mathbf{i}_{f,dq}$, to calculate a voltage setpoint, $\mathbf{v}^*_{t,dq}$. This voltage setpoint is then used to compute the modulation indices in the d-q frame, m_{dq}. These indices are subsequently transformed back to the abc frame and used to generate the PWM signals for the inverter.

2.2. False data injection attacks

As the controller computes the control signals based on the terminal measurements $\mathbf{y} := [\mathbf{i}_f^\top, \mathbf{i}_o^\top, \mathbf{v}_c^\top]^\top$, these measurements can be attractive targets for cyber attacks. For example, as shown in Fig. 1b, an attacker may launch FDIAs by forcing the sensors to report different measurements $\mathbf{z}$ ($\neq \mathbf{y}$) to the IBR controller. As a result, the performance of the IBR can be compromised. Since the IBR connects to the MG, the negative impacts of local FDI attacks may propagate to its host microgrid, leading to instability or sustained oscillations throughout the microgrid.

Examples of the FDIAs include noise injection attacks, replay attacks, destabilization attacks, and stealth attacks. In noise injection attacks, the attackers can superimpose large noise onto the actual measurements $\mathbf{y}$, compromising the efficiency of the IBRs [24]. In replay attacks, the attackers record a segment of measurements and replace the actual measurements with the pre-recorded ones. As a result, actual events, such as voltage/current changes, are invisible to the IBR controllers. In destabilization attacks [24], an attacker can change a negative feedback to a positive feedback by manipulating sensor readings, in order to destabilize the system. Since some advanced IBR controllers are designed based on the statistics of measurement noises, the stealth attacks [27] compromise the performance of these controllers by replacing the actual measurement noise with fake noise in the measurements.

This paper focuses on detecting FDIAs on terminal measurements $\mathbf{y}$, i.e., *given the reported measurement* $\mathbf{z}$*, decide if* $\mathbf{z}$ *equals* $\mathbf{y}$. The terminal measurements, $\mathbf{y}$, are measured by hall sensors and they can be manipulated by externally changing the magnetic field of the sensors. A motivation example where one can manipulate sensor readings in a non-intrusive manner is shown in [32].

3. Natural watermarking approach

This section presents the proposed natural watermarking approach for detecting cyber attacks on terminal measurements, $\mathbf{y}$. We first present the open-loop dynamics seen by an IBR controller. Based on the open-loop dynamics, we introduce the cyber attack detection criteria in the proposed natural watermark approach. Finally, we present a data-driven method to identify the open-loop dynamics.

3.1. Open-loop dynamics

Fig. 3 presents the open-loop dynamics seen by the controller of the jth IBR in the MG, modeled as a multi-input multi-output (MIMO) system. These dynamics capture the influence of the rest of the microgrid on the jth IBR and form the basis for implementing the natural watermarking approach in a microgrid context. The inputs of the open-loop dynamics are $\frac{1}{2}V_{DC}\mathbf{m}_{dq}$ where $\mathbf{m}_{dq} \in R^2$ collects the modulation indices in the d-q frame, and V_{DC} is the DC link voltage. The output of the open-loop dynamics is the terminal measurements, $\mathbf{y}$. The open-loop dynamics seen by the controller can be described by the following state-space model:

$$\dot{\mathbf{x}} = A\mathbf{x} + \frac{1}{2}BV_{DC}\mathbf{m}_{dq} + \epsilon, \qquad \mathbf{y} = C\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ collects the state variables governing the dynamics of the physical layer of the jth IBR and the rest of the MG; ϵ is process noise resulting from external disturbances; $\mathbf{n}$ is the measurement noise; and matrices A, B, and C are the system, input, and output matrices which can be analytically derived based on the differential equations in A. It is worth noting that the dynamics (1) are non-linear due to the bilinear term "$V_{DC}\mathbf{m}_{dq}$," if the DC link voltage V_{DC} fluctuates. Identifying localized state space models in this manner enables decentralized and scalable cyber attack detection across MGs, as it allows each IBR to locally capture the dynamics within inverter-based microgrids.

Based on the open-loop dynamics (1), we proceed to explain the basic idea of the natural watermarking approach. Changes in environmental conditions, such as temperature and irradiation levels, cause V_{DC} to fluctuate over time. The term $V_{DC}\mathbf{m}_{dq}$ also fluctuates, which enables $V_{DC}\mathbf{m}_{dq}$ to act as a natural watermark signal to detect cyber attacks in the terminal measurements $\mathbf{y}$. If there is no cyber attack in the reported measurements $\mathbf{z}$, i.e., $\mathbf{z} = \mathbf{y}$, the reported measurement $\mathbf{z}$ should reflect the signature of the renewable fluctuation, as the changes in V_{DC} propagate to the terminal measurements through the open-loop dynamics (1). If a cyber attack occurs, the signature of the renewable fluctuations will be distorted in the reported measurement $\mathbf{z}$. Therefore, for a renewable-powered IBR, to implement the dynamic watermarking approach, there is no need for external noise injections, which may compromise the control performance.

3.2. Implementation of the natural watermarking approach

How do we check the existence of the signature of renewable fluctuations in the reported measurements $\mathbf{z}$*?* Fig. 4 summarizes the overall implementation of the proposed natural watermarking approach. Implementing cyber attack detection using this approach requires key system parameters and signals, such as the renewable fluctuation, V_{DC}, control signal of the inverter, sensor measurements, and the state space model of the system. Using V_{DC} fluctuations and the control signal of the inverter as inputs to the system model, the outputs of the system can be accurately predicted. The measurements reported by the sensors are then compared with the measurements predicted by the model to identify any discrepancies. The discrepancy between the sensor-reported measurements and the model-predicted measurements is calculated by, $\Delta y_j[n] = |z_j[n] - \hat{y}_j[n]|$, where $\hat{y}_j[n]$ is the jth predicted measurement in

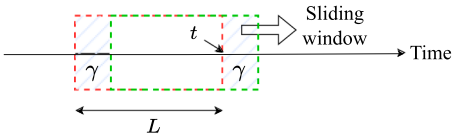


Fig. 5. Sliding window approach.

y at time n based on the open-loop dynamics (1), and $z_j[n]$ is the jth reported measurement in **z** at time n.

The two attack indicators, $\chi_{1j}[t]$ and $\chi_{2j}[t]$, at time t are computed using moving average and moving variance methods, respectively, in a sliding-window fashion, based on the following equations:

$$\chi_{1j}[t] = \frac{1}{L}\sum_{n=t-L+1}^{t} \Delta y_j[n], \qquad \chi_{2j}[t] = \frac{1}{L}\sum_{n=t-L+1}^{t} \left(\Delta y_j[n] - \chi_{1j}\right)^2 \quad (2)$$

where L is the window width. A graphical interpretation of the sliding window approach is depicted in Fig. 5. The current window at time t, which is indicated by the red dashed box, consists of L data points ranging from time $t-L+1$ to time t. $\chi_{1j}[t]$ and $\chi_{2j}[t]$ are then computed using the data collected in this window. As $\gamma \in \{1, 2, \ldots, L\}$ new data points are received, the sliding window is updated as indicated by the green dashed box by removing the oldest γ points and adding γ new points. The data points in the newly updated window are used to update the values of $\chi_{1j}[t]$ and $\chi_{2j}[t]$. This process is repeated to generate a sequence of χ_{1j} and χ_{2j}, which indicates the presence of an attack affecting the jth sensor in the system through a significant increase in their values compared to normal operation or no-attack scenario. This is how attacks are locally detected at each IBR system in the MG without centralized coordination.

3.3. System model identification

The remaining question is how to obtain the model for the open-loop dynamics (1). We leverage system identification techniques as in [33] to answer this question. The input for the system identification is $\mathbf{m}_{\text{dq}}\frac{V_{\text{DC}}}{2}$. The outputs for the system identification are the d-q components of the three-phase currents, i_f, i_o, and voltage, v_c. At the system identification stage, we perturb the input and measure the corresponding output. In our test system (will be introduced in Section 4), we change V_{DC}, which has a mean of 800 V, by adding normally distributed (Gaussian) fluctuations with a variance of 100 V every 10 ms. The System Identification Toolbox in MATLAB is used to obtain the state space model of the MG system from input-output data. To get the most accurate state space model for the system, several model estimations with different model orders are considered, and the one with the closest fit to the actual data is chosen. Fig. 6a and b show the performance of the system identification, where the predicted output captures the general trend of the actual output. This indeed highlights the accuracy of the predictions made by the identified model. The proposed local input-output based system identification can be extended to a recursive system identification as discussed in [27] to update the identified model online using recent local measurements.

4. Case study

4.1. Microgrid test system

The test system represents a renewable-rich multi-IBR system as shown in Fig. 7. It consists of five interconnected IBR systems, including solar PV systems and battery storage systems, all interfaced via inverters and supplying different loads.

The MG test system shown in Fig. 7 is modeled on the Typhoon HIL 506 real-time simulator. The solar PV systems are modeled as described in [30], each with different output voltages to represent diverse and realistic renewable generation profiles. Out of the five inverters, the three inverters interfaced with solar PV are modeled with switching dynamics at a switching frequency of 20 kHz while the inverters interfaced with battery storage are modeled using the average model inverters due to the real-time simulation constraints in the simulator.

All the IBR systems in the MG are modeled using the same control loops as in [30] and [34], consisting of droop, voltage, and current controls. A detailed block diagram of the controller is shown in Fig. 2. The combination of renewable generation using solar PV and battery storage systems provides a comprehensive test environment for evaluating the natural watermarking approach in detecting cyber attacks in microgrids.

4.2. Software simulations of a microgrid with 5 IBRs

4.2.1. Evaluation of natural watermarking approach using software-only simulations

As the initial step, the microgrid test system depicted in Fig. 7 is modeled in the Typhoon HIL schematic editor and run as real-time software-only simulations to test the performance of the proposed natural watermarking approach against different types of attacks, namely, noise injection attack, replay attack, simultaneous multi-sensor noise injection attack, gradual amplitude reduction attack, and stealth attack. An in-depth analysis of each attack scenario is provided as follows.

A noise injection attack is where Gaussian noise is superimposed onto the sensor measurements, distorting the signals used in the control loops and adversely affecting the system's normal operation. Such an attack scenario occurring at 10.5 s on the filter output current, i_{oa}, of IBR 1 is shown in Fig. 8a. The proposed attack indicators, χ_1 and χ_2, indicate a sharp rise from their baseline values after the attack is launched, signaling an attack in the system as shown in Fig. 8b and c.

A replay attack occurs when actual sensor measurements are replaced with previously recorded measurements and then replayed over some time interval to evade detection. This is a more sophisticated type of attack, as it does not produce any visible distortions in the sensor measurements. Fig. 9a shows how such an attack is launched at 10.5s on the filter output current, i_{oa}, of IBR 1. This waveform indicates normal operation of the system, despite an ongoing attack. However, as seen in Fig. 9b and c, the proposed attack indicators exhibit an increase in their values compared to their baseline values, indicating anomalous behavior in the system.

A simultaneous multi-sensor attack is where an attacker manipulates multiple sensor measurements at the same time. Such a simultaneous noise injection attack in which both the output current, i_{oa}, and the capacitor voltage, v_{cb}, of IBR 1, are manipulated starting at 10.5s by superimposing noise, is shown in Fig. 10a and d, respectively. As shown in Fig. 10b, c, e, and f, the attack indicators $\chi_{1_{i_{oa}}}$, $\chi_{2_{i_{oa}}}$, $\chi_{1_{v_{cb}}}$ and $\chi_{2_{v_{cb}}}$ corresponding to each of the compromised sensor measurements increase sharply from their baseline values indicating the presence of simultaneous attacks in multiple sensors.

A gradual amplitude reduction attack is an example of a false data injection attack in which the attacker slowly reduces the amplitude of the sensor measurements over time instead of abruptly introducing it. Here, a gradual amplitude reduction attack on the output current, i_{oa}, of IBR 1 starting at 5.5s is considered. As shown in Fig. 11a and b, the attacker slowly reduces the amplitude of the current fed into the controller, leading the controller to produce more current to fix the gradual reduction in the manipulated current feedback signal. Fig. 11c and d show the behavior of the attack indicators χ_1 and χ_2 during a gradual amplitude reduction attack. It can be seen that the attack is not detected immediately, but with a delay of about 4.5s. This is because the attack itself is a slow attack with gradual manipulations and it is evident that there will be a delay in attack detection. The gradual changes in the current amplitude introduce small discrepancies at each time step, and it requires sufficient accumulated data to exceed the set thresholds in order to indicate the presence of an attack. This shows that the natural watermarking approach can detect abruptly initiated attacks as well as

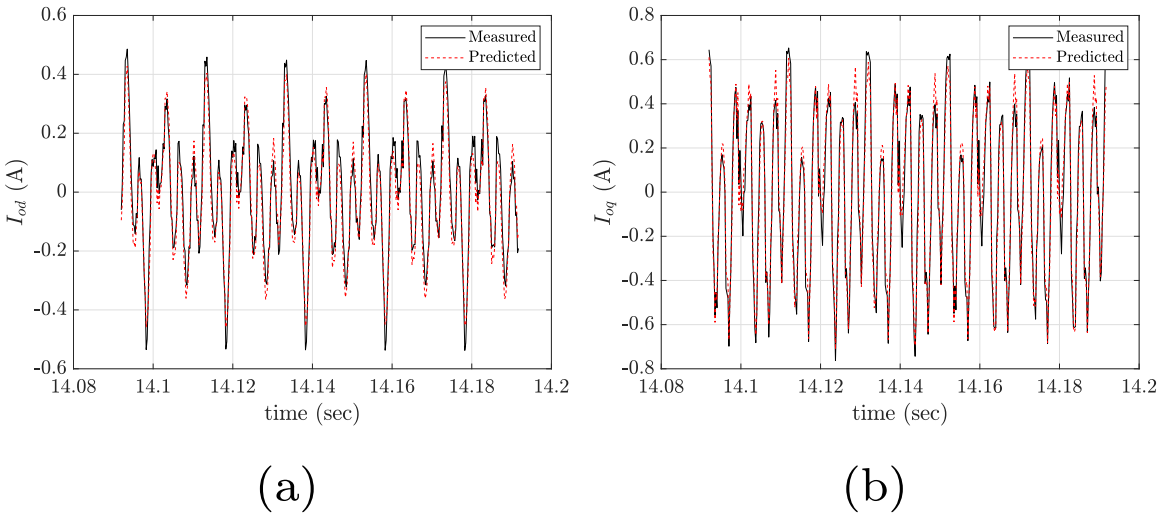


Fig. 6. Performance of system identification for the outputs (a) i_{od}; (b) i_{oq}.

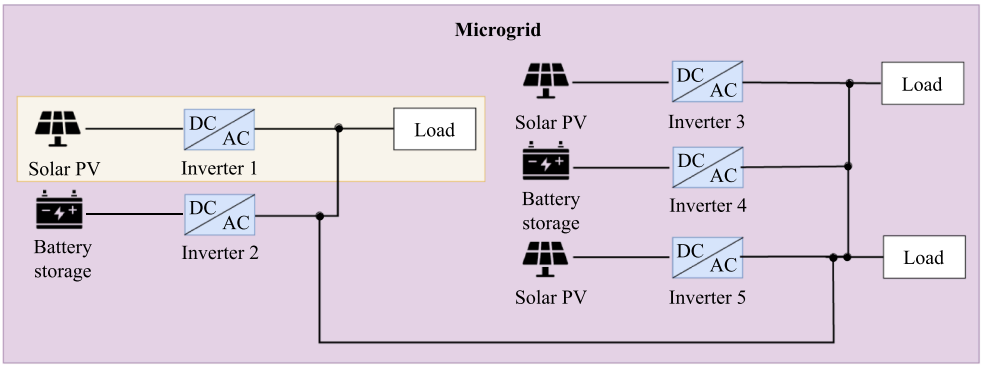


Fig. 7. Microgrid test system.

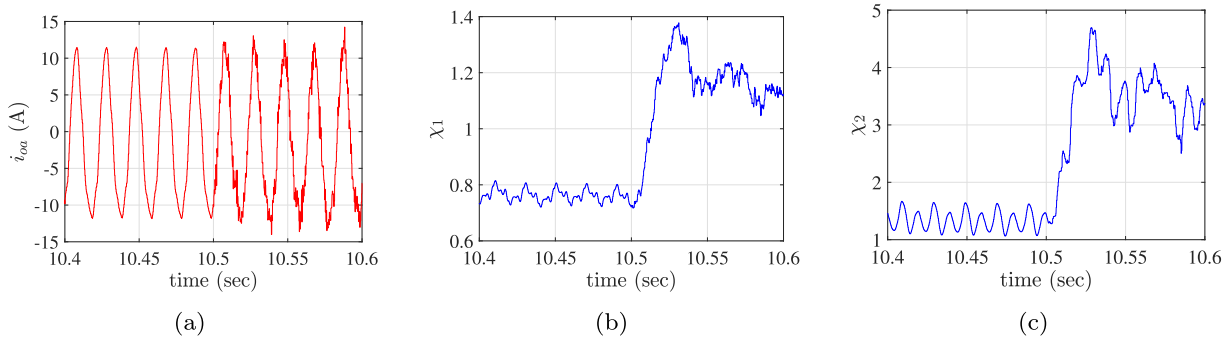


Fig. 8. Noise injection attack at 10.5 s and its detection in pure-software simulations: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

gradual or slow attacks. However, it should be noted that the detection sensitivity for gradual attacks is lower than that for abruptly initiated attacks due to unavoidable detection delay.

A stealth attack is an advanced type of cyber attack where the attacker uses knowledge of the system model and available measurements to predict the behavior of the system. Here, we consider a scenario where the attacker predicts sensor measurements based on their system model and its inputs, and attempts to forcefully send these false predictions to the controller. As a result, the compromised sensors will be reporting values of the form $\hat{\mathbf{z}} + \mathbf{n}'$, where $\hat{\mathbf{z}}$ represents the predicted measurements by the attacker and $\mathbf{n}'$ is a small noise which is different from the actual measurement noise $\mathbf{n}$. This makes stealth attack detection challenging compared to other attacks, as the attacker's predicted measurements, now fed into the controller, closely resemble the expected system behavior. The effectiveness of the proposed detection system against a stealth attack on the filter output current, i_{oa}, is shown in Fig. 12. Despite the absence of any visible anomalies in the waveform, the attack indicators, χ_1 and χ_2, show deviations from their baseline values, indicating the successful detection of this intrusion.

This raises the question of whether an attacker can model the exact system and predict sensor measurements based on the system model, and also precisely replicate the expected statistical patterns of renewable fluctuations to launch an advanced stealthy attack. While this may appear to be feasible in theory, it is extremely challenging in practice,

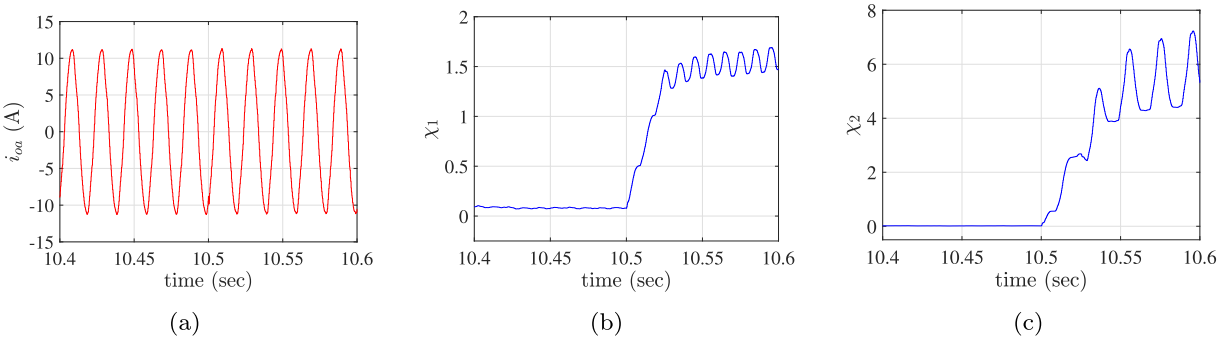


Fig. 9. Replay attack at 10.5 s and its detection in pure-software simulations: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

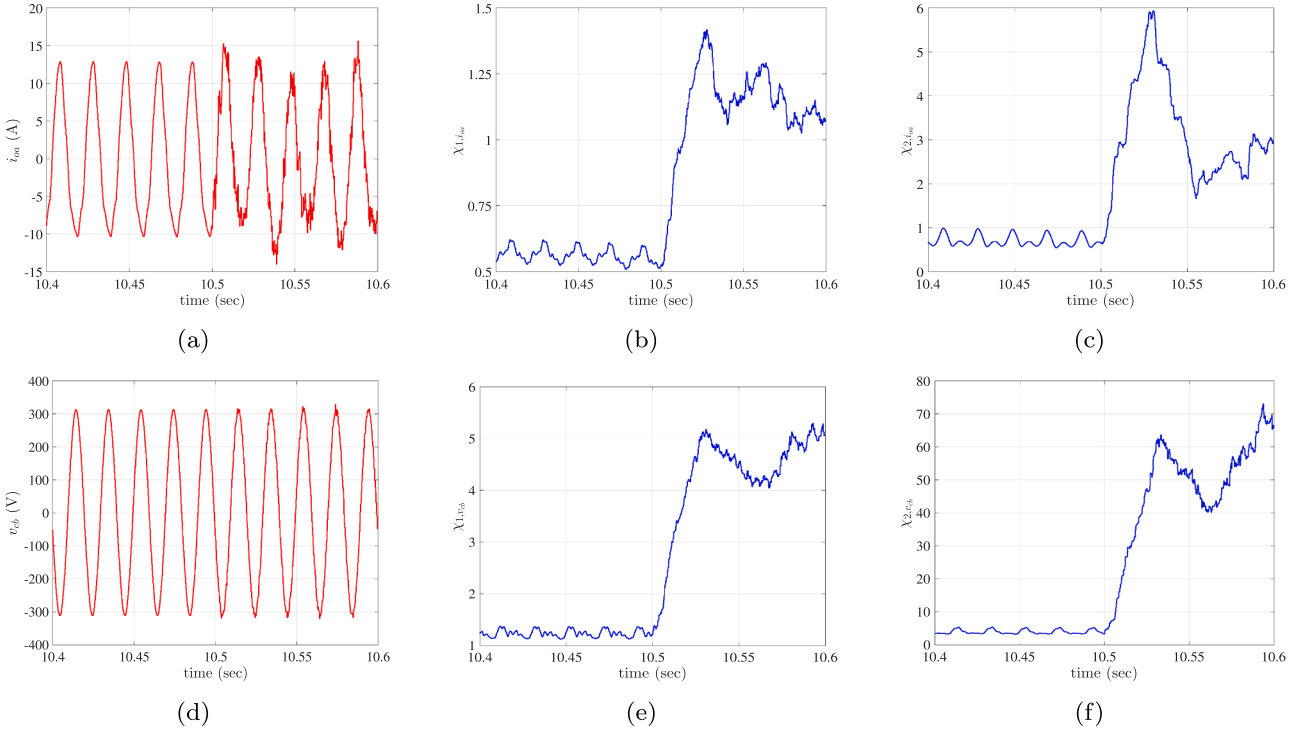


Fig. 10. Simultaneous multi-sensor noise injection attack starting at 10.5s: (a) attacked output current i_{oa}; (b) detection using $\chi_{1,i_{oa}}$; (c) detection using $\chi_{2,i_{oa}}$; (d) attacked capacitor voltage v_{cb}; (e) detection using $\chi_{1,v_{cb}}$; (f) detection using $\chi_{2,v_{cb}}$.

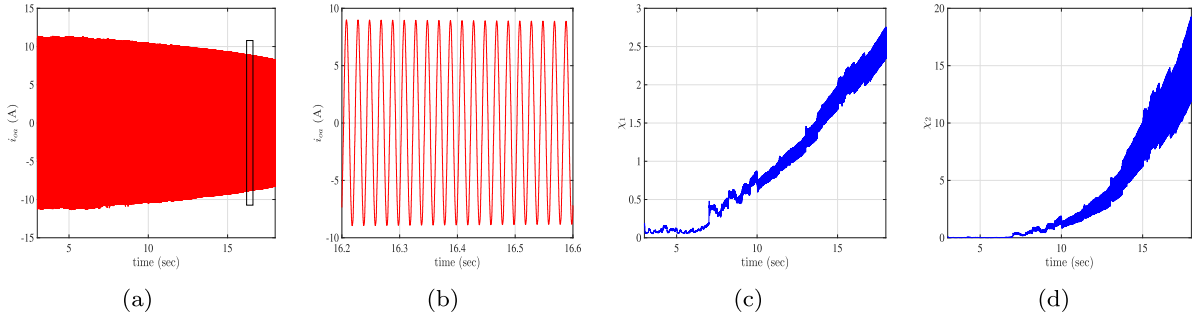


Fig. 11. Gradual amplitude reduction attack starting at 5.5s: (a) attacked output current i_{oa}; (b) zoomed-in plot of i_{oa}; (c) detection using χ_1; (d) detection using χ_2.

even for a sophisticated attacker, to replicate the behavior of V_{DC}. This is because V_{DC} depends not only on stochastic factors such as solar irradiation, temperature, partial shading, cloud movements, and wind speed, but also on panel-specific factors such as degradation of PV modules, accumulation of dust, dirt and other contaminants, cell damage or micro cracks and manufacturer-specific characteristics like cell material and configuration [35,36]. Even if an attacker could access or accurately predict the external stochastic factors, it is highly unlikely that they have access to panel-specific factors of the solar PV arrays in real-time, which makes it infeasible to replicate V_{DC} accurately.

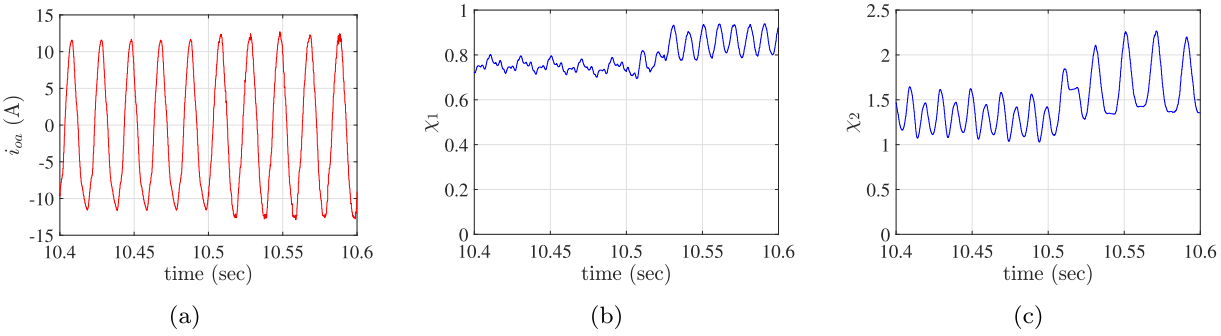


Fig. 12. Stealth attack at 10.5 s and its detection in pure-software simulations: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

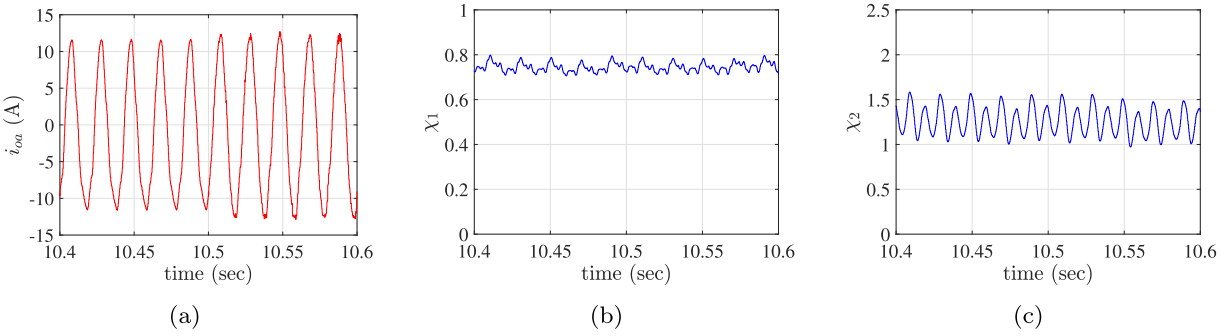


Fig. 13. Undetected stealth attack at 10.5 s due to the absence of the watermark signal in the sensor measurements in the pure-software simulation of the MG system: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

4.2.2. *Necessity of the natural watermark*

Next, we demonstrate the necessary condition for attack detection in our approach, which is the presence of the natural watermark. Let us consider a scenario where the natural watermark is not present in the sensor measurements, which means that our watermark signal, V_{DC}, remains constant without any fluctuations. The response of the detection system to a stealth attack in such a scenario is shown in Fig. 13. It can be seen that the attack is not detected by our indicators, χ_1 and χ_2, as they remain almost constant. Therefore, this indeed validates the necessity of having natural watermark signals for detecting cyber attacks using our approach. In the case of solar PV, when its output voltage, V_{DC}, which serves as the watermark signal, remains constant for a while, our detection algorithm can still detect cyber intrusions in the system as soon as the PV output changes due to changes in the environmental conditions.

This result highlights the impact of renewable variability on the attack detection capability of the proposed approach, i.e., higher variability in V_{DC} improves detection performance. To ensure timely attack detection during periods of minimal renewable fluctuation or relatively constant V_{DC}, a hybrid strategy can be employed. Here, natural watermarking can be used when there is sufficient renewable variability, while it can be switched to the conventional dynamic watermarking approach when renewable variability is low. This further indicates the fact that natural watermarking and conventional dynamic watermarking are complementary approaches.

4.2.3. *Comparison studies*

The dynamic watermarking approach is a well-known strategy for detecting cyber attacks in power systems. Therefore, a performance comparison is carried out to evaluate the effectiveness of the proposed natural watermarking approach against the original dynamic watermarking approach. To implement dynamic watermarking in the MG testbed, a small Gaussian noise with zero mean is added to the control signals of Inverter 1. Fig. 14a and b show the behavior of the above two methods during a replay attack, while Fig. 14c and d show the behavior during a noise injection attack. Overall results from this comparison study show that the natural watermarking approach shows similar detection performance to the dynamic watermarking approach, even without external noise injections into the system.

The impact of the externally injected noise in dynamic watermarking on the control performance of the inverter is evaluated quantitatively using total harmonic distortion (THD) as a key metric. The THD of both the filter output current, (i_{oa}), and the voltage at the point of common coupling, (v_{PCC}), are compared under both dynamic watermarking and natural watermarking approaches. As shown in Fig. 15a, the THD of i_{oa} in the dynamic watermarking approach has a higher magnitude and a higher variability than that of the natural watermarking approach. This indicates that the externally injected watermark signal in dynamic watermarking could interfere with current regulation standards, leading to a degradation in power quality and affecting the IBR control performance. A similar scenario is seen in Fig. 15b for v_{PCC}. This highlights the importance of employing a non-intrusive detection method, such as natural watermarking, to preserve the IBR control performance and maintain power quality.

To further evaluate the system performance under both dynamic watermarking and natural watermarking approaches, voltage regulation and transient response are examined using a change in the capacitor d-axis reference voltage, $v_{cd,ref}$, generated by a step change in the nominal voltage of the droop controller. The responses of v_{cd} under both approaches are then compared as shown in Fig. 16. To quantify voltage regulation in steady state, we compute the variance of v_{cd} before and after the change in the reference voltage. The results in Table 1 show that the dynamic watermarking approach produces a higher voltage variability than the natural watermarking approach. This is because dynamic watermarking uses external noise injection into the control command of the inverter. Therefore, natural watermarking achieves a better voltage regulation than dynamic watermarking. As shown in Fig. 16, the transient response of v_{cd} during the change in the reference voltage is

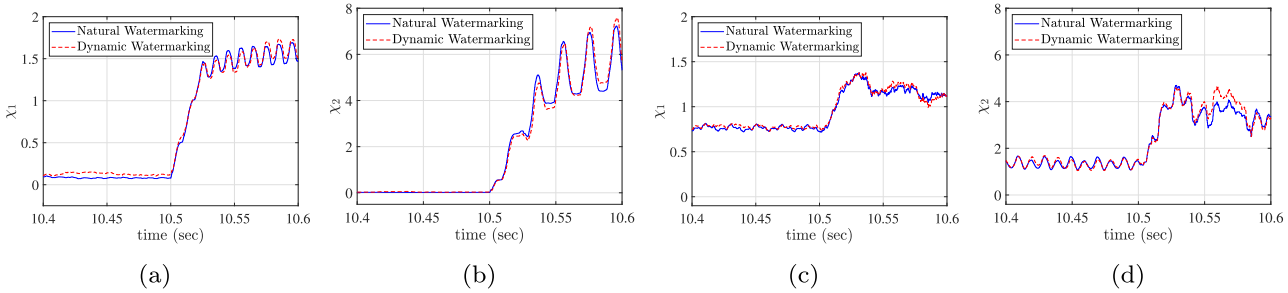


Fig. 14. Performance comparison between natural and dynamic watermarking approaches: (a) Replay attack: χ_1; (b) Replay attack: χ_2; (c) Noise injection: χ_1; (d) Noise injection: χ_2.

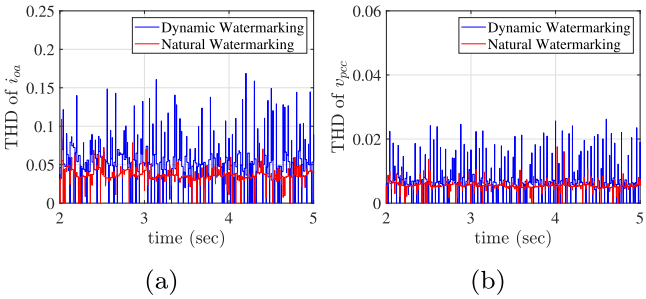


Fig. 15. Comparison of total harmonic distortion (THD) under dynamic and natural watermarking: (a) filter output current i_{oa}; (b) PCC voltage v_{PCC}.

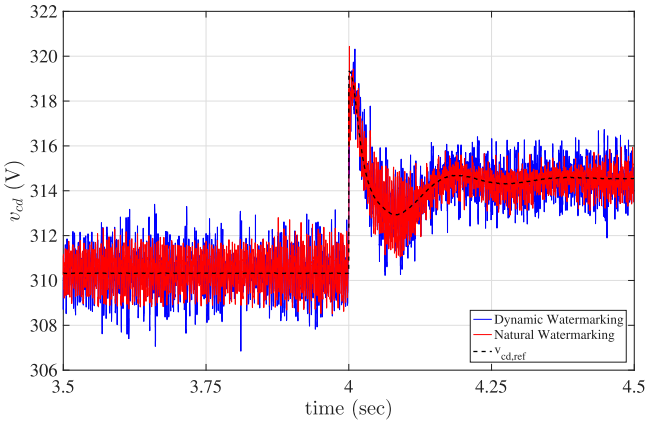


Fig. 16. Voltage reference tracking of v_{cd} under dynamic and natural watermarking following a step change in the voltage reference $v_{cd,\text{ref}}$.

Table 1
Steady state variability of v_{cd} under dynamic watermarking (DW) and natural watermarking (NW) before and after reference change.

	Variance of v_{cd} (V^2)	
	DW	**NW**
before step change	1.1348	0.6664
after step change	0.6677	0.2350

comparable between the dynamic watermarking approach and the natural watermarking approach.

In addition, key studies on conventional watermarking, such as [26, 27], do not account for the variability of renewable fluctuations when developing their prediction model. They typically use the modulation index as the only input to the system model, without considering the effect of V_{DC} variations caused by renewable fluctuations. As a result, natural fluctuations in V_{DC} lead to false alarms in attack detection as shown in Fig. 17, even when there is no attack in the system.

4.2.4. *Sensitivity of detection performance to the quality of the identified model*

To evaluate the sensitivity of attack indicators to inaccuracies in the identified open-loop state space model, a set of experiments is performed using state space models of different model accuracies defined by the "best fit" metric in the System Identification Toolbox of MATLAB. The best fit is the percentage of variations in the measured data that are reproduced by the identified model, which can be mathematically expressed as, fitting accuracy (%) $= \left(1 - \frac{\|Y-\hat{Y}\|}{\|Y-\text{mean}(Y)\|}\right) \times 100\%$, where Y and $\hat{Y}$ are actual and predicted measurements, respectively.

Fig. 18a and b show the behavior of the attack indicator, χ_1, in a 5-IBR system during a replay attack launched at 10.5s, when using state space models of accuracies 20% and 66%, respectively, in the detection system. The thresholds for each scenario are calculated by multiplying the highest value recorded by the indicator during normal operation of the system by a factor of two in order to keep a safety margin and are indicated by black dashed lines in Fig. 18a and b. As seen in Fig. 18a, when using a model of 20% accuracy, the attack detection begins around 11.11s when the value of χ_1 starts to exceed the threshold. This results in an attack detection delay of 0.61s. Similarly, for a model with an

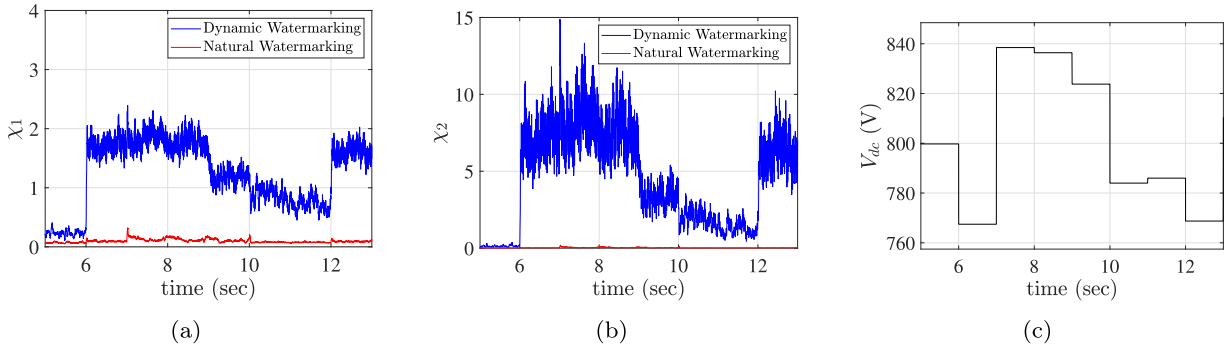


Fig. 17. Attack indicators in the dynamic watermarking approach exhibiting false detections due to V_{DC} fluctuations even though there is no attack in the system: (a) χ_1; (b) χ_2; (c) V_{DC}.

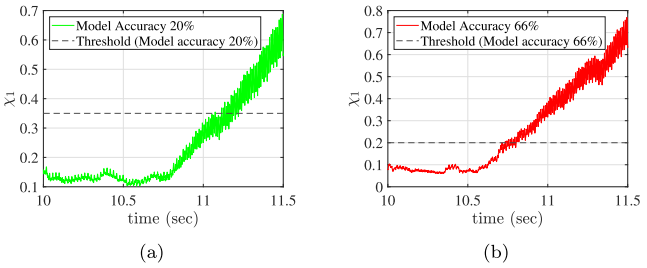


Fig. 18. Attack detection delay in χ_1 when using models with different accuracies during a replay attack: (a) model accuracy is 20%; (b) model accuracy is 66%.

accuracy of 66% as in Fig. 18b, the attack detection begins around 10.73s, leading to an attack detection delay of 0.23 seconds. This indicates that a higher-accuracy model would result in shorter detection delays. Hence, this highlights the sensitivity of the attack indicators to inaccuracies of the identified models.

4.3. Controller hardware-in-the-loop (C-HIL) simulations

4.3.1. C-HIL testbed configuration

The MG test system in Fig. 7 is implemented using a C-HIL approach as shown in Fig. 19. Here, the controller of one of the IBR systems, in this case IBR 1 of the microgrid, is executed on the Texas Instruments C2000 LAUNCHXL-F280049C microcontroller and is interfaced with a Typhoon HIL 506 real-time simulator on which the rest of the system is modeled. This creates a realistic testing environment where the actual controller interacts with a high-fidelity real-time simulation of the plant, bridging the gap between virtual and physical testing. The controller shown in Fig. 2 and the detection system shown in Fig. 4 are modeled using the C2000 microcontroller blockset in Simulink, enabling direct code generation and flashing onto the C2000. The rest of the MG system and the cyber attacks are modeled using Typhoon HIL's Schematic Editor software on the real-time simulator.

The sensor measurements of the first IBR system, including the inverter current, i_f, the capacitor voltage, v_c, and the filter output current, i_o, along with the solar PV output voltage, V_{DC}, from the Typhoon HIL are received by the analog inputs of the C2000 microcontroller. The PWM signals computed in real-time by the microcontroller are sent to the Typhoon HIL via the digital inputs of the simulator to ensure the operation of the switching model inverter. This completes the closed-loop system of the IBR 1 in the MG. The attack indicator signals, χ_{1i} and χ_{2i}, computed by the detection system on the microcontroller are also sent to the Typhoon HIL via the analog inputs of the simulator in order to issue the attack alerts.

The control loops on the C2000 and the signal processing components on the Typhoon HIL are executed at 4 kHz while the other electrical components on the Typhoon HIL are executed with a much higher resolution at a fixed simulation time step of 2 μs. This distinction between control and simulation resolution supports accurate modeling of electrical dynamics and seamless real-time controller interaction. To achieve small simulation time steps during the real-time emulation of the MG test system, which consists of a large complex electrical circuit with power electronics and power system models, the circuit is parallelized via electric circuit partitioning. Each partition is then executed on the Typhoon FPGA solver, which is an FPGA-based multi-core processor to support high-speed and real-time simulations.

Having established the C-HIL architecture, the next step is to evaluate the effectiveness of the proposed natural watermarking approach for cyber attack detection in a high-fidelity real-time testing environment that mimics real-world operating conditions. Here, we implement the same attack scenarios described in Section 4.2.1 and check the performance of the detection algorithm. As previously discussed, the attacks are modeled on the Typhoon HIL real-time simulator, targeting IBR 1 of the MG, while the detection system runs on the microcontroller. The microcontroller continuously monitors sensor measurements, computes attack indicators in real-time, and sends them to the Typhoon HIL for attack alerts. Next, we present experimental results for the aforementioned attack scenarios and their detection.

4.3.2. Noise injection attack

Here, a noise injection attack is launched at 5.5 s on the filter output current, i_{oa}, in IBR 1 as shown in Fig. 20a. The two attack indicators, χ_1 and χ_2, computed in real-time on the C2000 start to show a significant increase in their values when compared with the no-attack scenario as seen in Fig. 20b and c, indicating that our algorithm has successfully detected this attack in a more realistic environment.

4.3.3. Replay attack

Here, a replay attack is launched at 5.5 s on i_{oa} as before, which can be seen in Fig. 21a. The waveform does not show any visible distortions as a result of this attack and hence, it appears that the system is in normal operation. However, the behavior of the two attack indicators, χ_1 and χ_2, shown in Fig. 21b and c indicates a substantial rise from the nominal values signaling prompt attack detection.

4.3.4. Stealth attack

A stealth attack scenario similar to the one discussed in Section 4.2.1 is considered here. When the attack is launched at 5.5 s, the indicators start to deviate from their nominal values as in Fig. 22, further illustrating the effectiveness of our approach in detecting more sophisticated attacks. Therefore, the experimental results from C-HIL implementation, which demonstrate successful detection of each of these attacks, further validate the effectiveness of our natural watermarking approach for cyber attack detection in a more realistic and practical testing environment.

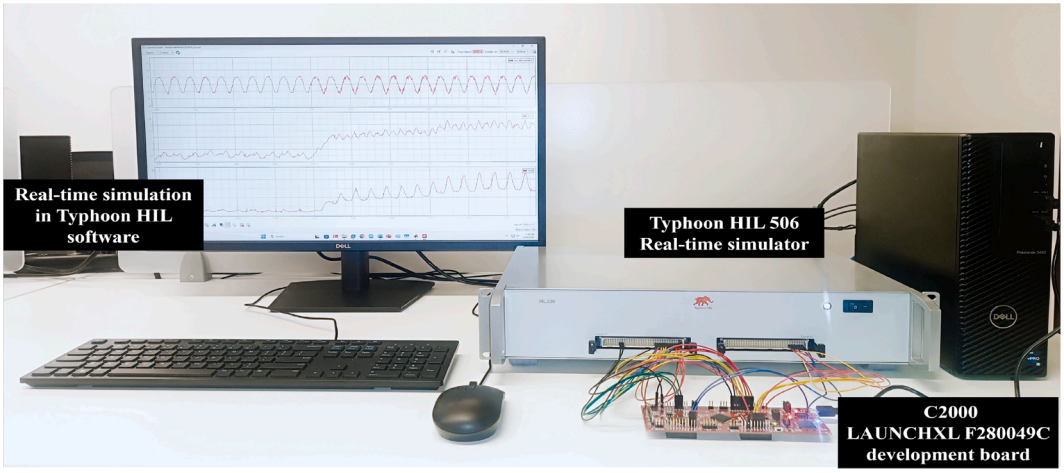


Fig. 19. Controller hardware-in-the-loop setup.

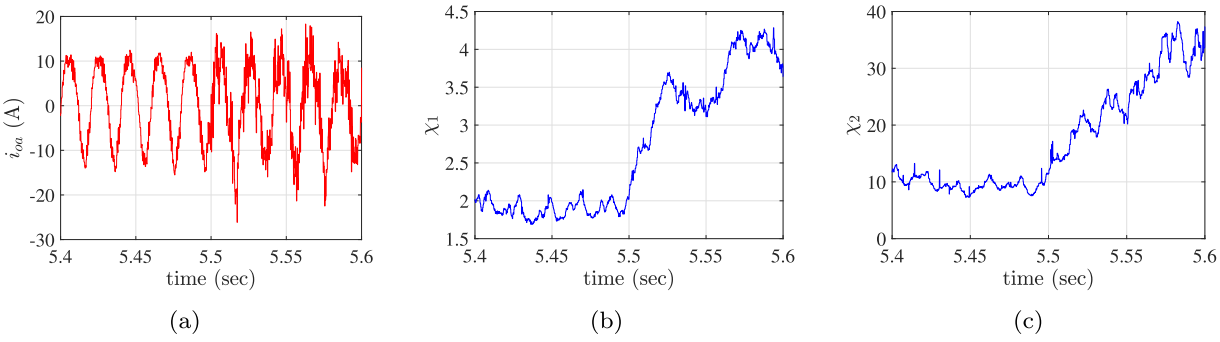


Fig. 20. Noise injection attack at 5.5s and its detection in the C-HIL implementation: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_{2_1} for sensor 1 which measures i_{oa}.

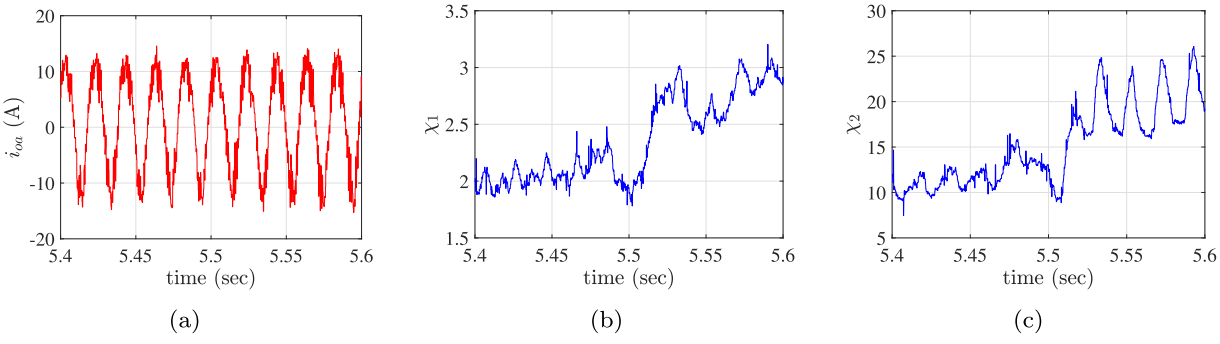


Fig. 21. Replay attack at 5.5s and its detection in the C-HIL implementation: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

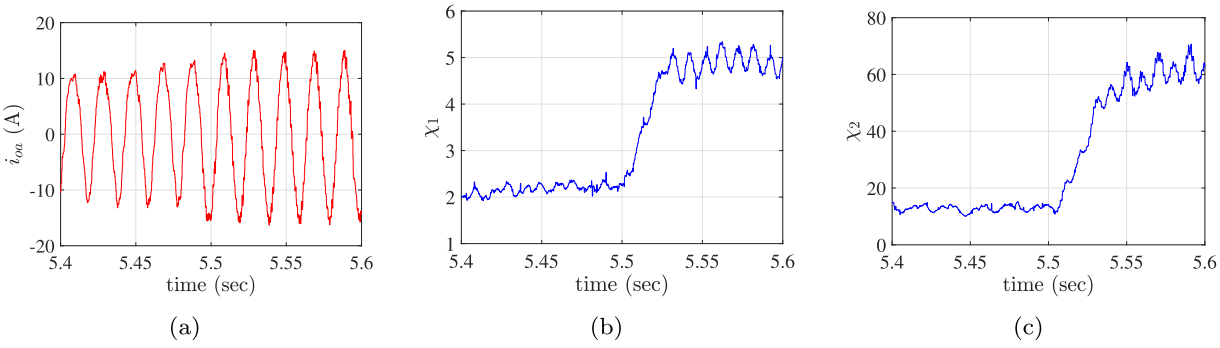


Fig. 22. Stealth attack at 5.5s and its detection in the C-HIL implementation: (a) Waveform of i_{oa} under attack; (b) Detection using χ_1 for sensor 1 which measures i_{oa}; (c) Detection using χ_2 for sensor 1 which measures i_{oa}.

5. Conclusion

This paper proposes a natural watermarking approach for cyber attack detection in renewable-rich MGs. The method leverages the inherent variability of renewable energy generation to naturally watermark the measurements feeding inverter-based resources, enabling each IBR to locally detect attacks by monitoring the statistics of the natural watermark embedded in the IBR measurements. This detection strategy improves scalability for large-scale MGs and reduces system complexity by eliminating the need for external noise injections onto the control signals of inverters, as in the conventional watermarking approach. The effectiveness of the proposed method is demonstrated through both pure-software simulations and real-time C-HIL simulations in a MG testbed, against noise injection, replay, gradual amplitude reduction, simultaneous multi-sensor noise injection and stealth attacks. Simulation and experimental results demonstrate that the natural watermarking approach can be effectively applied in MGs for cyber attack detection. Future work will test and validate the natural watermarking approach in a pure hardware testbed.

CRediT authorship contribution statement

Imasha Balahewa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis; **Lars Bjorndal:** Writing – review & editing, Methodology; **Chris Mi:** Writing – review & editing, Supervision; **Tong Huang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Dynamics of IBR-dominated microgrids

As shown in Fig. 2, the dynamics of the jth IBR are governed by an LCL filter, a droop controller, a voltage controller, and a current controller. Next, we elaborate on these blocks as follows.

A.1. Reference frame transformation

A vector $[x_d, x_q]^\top$ in the local dq frame can be transformed to the common DQ frame according to:

$$\begin{bmatrix} x_D \\ x_Q \end{bmatrix} = \begin{bmatrix} \cos\delta_j & -\sin\delta_j \\ \sin\delta_j & \cos\delta_j \end{bmatrix} \begin{bmatrix} x_d \\ x_q \end{bmatrix} \tag{A.1}$$

A.2. LCL filter

The LCL filter dynamics of the jth IBR can be described as below.

$$\frac{di_{fj,d}}{dt} = -\frac{r_{fj}}{L_{fj}} i_{fj,d} + \omega i_{fj,q} - \frac{1}{L_{fj}} v_{cj,d} + \frac{1}{L_{fj}} v_{invj,d} \tag{A.2a}$$

$$\frac{di_{fj,q}}{dt} = -\omega i_{fj,d} - \frac{r_{fj}}{L_{fj}} i_{fj,q} - \frac{1}{L_{fj}} v_{cj,q} + \frac{1}{L_{fj}} v_{invj,q} \tag{A.2b}$$

$$\frac{dv_{cj,d}}{dt} = \frac{1}{C_{fj}} i_{fj,d} + \omega v_{cj,q} - \frac{1}{C_{fj}} i_{oj,d} \tag{A.2c}$$

$$\frac{dv_{cj,q}}{dt} = \frac{1}{C_{fj}} i_{fj,q} - \omega v_{cj,d} - \frac{1}{C_{fj}} i_{oj,q} \tag{A.2d}$$

$$\frac{di_{oj,d}}{dt} = \frac{1}{L_{cj}} v_{cj,d} - \frac{r_{cj}}{L_{cj}} i_{oj,d} + \omega i_{oj,q} - \frac{1}{L_{cj}} v_{bj,d} \tag{A.2e}$$

$$\frac{di_{oj,q}}{dt} = \frac{1}{L_{cj}} v_{cj,q} - \omega i_{oj,d} - \frac{r_{cj}}{L_{cj}} i_{oj,q} - \frac{1}{L_{cj}} v_{bj,q} \tag{A.2f}$$

Here, $i_{fj,d}, i_{fj,q}$ are the dq components of the inverter-side inductor current, $v_{cj,d}, v_{cj,q}$ are the dq components of the capacitor voltage, and $i_{oj,d}, i_{oj,q}$ are the dq components of the load-side inductor current in the LCL filter of the jth IBR. $v_{invj,d}, v_{invj,q}$ are the dq components of the inverter output voltage, and $v_{bj,d}, v_{bj,q}$ are the dq components of the voltage at the point of common coupling of the jth IBR.

A.3. Droop controller

Droop controller dynamics are described using the differential equations below.

$$\dot{P}_j = -\omega_c P_j + \omega_c (v_{cdj} i_{odj} + v_{cqj} i_{oqj}) \tag{A.3a}$$

$$\dot{Q}_j = -\omega_c Q_j + \omega_c (v_{cdj} i_{oqj} - v_{cqj} i_{odj})) \tag{A.3b}$$

$$\dot{\delta}_j = \omega_j - \omega_{common} \tag{A.4}$$

A.4. Voltage controller

The state and algebraic equations for the voltage controller are described below.

$$\dot{\phi}_{dj} = v^*_{cdj} - v_{cdj}, \quad \dot{\phi}_{qj} = v^*_{cqj} - v_{cqj} \tag{A.5a}$$

$$i^*_{idj} = F_j i_{odj} - \omega_n C_{fj} v_{cqj} + K_{pvj}(v^*_{cdj} - v_{cdj}) + K_{ivj}\phi_{dj} \tag{A.5b}$$

$$i^*_{iqj} = F_j i_{oqj} + \omega_n C_{fj} v_{cdf} + K_{pvj}(v^*_{cqj} - v_{cqj}) + K_{ivj}\phi_{qj} \tag{A.5c}$$

A.5. Current controller

The state and algebraic equations for the current controller are described below.

$$\dot{\gamma}_{dj} = i^*_{idj} - i_{idj}, \quad \dot{\gamma}_{qj} = i^*_{iqj} - i_{iqj} \tag{A.6}$$

$$v^*_{invj,d} = K_{pcj}(i^*_{idj} - i_{idj}) + K_{icj}\gamma_{dj} - \omega_n L_{fj} i_{iqj} \tag{A.7}$$

$$v^*_{invj,q} = K_{pcj}(i^*_{iqj} - i_{iqj}) + K_{icj}\gamma_{qj} + \omega_n L_{fj} i_{idj} \tag{A.8}$$

A.6. RL tie-line connecting the jth and kth IBR systems

The dynamics of the RL tie line can be described as below.

$$\frac{di_{line,D}}{dt} = \frac{1}{L_{line}}(v_{bj,D} - v_{bk,D}) - \frac{r_{line}}{L_{line}} i_{line,D} + \omega i_{line,Q} \tag{A.9a}$$

$$\frac{di_{line,Q}}{dt} = \frac{1}{L_{line}}(v_{bj,Q} - v_{bk,Q}) - \frac{r_{line}}{L_{line}} i_{line,Q} - \omega i_{line,D} \tag{A.9b}$$

A.7. Load model

The dynamics of the RL load in the jth IBR system are described below.

$$\frac{di_{\mathrm{load,D}}}{dt} = -\frac{R_{\mathrm{load}}}{L_{\mathrm{load}}} i_{\mathrm{load,D}} + \omega i_{\mathrm{load,Q}} + \frac{1}{L_{\mathrm{load}}} v_{bj,D} \tag{A.10a}$$

$$\frac{di_{\mathrm{load,Q}}}{dt} = -\frac{R_{\mathrm{load}}}{L_{\mathrm{load}}} i_{\mathrm{load,Q}} - \omega i_{\mathrm{load,D}} + \frac{1}{L_{\mathrm{load}}} v_{bj,Q} \tag{A.10b}$$

Using the aforementioned dynamics of each component in the individual IBR subsystems, the state space model of the MG system can be obtained by combining the state space models of the subsystems, as described in [34]. However, due to the complexity of analytically deriving the full state space model for large-scale MGs, the System Identification Toolbox in MATLAB is used in this paper to obtain the overall state space model of the MG from simulation data.

References

[1] Q. Zhou, M. Shahidehpour, A. Paaso, S. Bahramirad, A. Alabdulwahab, A. Abusorrah, Distributed control and communication strategies in networked microgrids, IEEE Commun. Surv. Tutor. 22 (4) (2020) 2586–2633. https://doi.org/10.1109/COMST.2020.3023963

[2] P.S. Tadepalli, D. Pullaguram, Distributed control microgrids: cyber-attack models, impacts and remedial strategies, IEEE Trans. Signal Inf. Process. Netw. 8 (2022) 1008–1023. https://doi.org/10.1109/TSIPN.2022.3230562

[3] N.D. Tuyen, N.S. Quan, V.B. Linh, V. Van Tuyen, G. Fujita, A comprehensive review of cybersecurity in inverter-based smart power system amid the boom of renewable energy, IEEE Access 10 (2022) 35846–35875. https://doi.org/10.1109/ACCESS.2022.3163551

[4] A. Takiddin, S. Rath, M. Ismail, S. Sahoo, Data-driven detection of stealth cyber-attacks in DC microgrids, IEEE Syst. J. 16 (4) (2022) 6097–6106. https://doi.org/10.1109/JSYST.2022.3183140

[5] D. Du, M. Zhu, X. Li, M. Fei, S. Bu, L. Wu, K. Li, A review on cybersecurity analysis, attack detection, and attack defense methods in cyber-physical power systems, J. Mod. Power Syst. Clean Energy 11 (3) (2023) 727–743. https://doi.org/10.35833/MPCE.2021.000604

[6] A.A. Khan, O.A. Beg, M. Alamaniotis, S. Ahmed, Intelligent anomaly identification in cyber-physical inverter-based systems, Electr. Power Syst. Res. 193 (2021) 107024. https://doi.org/10.1016/j.epsr.2021.107024

[7] M.A. Hasnat, M. Rahnamay-Naeini, Detecting and locating cyber and physical stresses in smart grids using the k-nearest neighbour analysis of instantaneous correlation of states, IET Smart Grid 4 (3) (2021) 307–320. https://doi.org/10.1049/stg2.12030

[8] R. Punmiya, S. Choe, Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing, IEEE Trans. Smart Grid 10 (2) (2019) 2326–2329. https://doi.org/10.1109/TSG.2019.2892595

[9] P. Ganesh, X. Lou, Y. Chen, R. Tan, D.K.Y. Yau, D. Chen, M. Winslett, Learning-based simultaneous detection and characterization of time delay attack in cyber-physical systems, IEEE Trans. Smart Grid 12 (4) (2021) 3581–3593. https://doi.org/10.1109/TSG.2021.3058682

[10] M.R. Habibi, H.R. Baghaee, F. Blaabjerg, T. Dragičević, Secure MPC/ANN-based false data injection cyber-attack detection and mitigation in DC microgrids, IEEE Syst. J. 16 (1) (2022) 1487–1498. https://doi.org/10.1109/JSYST.2021.3086145

[11] M. Ismail, M.F. Shaaban, M. Naidu, E. Serpedin, Deep learning detection of electricity theft cyber-attacks in renewable distributed generation, IEEE Trans. Smart Grid 11 (4) (2020) 3428–3437. https://doi.org/10.1109/TSG.2020.2973681

[12] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, I. Chueiri, A tunable fraud detection system for advanced metering infrastructure using short-lived patterns, IEEE Trans. Smart Grid 10 (1) (2019) 830–840. https://doi.org/10.1109/TSG.2017.2753738

[13] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, X. Duan, Distributed framework for detecting PMU data manipulation attacks with deep autoencoders, IEEE Trans. Smart Grid 10 (4) (2019) 4401–4410. https://doi.org/10.1109/TSG.2018.2859339

[14] S. Ahmed, Y. Lee, S.-H. Hyun, I. Koo, Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest, IEEE Trans. Inf. Forensics Secur. 14 (10) (2019) 2765–2777. https://doi.org/10.1109/TIFS.2019.2902822

[15] Y. Zhang, J. Wang, B. Chen, Detecting false data injection attacks in smart grids: a semi-supervised deep learning approach, IEEE Trans. Smart Grid 12 (1) (2021) 623–634. https://doi.org/10.1109/TSG.2020.3010510

[16] M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far, M. Saif, M. Parvania, Adversarial semi-supervised learning for diagnosing faults and attacks in power grids, IEEE Trans. Smart Grid 12 (4) (2021) 3468–3478. https://doi.org/10.1109/TSG.2021.3061395

[17] T. Zheng, Y. Liu, Y. Yan, S. Xiong, T. Lin, Y. Chen, Z. Wang, X. Jiang, RSSPN:robust semi-supervised prototypical network for fault root cause classification in power distribution systems, IEEE Trans. Power Deliv. 37 (4) (2022) 3282–3290. https://doi.org/10.1109/TPWRD.2021.3125704

[18] S. Tan, J.M. Guerrero, P. Xie, R. Han, J.C. Vasquez, Brief survey on attack detection methods for cyber-physical systems, IEEE Syst. J. 14 (4) (2020) 5329–5339. https://doi.org/10.1109/JSYST.2020.2991258

[19] J. Ye, A. Giani, A. Elasser, S.K. Mazumder, C. Farnell, H.A. Mantooth, T. Kim, J. Liu, B. Chen, G.-S. Seo, W. Song, M.D.R. Greidanus, S. Sahoo, F. Blaabjerg, J. Zhang, L. Guo, B. Ahn, M.B. Shadmand, N.R. Gajanur, M.A. Abbaszada, A review of cyber-physical security for photovoltaic systems, IEEE J. Emerg. Sel. Top. Power Electron. 10 (4) (2022) 4879–4901. https://doi.org/10.1109/JESTPE.2021.3111728

[20] T. Vu, B.H.L. Nguyen, T.A. Ngo, M. Steurer, K. Schoder, R. Hovsapian, Distributed optimal dynamic state estimation for cyber intrusion detection in networked DC microgrids, in: Proc. Annu. Conf. IEEE Ind. Electron. Soc. (IECON), 2019, pp. 4050–4055.

[21] N. Muralidhar, ILLIAD: intelligent invariant and anomaly detection in cyber-physical systems, ACM Trans. Intell. Syst. Technol. 9 (3) (2018) 1–20.

[22] G. Anagnostou, F. Boem, S. Kuenzel, B.C. Pal, T. Parisini, Observer-based anomaly detection of synchronous generators for power systems monitoring, IEEE Trans. Power Syst. 33 (4) (2018) 4228–4237.

[23] B. Satchidanandan, P.R. Kumar, Dynamic watermarking: active defense of networked cyber-physical systems, Proc. IEEE 105 (2) (2017) 219–240. https://doi.org/10.1109/JPROC.2016.2575064

[24] T. Huang, B. Satchidanandan, P.R. Kumar, L. Xie, An online detection framework for cyber attacks on automatic generation control, IEEE Trans. Power Syst. 33 (6) (2018) 6816–6827. https://doi.org/10.1109/TPWRS.2018.2829743

[25] T. Huang, D. Wu, M. Ilić, Cyber-resilient automatic generation control for systems of AC microgrids, IEEE Trans. Smart Grid 15 (1) (2024) 886–898. https://doi.org/10.1109/TSG.2023.3272632

[26] H.A.J. Ibrahim, et al., Detection of cyber attacks in grid-tied PV systems using dynamic watermarking, IEEE Trans. Ind. Appl. 60 (1) (2024) 819–827. https://doi.org/10.1109/TIA.2023.3321588

[27] W.-H. Ko, J.A. Ramos-Ruiz, T. Huang, J. Kim, H. Ibrahim, P.N. Enjeti, P.R. Kumar, L. Xie, Robust dynamic watermarking for cyber-physical security of inverter-based resources in power distribution systems, IEEE Trans. Ind. Electron. 71 (7) (2024) 7106–7116. https://doi.org/10.1109/TIE.2023.3303614

[28] T. Huang, J. Ramos-Ruiz, W.-H. Ko, J. Kim, P. Enjeti, P.R. Kumar, L. Xie, Enabling secure peer-to-peer energy transactions through dynamic watermarking in electric distribution grids: defending the distribution system against sophisticated cyberattacks with a provable guarantee, IEEE Electrif. Mag. 9 (3) (2021) 55–64. https://doi.org/10.1109/MELE.2021.3093600

[29] T. Huang, B. Wang, J. Ramos-Ruiz, P. Enjeti, P.R. Kumar, L. Xie, Detection of cyber attacks in renewable-rich microgrids using dynamic watermarking, in: 2020 IEEE Power & Energy Society General Meeting (PESGM), 2020, pp. 1–5. https://doi.org/10.1109/PESGM41954.2020.9282071

[30] I. Balahewa, L. Bjorndal, C. Mi, T. Huang A Natural Watermarking Approach to Cyber Attack Detection for Power Electronics- Interfaced Renewables, in: 56th North American Power Symposium (NAPS), El Paso, TX, USA, 2024, pp. 1–6. https://doi.org/10.1109/NAPS61145.2024.10741694

[31] T. Huang, Non-intrusive Enforcement of Decentralized Stability Protocol for IBRs in AC Microgrids, arXiv preprint arXiv:2310.09450 (2023).

[32] A. Barua, M.A. Al Faruque, Hall spoofing: a {non-invasive}{DoS} attack on {grid-tied} solar inverter, in: 29th USENIX Security Symposium (USENIX Security 20), 2020, pp. 1273–1290.

[33] The MathWorks, Inc., System Identification Toolbox Documentation, 2025, (https://www.mathworks.com/help/ident/index.html). Accessed 14 March 2025.

[34] N. Pogaku, M. Prodanovic, T.C. Green, Modeling, analysis and testing of autonomous operation of an inverter-based microgrid, IEEE Trans. Power Electron. 22 (2) (2007) 613–625. https://doi.org/10.1109/TPEL.2006.890003

[35] A.K. Singh, R.R. Singh, An overview of factors influencing solar power efficiency and strategies for enhancing, in: 2021 Innovations in Power and Advanced Computing Technologies (I-PACT), 2021, pp. 1–6. https://doi.org/10.1109/i-PACT52855.2021.9696845

[36] A. Pradhan, B. Panda, Analysis of ten external factors affecting the performance of PV system, in: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, pp. 3093–3098. https://doi.org/10.1109/ICECDS.2017.8390025

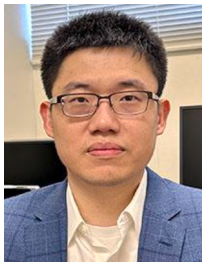
Imasha Balahewa is a joint doctoral student in the Department of Electrical and Computer Engineering at San Diego State University (SDSU) and the University of California San Diego (UCSD). She received her Bachelor of Science of Engineering Honours degree in Electrical Engineering from the University of Moratuwa, Sri Lanka. Her research focuses on cyber-physical security in inverter-based resource (IBR) dominated power systems, for enabling the secure and stable operation of renewable-rich grids under both cyber threats and physical disturbances.

Lars Bjorndal is a joint doctoral student in the Department of Electrical and Computer Engineering at San Diego State University and the University of California San Diego. He received the BS degree from the University of Florida. His research interests include cybersecurity for power-electronics-based power systems, with emphasis on attack detection and anomaly detection using data-driven and model-based methods.

Chris Mi (Fellow, IEEE and Fellow, SAE) received the BSEE and MSEE degrees in electrical engineering from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, Ontario, Canada, in 1985, 1988, and 2001, respectively.
He is a Distinguished Professor in the Department of Electrical and Computer Engineering and the Director of Caili and

Daniel Chang Center of Electrical Drive Transportation, San Diego State University (SDSU), San Diego, USA. Prior to joining SDSU, he was with the University of Michigan, Dearborn, from 2001 to 2015. His research interests include electric drives, power electronics, electric machines, electrical and hybrid vehicles, wireless power transfer, and power electronics.

In 2019, he received the Inaugural IEEE Power Electronics Emerging Technology Award. In 2022, he received the Albert W. Johnson Research Lectureship and was named the Distinguished Professor, the highest honor given to an SDSU faculty member, and only one award is given each year. In 2023, he received the IEEE PELS Vehicle and Transportation Systems Achievement Award, the IEEE Transactions on Industry Applications Best Paper Award, and the SDSU Innovator of the Year Award. In 2024, he received the prestigious Alumni Distinguished Faculty Award from SDSU. Most recently, he received the Wang Family Excellence Award from the California State University System.



Tong Huang is an Assistant Professor in the Department of ECE at San Diego State University. Before joining SDSU, he was a postdoctoral associate in the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT). He received his PhD degree from Texas A&M University. His industry experience includes R&D roles with ISO-New England and Mitsubishi Electric Research Laboratories. As the first author, he received two Best Paper Awards at the 2020 IEEE PES General Meeting and the 54-th Hawaii International Conference on System Sciences. His research focuses on cyber-physical resilience enhancement of power electronics-dominated electricity infrastructure via both data-driven and model-based approaches.