

Cyber-Resilient Frequency Control of VSGs Using Adaptive Learning Control

Mohammad Raeispour*, Shuo Yan*, Lasantha Meegahapola*

*Department of Electrical & Electronic Engineering, RMIT University, Melbourne, Australia

Abstract—This paper proposes a cyber-resilient secondary frequency control strategy for grid-forming inverters (GFIMs) operating under the virtual synchronous generator (VSG) control architecture. The reliance of VSG secondary control on inter-unit communication leaves it susceptible to cyber threats such as replay attacks that can create vulnerabilities and degrade dynamic behavior. Building on the power-balance swing dynamics with a virtual-governor model, we represent the frequency channel as a nominal nonlinear system, wherein link-level replay injects stale, lagged neighbor states that bias the secondary consensus input. A local predictor-based detector enforces time-order consistency in the received timestamps to detect and isolate compromised edges. Upon detection, an online deep neural network (DNN) estimator, driven by local residuals and the frequency error, learns and mitigates the unknown replay attack in real time. The resulting resilient control input restores frequency while preserving distributed active-power sharing. Simulation case studies demonstrate that the scheme maintains frequency restoration and sharing accuracy under replay.

Index Terms—Deep neural networks, frequency restoration, replay attack, virtual synchronous generator.

I. INTRODUCTION

Power systems with high shares of inverter-based resources (IBRs) increasingly rely on grid-forming (GFM) converters to maintain stability during disturbances. Virtual synchronous generator (VSG) control has become a practical way to equip converters with inertia- and damping-like responses while preserving fast control bandwidths and black-start capability [1]. While VSGs provide inertia-like response, they do not by themselves remove steady-state frequency offsets. Secondary frequency control is therefore layered above the primary loop to restore the nominal setpoint and coordinate active-power sharing among multiple units. This layer implementations have moved toward distributed designs, which eliminate single points of failure, scale naturally, and rely on simple neighbor messaging [2].

While inter-agent messaging improves efficiency, it simultaneously introduces an integrity risk, one that is not sufficiently addressed by schemes that prioritize performance. In practical networks, a capable adversary can replay previously valid neighbor states on selected links, injecting stale but authentic-looking information that biases the secondary consensus without triggering value-only checks. Unlike denial-of-service (DoS) attacks, replay does not involve disrupting communication or breaking encryption. Unlike sensor noise, replay attack is time-structured and targets the decision channel that coordinates frequency restoration and power sharing. This results in the slow down frequency recovery, deteriorated power sharing, and limited visibility of operators into the compromised edges [3], [4].

Recent work on secondary control for grid-forming inverters (GFIMs) has predominantly focused on achieving stable performance and efficient communication in normal operating conditions [5], [6]. While distributed secondary control schemes have gained attention for their ability to maintain frequency restoration and active-power sharing without relying on centralized supervision, the field remains relatively underdeveloped, especially in the context of VSGs [7]. This underdevelopment is compounded by growing concerns about adversarial integrity in distributed coordination, which motivates a closer look at attack-resilient control methods.

The attack-resilient control literature can be grouped into model-based, data-driven, and graph-theoretic approaches. Model-based methods rely on system models to detect and mitigate attacks, often providing strong theoretical guarantees such as stability and robustness. However, they typically require accurate system models and can struggle with nonlinearities or unmodeled attack behaviors, such as delayed inputs. Data-driven methods, particularly those leveraging machine learning, can adapt to unknown disturbances and handle complex, nonlinear system dynamics without relying on explicit models. However, these methods often lack theoretical stability guarantees and may require extensive training data. Graph-theory-based methods focus on the network structure, using metrics such as connectivity and topology reconfiguration to address compromised communication links. These methods are effective in managing distributed control systems but can lead to communication delays or loss of connectivity, especially when fine-grained edge isolation is needed [8].

Within this taxonomy, integrity threats do not affect the loop uniformly. Replay keeps links available and values plausible, yet timestamps are wrong, a realistic field condition that value-only filters often miss and that motivates timing-aware, edge-level control [9]. Despite recent progress, important gaps remain. Secondary control for VSGs is still underdeveloped from a resilience viewpoint, with most efforts centered on primary control or benign-operation performance. Replay attacks, though underrepresented in the literature, are likely in practice and particularly damaging for secondary consensus because they bias the coordination input through time-shifted neighbor states. Existing defenses emphasize DoS and FDIA, tend to act at the node level or via coarse graph reconfiguration, and rarely provide edge-level isolation or integrate online estimation to cancel replay-induced disturbances.

To close this gap, we treat replay as edge tampering and intervene exactly where stale neighbor states enter the secondary loop. We model the VSG secondary layer as a distributed system in which a compromised link appears as a

II. SYSTEM MODELING AND CONTROL OBJECTIVES

$$P_i^m - P_i^e = J_i \omega_{ni} \frac{d(\omega_i - \omega_{ni})}{dt} + D_i(\omega_i - \omega_{ni}) \quad (1)$$

Fig. 1. VSG primary control structure.

adopted, which regulates power injection based on frequency

$$\omega_i - \omega_{ni} = k_i^c P_i^c \quad (2)$$

Since primary control cannot fully eliminate steady-state frequency deviations, a secondary control layer is introduced to restore system frequency and enforce accurate active power sharing. This work designs the secondary control strategy based on a virtual governor framework, integrating both frequency regulation and power coordination.

The secondary control enforces (i) frequency restoration, $\lim_{t \rightarrow \infty} \omega_i = \omega_{\text{ref}} \forall i \in \mathcal{V}$, and (ii) proportional active-power sharing, $\lim_{t \rightarrow \infty} \frac{P_i^{\text{ref}} - P_i^e}{D_i} = \frac{P_j^{\text{ref}} - P_j^e}{D_j} \forall i, j \in \mathcal{V}$.

$$\dot{\omega}_{ni} = \dot{\omega}_i - k_i^c \dot{P}_i^c \quad (3)$$
$$\dot{\omega}_i = \frac{1}{J_i \omega_{ni}} (P_i^m - P_i^e - D_i(\omega_i - \omega_{ni})) + \dot{\omega}_{ni}. \quad (4)$$
$$g_i := \frac{1}{J_i \omega_{n_i}} (P_i^m - P_i^e - D_i(\omega_i - \omega_{n_i})),$$
$$\dot{\omega}_i = g_i + u_i^\omega \quad (5)$$
$$k_i^c \dot{P}_i^c = u_i^p \quad (6)$$
$$\omega_{ni} = \int (g_i + u_i^\omega - u_i^p) dt \quad (7)$$

A. Edge tampering via looped segment replay

Secondary-layer consensus relies on frequent and timely exchange of neighbors’ states across communication links. Many practical deployments use legacy or low-cost channels that lack strong authentication or freshness guarantees, which creates an opportunity for an adversary to intercept and later retransmit previously captured measurements. We refer to such link-targeted interventions as *edge tampering*. Here we consider a replay variant in which a finite recorded segment of a neighbor’s signal is forwarded to the target node in a loop. Replay attacks are straightforward to mount and can be difficult to detect with purely local checks.

To analyze this effect, consider the nominal secondary distributed consensus signal for frequency control of VSGs

$$u_{i,nom}^\omega = c_\omega \left(\sum_{j \in \mathcal{N}_i} a_{ij} (\omega_j - \omega_i) + (\omega_{\text{ref}} - \omega_i) \right) - g_i, \quad (8)$$

with dynamics $\dot{\omega}_i = g_i + u_i^\omega$. We omit the time argument when clear from context and write ω_i for $\omega_i(t)$.

Consider a directed edge $(j \rightarrow i)$, a record window $h_{ij} = [T_{r,ij}, T_{r,ij} + L_{ij})$ with length $L_{ij} > 0$, and an activation time $T_{a,ij} \geq T_{r,ij} + L_{ij}$. After activation, the attacker replaces fresh data with a looped replay of the recorded segment

$$\bar{\omega}_{j \rightarrow i}(t) = \begin{cases} \omega_j(t), & t < T_{a,ij}, \\ \omega_j(T_{r,ij} + \text{wrap}_{L_{ij}}(t - T_{a,ij})), & t \geq T_{a,ij}. \end{cases}$$

where $\text{wrap}_L(x) \triangleq x - L \lfloor x/L \rfloor \in [0, L)$. So, the received signal can be written as $\bar{\omega}_{j \rightarrow i} = \omega_j(t - \tau_{ij}(t))$ with effective delay values $\tau_{ij}(t) = t - T_{r,ij} - \text{wrap}_{L_{ij}}(t - T_{a,ij})$.

Within each replay cycle the delay is constant and it jumps by L_{ij} at loop resets. If attacked data are forwarded without changing link trust, the computed secondary input at node i is

$$\bar{u}_i^\omega = c_\omega \left(\sum_{j \in \mathcal{N}_i} a_{ij} (\bar{\omega}_{j \rightarrow i} - \omega_i) + (\omega_{\text{ref}} - \omega_i) \right) - g_i. \quad (9)$$

Subtracting the nominal input yields the exact replay-induced control error

$$f_i^{\text{replay}} = c_\omega \sum_{j \in \mathcal{N}_i} a_{ij} (\omega_j(t - \tau_{ij}) - \omega_j), \quad (10)$$

consequently, the attacked continuous-time normalized dynamics at node i can be compactly written as:

$$\dot{\omega}_i = g_i + u_i^\omega + f_i^{\text{replay}} \quad (11)$$

III. DETECTION AND DNN-BASED MITIGATION

A local detection mechanism and an online learning-based estimator structured as a DNN are developed to mitigate replay-induced disturbances. The detection model assumes the effective delays τ_{ij} are uniformly bounded, and under the looped replay model τ_{ij} is piecewise constant within each cycle and jumps by L_{ij} at loop resets. Replay episodes are intermittent, so persistent isolation is feasible. When a replay window begins, the link is initially treated as trusted ($\bar{a}_{ij} = a_{ij}$) until detection triggers. Where useful, the small- τ linearization $\omega_j - \omega_j(t - \tau_{ij}) \approx \tau_{ij} \dot{\omega}_j$ may be used only when τ_{ij} is small relative to the dominant electromechanical time constants. The estimator operates directly on reconstructed disturbance samples and does not rely on small-delay approximations.

A. Predictor-based event detection

Each node i maintains, for every incoming link $(j \rightarrow i)$, a one-step predictor, a residual test and two simple counters for misses and flags. The most recent received sample and its reception time are stored as $\omega_{j \rightarrow i}^{\text{last}}$ and t_{ij}^{last} . The predictor is initialized at t_{ij}^{last} with $\hat{\omega}_{j \rightarrow i}^{\text{pred}} = \omega_{j \rightarrow i}^{\text{last}}$ and propagated for $t \geq t_{ij}^{\text{last}}$ according to

$$\frac{d}{dt} \hat{\omega}_{j \rightarrow i}^{\text{pred}} = g_j + \hat{u}_{j,nom}^\omega, \quad (12)$$

where $\hat{u}_{j,nom}^\omega$ is the nominal control terms derived from node j 's local model. For each received measurement $\omega_{j \rightarrow i}^{\text{rec}}$, with timestamp t_{tag} , form the raw residual

$$r_{ij} = \omega_{j \rightarrow i}^{\text{rec}} - \hat{\omega}_{j \rightarrow i}^{\text{pred}}. \quad (13)$$

To reduce false alarms from measurement noise the raw residual is passed through a simple exponential filter,

$$\bar{r}_{ij} = \lambda_r \bar{r}_{ij}^{\text{pred}} + (1 - \lambda_r) r_{ij}, \quad 0 < \lambda_r < 1, \quad (14)$$

where $\bar{r}_{ij}^{\text{pred}}$ is the predicted filtered residual at reception, from propagating $\bar{r}_{ij}$ between receptions. If a measurement carries a timestamp number $t_{\text{tag}} \leq t_{ij}^{\text{last}}$ it is immediately counted as suspicious. Otherwise the filtered residual is compared to a threshold η_{ij} and the flag counter c_{ij}^{flag} is incremented when $|\bar{r}_{ij}| > \eta_{ij}$ (reset to zero otherwise). When no data arrives within the expected interval Δt_{exp} (the nominal inter-arrival time), the miss counter c_{ij}^{miss} is incremented and the predictor continues propagation (last-value hold if $g_j, \hat{u}_{j,nom}^\omega$ are unavailable). Isolation is triggered when either counter exceeds its design threshold:

$$\text{if } c_{ij}^{\text{miss}} \geq N_{\text{miss}} \text{ or } c_{ij}^{\text{flag}} \geq N_{\text{flag}} \Rightarrow \bar{a}_{ij} \leftarrow 0. \quad (15)$$

A previously isolated link is reinstated only after M_{recov} consecutive healthy receptions (no flags) and strictly increasing timestamps. M_{recov} is a design parameter that determines how many consecutive valid receptions (without flags or timestamp violations) are required before a previously isolated link can be trusted again. The threshold η_{ij} and filter parameter λ_r are tuned from nominal predictor-error statistics to balance sensitivity and false alarms.

B. DNN-based Replay Attack Estimation and Mitigation

To counteract the additive disturbance induced by replay attacks, each node i is equipped with an adaptive DNN estimator integrated within the secondary control loop. Leveraging universal approximation function, the replay effect is modeled as

$$f_i^{\text{replay}} = W_i^\top \sigma(\Psi_i(\phi_i)) + \varepsilon_i, \quad (16)$$

where $\phi_i \in \mathbb{R}^p$ denotes the local feature vector, $\Psi_i : \mathbb{R}^p \rightarrow \mathbb{R}^{L_m}$ is a nonlinear feature extractor implemented by a multilayer feedforward network, $\sigma(\cdot)$ is a bounded activation function, $W_i \in \mathbb{R}^{L_m}$ is the output-layer weight matrix, and ε_i is the bounded approximation error. The inner mapping is

$$\Psi_i(\phi_i) = \underbrace{\psi_{m,i}(V_{m,i}^\top \psi_{m-1,i}(V_{m-1,i}^\top \cdots \psi_{1,i}(V_{1,i}^\top \phi_i) \cdots))}_{m \text{ hidden layers}},$$

where the layer widths are $\{L_k\}_{k=0}^m$ with $L_0 = p$ and L_m the last hidden width, $V_{k,i} \in \mathbb{R}^{L_{k-1} \times L_k}$ for $k = 1, \dots, m$, and each $\psi_{k,i} : \mathbb{R}^{L_k} \rightarrow \mathbb{R}^{L_k}$ acts elementwise. We adopt the following designer-specified bounds for weights and activations:

$$\|W_i\|_2 \leq \bar{W}, \quad \|V_{k,i}\|_F \leq \bar{V}, \quad \|\sigma(z)\|_\infty \leq \bar{\sigma}$$

with $\|\cdot\|_2$ (vectors) and $\|\cdot\|_F$ (matrices). The local feature vector is selected to capture both replay-related residuals and local operating conditions:

$$\phi_i = [\{\bar{r}_{ij}\}_{j \in \mathcal{N}_i}, \omega_i, u_{i,nom}^\omega]^\top \in \mathbb{R}^p,$$

where $u_{i,\text{nom}}^\omega$ is the nominal consensus input, and $p = |\mathcal{N}_i| + 2$. These features provide direct indicators of link anomalies and their impact on the local frequency channel.

The mitigation requires a real-time estimate of the replay disturbance from locally available features. The DNN provides this estimate as

$$\hat{f}_i = \hat{W}_i^\top \sigma(\hat{\Psi}_i(\phi_i)), \quad (17)$$

with $\hat{W}_i$ and $\hat{\Psi}_i$ the online output and inner-layer parameters, respectively. The DNN's objective is to minimize the replay disturbance by continuously adapting its weights using the local frequency error as a training signal. To enable the DNN to learn the unknown disturbance in real time, both the outer- and inner-layer weights are continuously adapted based on the local frequency error $e_i = \omega_i - \omega_{\text{ref}}$. The adaptation follows a gradient-descent-like update that drives the estimated disturbance $\hat{f}_i$ to counteract the replay effect observed through e_i . Motivated by results in [10], the output-layer weights evolve as

$$\dot{\hat{W}}_i = \Gamma_{W_i} \sigma(\hat{\Psi}_i(\phi_i)) e_i, \quad (18)$$

where $\Gamma_{W_i} > 0$ is the learning-rate matrix governing the speed of adaptation. The inner-layer weights are updated in real-time using the feature vector ϕ_i and the frequency error e_i , ensuring that the DNN accurately tracks the disturbances over time. The inner-layer parameters are tuned through a continuous adaptation law of the form

$$\dot{\hat{\nu}}_{k,i} = \nu_{k,i}(\phi_i, e_i), \quad \|\nu_{k,i}\|_F \leq \rho(\|e_i\|) \|e_i\|, \quad (19)$$

where $\rho(\cdot)$ is a user-defined nondecreasing bounding function that constrains the adaptation rate and ensures smooth parameter evolution. Conditioned on detector activation, the nominal input is corrected to reject replay effects and restore frequency by subtracting the DNN estimate and adding proportional damping:

$$u_i^\omega = u_{i,\text{nom}}^\omega - \hat{f}_i - k_d e_i, \quad (20)$$

with $k_d > 0$.

TABLE I: MG parameters

Description	GFM 1&2	GFM 3&4
Active power droop gain (k_i^c)	20.3×10^{-2}	25.4×10^{-2}
Virtual damping coefficient (D_i)	52.6×10^3	52.6×10^3
Virtual moment inertia (J_i)	1.5	1.5
Voltage loop (K_{PV}, K_{IV})	(0.05, 390)	(0.1, 410)
Current loop (K_{PC}, K_{IC})	(10.5, 16000)	(12, 18000)
Loads ($P_{1,2}/P_{3,4}$)	45×10^3	35×10^3

IV. SIMULATION RESULTS

To assess the effectiveness of the proposed method, simulations are performed on the IEEE 33-bus feeder using MATLAB/Simulink. The system comprises four VSG units positioned at various nodes within the microgrid, each supplying power to a specific load. Primary and secondary control layers are employed to manage frequency regulation and active power sharing. The GFMI communication network is represented as an undirected ring graph, facilitating distributed consensus-based secondary control. Key system parameters

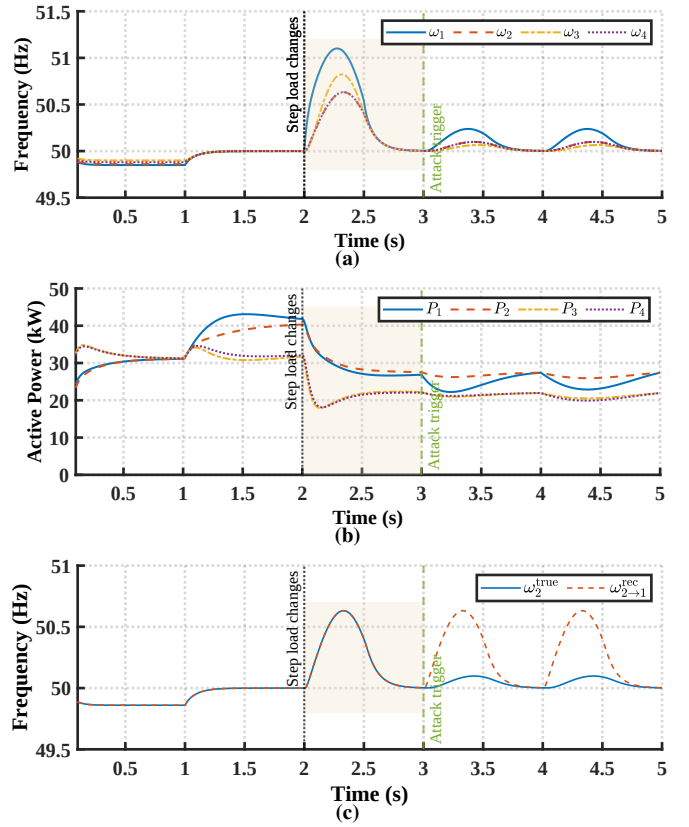


Fig. 2. Performance of conventional distributed method under replay attack on in GFMI2: (a) frequency, (b) active power (c) GFMI2 frequency during attack.

are provided in Table I. The study includes two cases: first, examining the performance of the secondary VSG in providing effective frequency control and fair active power sharing, and its limitations in handling replay attacks; second, evaluating the performance of the proposed method.

In the first case, the system initially operates under primary control with a quasi-steady-state frequency error. At $t = 0.5s$, the secondary control layer is activated to restore the nominal frequency. After a short activation period, the system frequency reaches the nominal value of 50 Hz, and active power sharing is achieved based on the droop coefficient (see Fig. 2). A step load change at $t = 2s$ decrease Load 1 and 3 by 50%, causing an decrease in active power and frequency overshoot, where the effect is quickly damped by the secondary control. At $t = 3s$, a replay attack is introduced, replaying frequency data from VSG 2 to VSG 1. This creates disturbances, leading to additional oscillations in the VSG units' frequencies. Although no step load changes occur, the frequency signal displays oscillations resembling those from a load change, which could deceive the system operator or trigger unintended actions. Without detection or mitigation, the system's performance deteriorates, highlighting its vulnerability to replay attacks.

In the second case, the system again undergoes a step load change at $t = 2s$, followed by a replay attack at $t = 3s$. However, this time, the system incorporates the proposed detection and mitigation method. In this way, the

DNN estimator is configured as a 4-layer tanh MLP with layer sizes of $[8, 9, 11, 9]$, an outer-layer learning gain of $\Gamma_{W_i} = 5 \times 10^3$, $\mathbf{1}_{L \times L}$, inner-layer gains $\Gamma_{V_{j,i}} = 50$, inner-layer weight bounds are set between $[10^{-4}, 10^2]$, with randomized initial weights $\hat{W}_i(0), \hat{V}_{j,i}(0) \sim \mathcal{U}[-10^{-3}, 10^{-3}]$, and secondary control gains are $k_d = 5$ and $c_\omega = 10$. The attack detection and isolation rely on thresholds $N_{miss} = 5$ and $N_{flag} = 5$. The detection mechanism, based on a local predictor, identifies the replay attack by comparing the incoming data to predicted values. Once the attack is detected, the compromised communication link is isolated. After isolation, the DNN estimator is activated to mitigate the replay-induced disturbances. The DNN dynamically adjusts in real-time, learning from residuals and frequency errors to counteract the replay effect. This method successfully restores the system frequency to its nominal value and ensures accurate active power sharing, despite the ongoing replay attack. The results in Fig. 3 demonstrate the effectiveness of the proposed method in isolating the compromised link and mitigating the replay attack, maintaining stable frequency and power sharing.

V. CONCLUSION

A detection-and-mitigation control scheme for VSG-based secondary frequency control is developed to effectively counter edge-level replay tampering. Mathematically, replay enters the closed loop as an additive input disturbance. Our one-step predictor and filtered residual provide a scalar decision signal whose thresholding enables link isolation with minimal modeling burden. On top of this, an adaptive DNN uses locally available features, the error, and a constrained continuous-time update to approximate and compensate the attack online, yielding a resilient law that preserves the nominal consensus objectives: frequency restoration and proportional active-power sharing. The algorithm requires only neighbor messages, local clocks for monotonic-timestamp checks, and lightweight counters, making it compatible with legacy communication links. Simulations confirm that the controller maintains stability and sharing performance under bounded, intermittent replay windows.

REFERENCES

- [1] R. Rosso, X. Wang, M. Liserre, X. Lu, and S. Engelken, "Grid-forming converters: Control approaches, grid-synchronization, and future trends—A review," *IEEE Open J. Ind. Appl.*, vol. 2, pp. 93–109, 2021.
- [2] M. Raeispour, H. Atrianfar, H. R. Baghaee, and G. B. Gharehpetian, "Resilient H_∞ consensus-based control of autonomous AC microgrids with uncertain time-delayed communications," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 3871–3884, 2020.
- [3] A. J. Gallo, M. S. Turan, F. Boem, T. Parisini, and G. Ferrari-Trecate, "A distributed cyber-attack detection scheme with application to DC microgrids," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3800–3815, 2020.
- [4] M. Raeispour, S. Yan, L. Meegahapola, and X. Yu, "Robust distributed learning framework for FDIA estimation and mitigation in grid-forming converters," *IEEE Trans. Ind. Electron.*, 2025.
- [5] K. Feng and C. Liu, "Distributed hierarchical control for fast frequency restoration in VSG-controlled islanded microgrids," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 496–506, 2022.
- [6] S. Yan, L. Meegahapola, Y. Yang, and F. Blaabjerg, "Grid-supporting renewable energy systems with power electronics interfaces," *IEEE Open J. Power Electron.*, 2025.
- [7] X. Huang, D. Qi, Y. Chen, Y. Yan, S. Yang, Y. Wang, and X. Wang, "Distributed self-triggered privacy-preserving secondary control of VSG-based AC microgrids," *IEEE Trans. Smart Grid*, 2024.

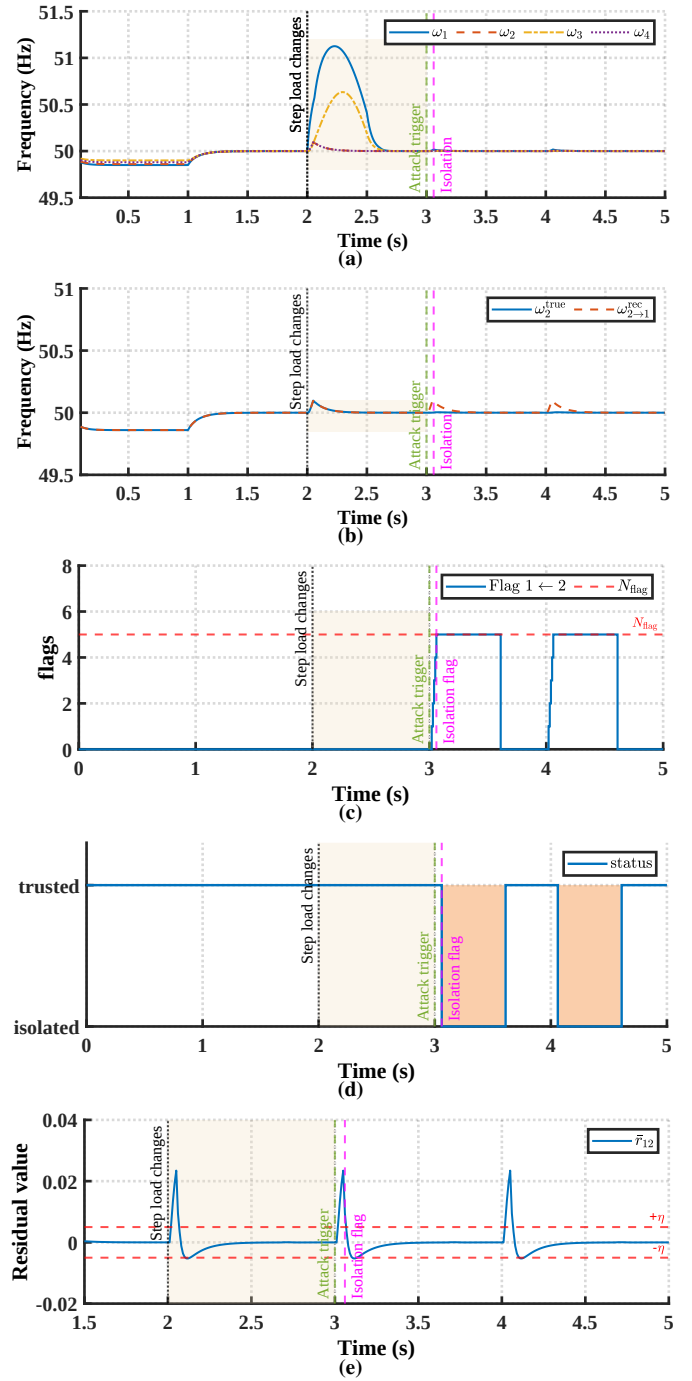


Fig. 3. Performance of proposed method under replay attack on in GFMI2: (a) frequency, (b) GFMI2 frequency during attack. (c) attacked link status, (d) detection flag, (e) residual error.

- [8] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakraborty, "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, 2019.
- [9] A. Naha, A. Teixeira, A. Ahlén, and S. Dey, "Sequential detection of replay attacks," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1941–1948, 2022.
- [10] M. Raeispour, S. Yan, and L. Meegahapola, "Resilient deep learning-based voltage control for grid-forming inverters under cyber intrusion," in *2025 IEEE International Conference on Energy Technologies for Future Grids (ETFG)*, Wollongong, NSW, Australia, 2025, pp. 1–6.