Detection and Mitigation of False Data Injection Attacks Against Wind Farm Active Power Controllers in Power Grids

Mostafa Ansari, Mohsen Ghafouri, Amir Ameli, *Senior Member*, *IEEE*, Ulas Karaagac, and Ilhan Kocar, *Senior Member*, *IEEE*

***Abstract*—The recent growing integration of wind farms (WFs), particularly variable speed wind turbines (WTs), results in several operational challenges to power grids integrated with WFs, such as low grid inertia and the reduced performance of measurement-based fast frequency response. To deal with such challenges, grid operators use WF active power controllers (WFAPCs) to enhance frequency control support from WTs and improve the frequency stability of the grid. However, the operation of WFAPC relies on measurements received through communication networks and cyber layers of WFs, which consequently makes them prone to cyber threats, e.g., false data injection (FDI). On this basis, firstly, this paper analyzes the cybersecurity vulnerabilities of WFAPCs and the possible impacts of exploiting cybersecurity vulnerabilities on the frequency response of WF and frequency stability of the grid. Then, based on the knowledge of intruders, two attacks, i.e., white-box and black-box FDI attacks, are developed against WFAPCs. Afterward, to detect these attacks, a novel bi-level detection and mitigation technique based on support vector machine (SVM)-based technique and long short-term memory (LSTM) -based technique is developed, which is implemented at the control center of the WF (primary detector) and at the dispatch center of the power grid (secondary detector), respectively. These detectors classify real-time measurements into attack and normal operation. Additionally, a hierarichical mitigation technique is proposed to counter the developed cyber attacks by replacing the active power reference signal of WF with new values obtained based on the droop control theory. The impacts of the attacks and the effectiveness of the proposed bi-level technique are evaluated using the modified 39-bus benchmark.**

***Index Terms*—Wind turbine (WT), wind farm (WF), cybersecurity, active power controller, attack detection, attack mitigation, frequency stability, false data injection (FDI), machine learning.**

Nomenclature

A. Subscripts

CIG	Converter interfaced generator
FL	Flexible load
GG	Governor
GS	Steam turbine generator
GT	Gas turbine generator
j	Index of wind farm (WF)
RH	Reheater
ST	Steam turbine
WT	Wind turbine

B. Variables and Parameters

α	Weight of frequency nadir ratio (FNR)
$\delta(t)$	Auxiliary time-variant function
ϵ_t	Normal distribution with zero mean
Λ	Forgetting factor
μ	Penalty factor
$\sigma, \tanh$	Sigmoid and hyperbolic tangent functions
τ_{tr}, τ_{ts}	Limits for training and testing time
$\tilde{\tau}$	Attack start time
$\Delta\omega$	Rotor speed variation
ω	Rotor speed
ω_m	Electromagnetic rotor speed
$\omega_0, \omega_{\min}$	Pre-event and post-support rotor speeds
ψ	Set of hyper-parameters
μ	Penalty factor
v_{tr}, v_{ts}	Auxiliary binary variables
$\boldsymbol{c}_t$	Cell state vector
D	Damping coefficient
d	Index of day
e	Amount of disturbance in power grid
$\hat{e}$	Estimated e
$\boldsymbol{err}$	Kalman filter error vector

Manuscript received: December 21, 2024; revised: April 29, 2025; accepted: June 11, 2025. Date of CrossCheck: June 11, 2025. Date of online publication: September 29, 2025.

This work was supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC) (No. RGPIN-2022-05048).

M. Ansari and M. Ghafouri (corresponding author) are with Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada (e-mail: mostafa.ansari@concordia.ca; mohsen.ghafouri@concordia.ca).

A. Ameli is with the Electrical Engineering Department, Lakehead University, Thunder Bay, Canada (e-mail: aameli@lakeheadu.ca).

U. Karaagac is with the Electrical-Electronics Engineering Department, Middle East Technical University, Ankara, Turkey (e-mail: ulaskar@metu.edu.tr).

I. Kocar is with the Electrical Engineering Department, Polytechnique Montreal, Montreal, Canada (e-mail: ilhan.kocar@polymtl.ca).

DOI: 10.35833/MPCE.2024.001368

Journal of Modern Power Systems and Clean Energy

f	Frequency
f_0	Initial frequency
$\boldsymbol{f}_t$	Activation vector of forget gate
$\tilde{f}$	Manipulated frequency
Δf	Frequency deviation
$\overline{\Delta f}$	Estimated Δf
$\widetilde{\Delta f}$	Measured Δf
Δf_{DB}	Frequency deadband
f_{DB}^{UF}, f_{DB}^{OF}	Under- and over-frequency deadbands
f_{LSh1}	Frequency threshold of first-step load-shedding
$f_{\max}$	Over-frequency limit
$f_{\min}$	Under-frequency limit
f_{nom}	Nominal frequency of grid
f_{thr}	Frequency threshold
Δf_{thr}	Frequency threshold deviation
f_{sample}	Frequency of sample
F_{HP}	Fraction of high-pressure power in steam turbine
$\boldsymbol{g}_t$	Cell input activation vector
$\boldsymbol{h}_t$	Hidden state vector
H	Grid inertia
H_w	Wind turbine inertia
$\boldsymbol{i}_t$	Activation vector of input gate
k	Index of time step
k_p, k_i	Proportional-integral (PI) controller parameters of automatic generation control (AGC) system
K_X	Participation factor of each frequency service provider X
K_δ	Constant factor
$LDWI$	Defined dependency index
n	Number of sampling time interval t_{sample}
$\boldsymbol{o}_t$	Activation vector of output gate
ΔP	Active power variation
P_e	Electromagnetic power
P_L	Load peak
P_{ref}	Active power setpoint of WF
P_W	Power output of WF
$\bar{P}_W$	Estimated value of P_W
P_W^{ref}	Reference of P_W
ΔP_W^{ref}	Deviation of P_W^{ref}
$\overline{\Delta P_W^{ref}}, \overline{\overline{\Delta P_W^{ref}}}$	Estimated values of ΔP_W^{ref} from primary level and secondary level
R_X	Active power droop of each frequency service provider X
$RoCoF$	Rate of change of frequency
$\widetilde{RoCoF}$	Manipulated $RoCoF$
$RoCoF_{\max}$	The maximum $RoCoF$
t_{delay}	The maximum communication delay in receiving data and sending commands
t_n	Time window length
t	Index of time
t_{sup}	Support time
t_{tr}, t_{ts}	Training time and testing time
T_X	Time constant of frequency response of each frequency service provider X
T_n	Moving window length of frequency measurements
T_e	Electromagnetic torque
T_{ML}	Time window of mechine learning classification
T_m	Mechanical torque
v_{dq}^*	Voltage reference of rotor-side converter (RSC)
$v_{r,j}, \boldsymbol{i}_{r,j}$	Measured voltage and current of rotor of doubly-fed induction generator (DFIG)
$\boldsymbol{x}_t$	Input vector to long short-term memory (LSTM) unit
$\boldsymbol{x}$	State vector of power grid
z	Z-transform variable

C. Matrices

$\boldsymbol{A}$	State matrix of power grid
$\boldsymbol{A}_r$	State matrix of reduced-order power grid
$\hat{\boldsymbol{A}}$	Estimated state matrix
$\boldsymbol{B}$	Input matrix of power grid
$\boldsymbol{B}_r$	Input matrix of reduced-order power grid
$\hat{\boldsymbol{B}}$	Estimated input matrix
$\boldsymbol{C}$	Output matrix of power grid
$\boldsymbol{C}_r$	Output matrix of reduced-order power grid
$\hat{\boldsymbol{C}}$	Estimated output matrix
$\boldsymbol{K}$	Kalman gain vector
$\boldsymbol{L}$	Observer gain vector
$\boldsymbol{M}$	Error covariance matrix
$\boldsymbol{Q}_r, \boldsymbol{R}_r$	Bryson's cost function matrices
$\boldsymbol{R}$	Recurrent weight matrix
$\boldsymbol{S}$	Bryson's solution matrix

I. Introduction

OVER the last decade, the total installed wind power capacity has grown annually by 17.5% in Canada and 21.1% globally [1]. Additionally, grid reliance on wind energy is becoming a new trend, e.g., on April 27, 2020, wind energy supplied 72% of the electricity demand in the Southwest Power Pool, Little Rock, USA, while only about 22% of its generation capacity is wind energy [2]. Therefore, even a relatively small share of wind energy can play a critical role at key moments in power grids.

Among various wind energy generation technologies, doubly-fed induction generator (DFIG) has been widely used in power grids due to its low cost and capability to harvest the maximum energy at different wind speeds [3]. However, the interfacing converters between the generator and the grid decouple DFIGs from the rest of the system, reducing the inertia of the entire grid. For example, the inertia of the European power grid in 2016 has been reduced by around 20% compared with that in 1996, primarily due to the integration of grid-decoupled renewable energy sources (RESs), most of which are DFIGs [4]. The inertia reduction makes the frequency stability of the grid more vulnerable to frequency-related events, e.g., load changes and generator trips. To tackle this problem, grid codes in power grids with considerable wind energy share, e.g., National Grid Code in UK [5], have enforced large-scale wind farms (WFs) to participate in load frequency control (LFC) schemes [3]. Although the integration of DFIGs leads to a reduction in the grid inertia, they possess the capability to offer fast frequency support. To provide such services by WFs, the techniques proposed in [6]-[13] can be classified into two major groups: installing auxiliary devices in the grid close to WFs [6]-[8] and modifying WF control schemes [9]-[13]. Using the second group of techniques, WFs can operate in primary frequency support mode (PFSM) to arrest the frequency changes after a disturbance. The techniques used in PFSM rely on supervisory control and data acquisition (SCADA) systems and communication links to take control actions and transfer data/command [14].

In a WF, communication protocols used to transfer data/command, e.g., IEC 61400-25 [15], are designed for fast data exchanges and do not include security features, such as encryption, as highlighted by the U.S. Department of Energy [16]. Therefore, the important communications are prone to various forms of cyber attacks. Such attacks should be differentiated from threats against conventional generators and handled separately since WFs ① are spread across remote areas with limited physical security and numerous remote data transfers, resulting in an extensive attack surface; ② have unique cyber layers and protocols designed without security consideration; and ③ exhibit fast and unique transient behaviors and stability issues that can quickly affect grid operations.

Over the past decade, adversaries have exploited existing vulnerabilities in WFs, leading to several events with considerable wind turbine (WT) outages. For instance, in March 2019, a denial-of-service (DoS) attack is launched against the communication between the control center and WFs in Utah, U.S., which results in unexpected reboots of the devices after exploiting the vendor firewall [16]. Another attack in Germany in February 2022 results in an outage of 5800 WTs [17]. Based on the above discussions, the security analysis of the WFs is of paramount importance, particularly when WTs are used for sensitive applications such as frequency control.

Despite the importance, only a limited number of studies focus on security analysis of the WFs. In [18], the cyber-physical model of WF is analyzed, and various scenarios are studied in which the attacker sends false shutdown commands to the WTs. In [19], manipulation of WT setpoints is considered as the attacker target, and the consequences of such attacks on the WTs operation are investigated. In [20], the sub-synchronous damping controller for series-compensated DFIG is targeted by attackers in various scenarios. Attacks that result in the disconnection of turbines are detected in [21] using time series data of WF power generation in the long term. Moreover, mitigation techniques for delay and DOS cyber attacks against WTs are studied in [22] and [23]. Although the cybersecurity of LFC is extensively studied in [24], only limited studies focus on the potential impacts of cyber attacks targeting WFs in the LFC scheme. For instance, a cyber attack against measured frequency is discussed in [25], which aims to trigger load-shedding schemes. However, the diagnosis and mitigation framework proposed in [25] overlooks critical grid operation aspects, such as LFC response, frequency limits (e.g., deadband and rate of change of frequency (RoCoF)), and time limitations of the PFSM. Consequently, the anomaly-based intrusion detection and diagnosis system relies on unrealistic model residuals, potentially leading to inaccurate decision-making. In addition, cyber vulnerabilities of WF active power controller (WFAPC) in PFSM and their impacts on the frequency stability of grid have not been studied in the literatures yet.

Based on the above discussion, in this paper, the cybersecurity vulnerabilities of WFAPC in PFSM are analyzed and effective solutions are proposed to enhance the security of WFs against the developed attacks. First, using a cyber-physical model of DFIG-based WFs, the entry points of the attacker are identified and two types of attacks, i.e., black-box and white-box false data injection (FDI) attacks, are developed based on the attacker's knowledge. In the white-box FDI attack, the attackers have sufficient knowledge about the grid parameters, whereas, in the black-box FDI attack, their information is limited to the historical data of the grid. Then, a bi-level detection and mitigation technique is proposed to maintain the grid security, offering superior performance compared with existing techniques. At the control center of the WF (primary detector), a support vector machine (SVM)-based technique is employed with the help of a well-tailored observer to detect and mitigate adverse measurement manipulations. At the dispatch center of the power grid (secondary detector), an long short-term memory (LSTM)-based technique is developed, employing frequency measurements, WF generation, and their respective rates of change to identify the attacks. Finally, to mitigate the attacks, the primary detector replaces the manipulated signal with an estimated one, whereas when the primary detector is not able to mitigate the attack, the secondary detector may rebuild the active power reference signal of WF based on frequency deviation and droop control theory. The contributions of this paper can be summarized as follows.

1) This paper identifies novel cybersecurity vulnerabilities of WFAPC, specifically focusing on their impact on frequency stability of power grid, and develops two FDI attacks tailored to varying levels of the attacker's knowledge.

2) A bi-level detection and mitigation technique is proposed to effectively identify the developed FDI attacks, em-

ploying an SVM-based technique at the control center of the WF as the primary detector and an LSTM-based technique at the dispatch center of the power grid as the secondary detector.

3) A hierarchical mitigation technique is presented, utilizing a tailored observer at the control center of the WF for initial response and a novel backup technique at the dispatch center of the power grid that bypasses compromised WF controls by directly replacing the active power reference signal.

II. LFC and Cyber-Physical Model of WFs

A. Participation of WFs in LFC

Three distinct periods during an under-frequency (UF) event in the grid are classified as: ① the inertial frequency response of all rotating masses; ② the primary frequency response of generators and load damping; and ③ automatic generation control (AGC) operation (secondary and tertiary frequency responses) [26] - [33]. WFs can provide all these frequency supports if they are empowered and enabled by WFAPC. Figure 1 presents a taxonomy of wind-based frequency support techniques, categorized by their underlying mechanisms and control strategies. While Fig. 1 illustrates the broader landscape of frequency support, the primary focus of this paper is specifically on inertial and primary frequency responses. This emphasis stems from the practical advantages of implementing such techniques without relying on auxiliary devices, which ultimately translates into lower system costs. To clarify the connection, Fig. 1 highlights the branches of the taxonomy directly related to inertial and primary frequency responses. These branches demonstrate how WTs can contribute to frequency stability by emulating synchronous generator inertia and providing rapid power injections in response to frequency drops. These techniques are crucial for maintaining frequency stability, which is the central theme explored in this paper. Additionally, among these techniques, de-loading mechanisms may result in wind energy curtailment. Thus, similar to many studies, we use power unreserved control mechanism [9]-[11].

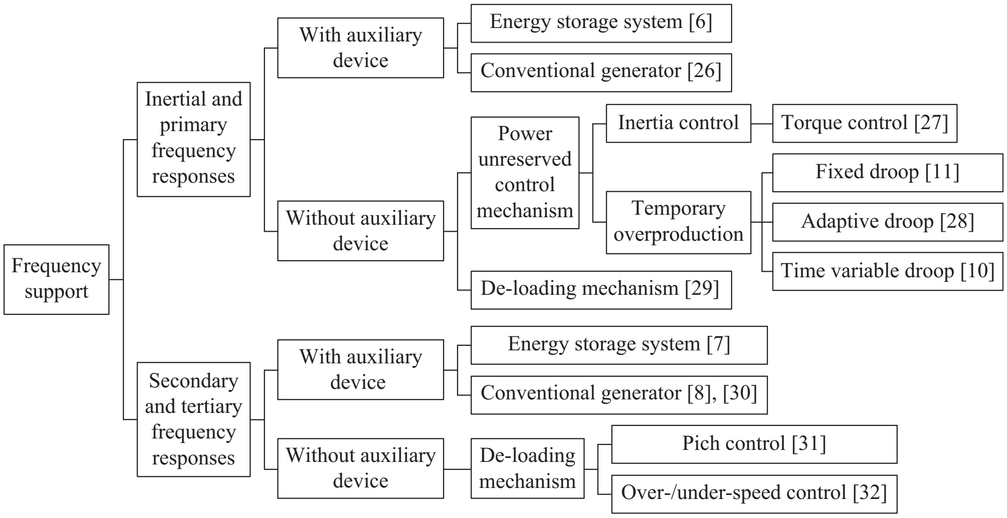

Frequency support
Inertial and primary frequency responses
With auxiliary device
Energy storage system [6]
Conventional generator [26]
Without auxiliary device
Power unreserved control mechanism
Inertia control
Torque control [27]
Temporary overproduction
Fixed droop [11]
Adaptive droop [28]
Time variable droop [10]
De-loading mechanism [29]
Secondary and tertiary frequency responses
With auxiliary device
Energy storage system [7]
Conventional generator [8], [30]
Without auxiliary device
De-loading mechanism
Pich control [31]
Over-/under-speed control [32]

Fig. 1. Taxonomy of wind-based frequency support techniques.

B. Communication System of SCADA System

The SCADA system of a WF is often composed of monitoring and control mechanisms and a communication system. Monitoring and control mechanisms are often deployed at three levels [3], [21]: bottom level (WT level), middle level (WF level), and top level (grid level). In addition, the communication system of the SCADA system can be divided into five sub-networks to transfer data/commands, as shown in Fig. 2, and is explained in Supplementary Material A [21]. In Fig. 2, LAN is short for local area network; WAN is short for wide-area network; SCU is short for substation control unit; GDA is short for grid data acquisition; PDI is short for process data interface; VCS is short for voltage control system; IED is short for intelligent electronic device; WTCP is short for WT control panel; Meteo is short for meteorological station; ICCP is short for inter-control center communications protocol; SL is short for substation level; PL is short for process level; BL is short for by level; LN is short for logical node; FC denotes function corresponding to an LN; DO is short for data object; and DA is short for data attribute.

C. Model of WF

1) Model of WT

The model of captured mechanical wind power P_m and the active power reference of the DFIG P_{MPPT} are adopted from [34]. Accordingly, the drive-train system (containing the shaft and the gearbox) is modeled as a single-mass block with inertia H_w, which can be expressed as:

$$\begin{cases} 2H_w \dfrac{\mathrm{d}\omega}{\mathrm{d}t} = T_m - T_e \\ T_m = \dfrac{P_m}{\omega} \\ T_e = \dfrac{P_e}{\omega} \end{cases} \tag{1}$$

Due to the fast dynamics of the back-to-back converter in DFIG, P_e can be considered to follow P_{MPPT} [10].

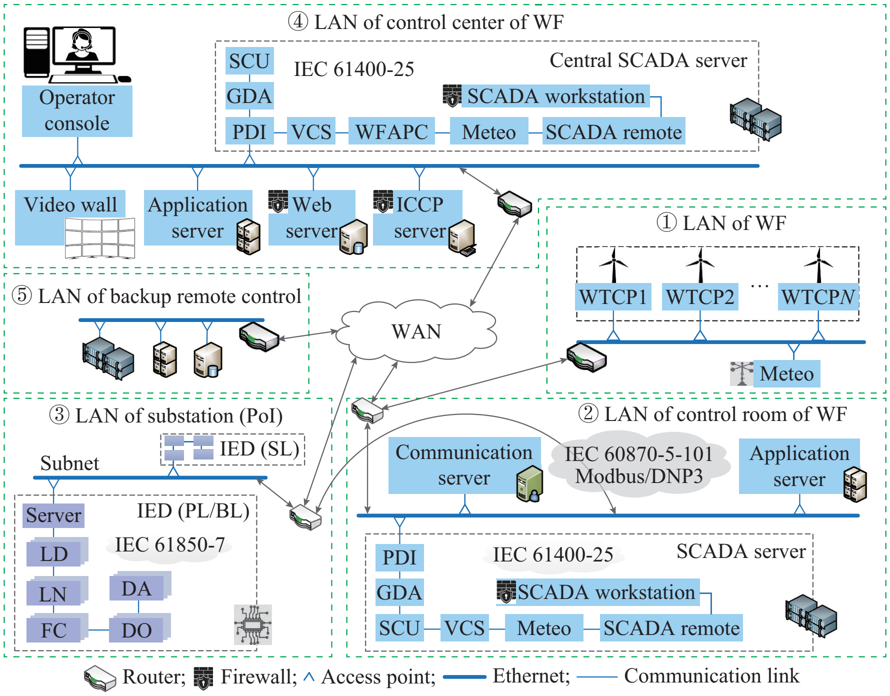


Fig. 2. Communication system of SCADA system.

2) Model of WFAPC

In this paper, a time-variable droop strategy [10], which is enhanced by an adaptive recovery time technique, is used to emulate frequency response and also alleviate secondary frequency dip.

The active power control scheme of WT, which is also the model of WFAPC, is shown in Fig. 3 and explained in Supplementary Material B, where MPPT is short for maximum power point tracking.

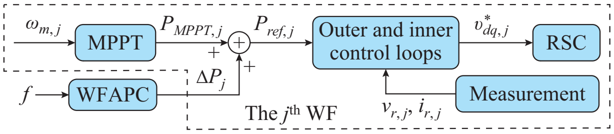


Fig. 3. Active power control scheme of WT.

D. Model of Power Grid Integrated with WFs

The frequency response model of the power grid is generally built neglecting nonlinearities and all relatively small time constants, which is a common assumption in the studies related to LFC [35]. An approximate model for the frequency stability study in a single-area power grid following an active power disturbance ΔP_{dist} is proposed in Fig. 4. In Fig. 4, the model of WT includes WFAPC and DFIG drive-train dynamics during and after the support time. The disturbance can be either a sudden loss of a load ($\Delta P_{dist}>0$) or a generating unit ($\Delta P_{dist}<0$). In this model, all the frequency service providers are divided into five categories, which are as follows.

1) WT: dynamic behaviors of WTs are characterized by the PFSM and the drive-train response with the time constant T_{WT}. Assuming the mechanical power is constant ($\Delta P_m=0$), impacts of WTs on system frequency during and after the support period are considered in the model proposed in [10] and [11].

2) Steam turbine: the fraction of steam turbine power generated by high-pressure section is represented by F_{HP} [35].

3) Gas turbine: the transfer function of combined cycle and open cycle gas turbines has two considerable time constants including T_{GT} and T_{GG} [35].

4) Flexible load (FL): as emerging frequency service providers, FL can be modeled by their time constant T_{FL}.

5) Converter interfaced generator (CIG): the transfer function of CIGs that provide virtual frequency response, except WTs, can be modeled considering the time constant T_{CIG} [36]. In Fig. 4, the AGC system is modeled as a proportional-integral (PI) controller (k_p and k_i), and the transfer function of power grid is characterized by the grid equivalent inertia H and the load damping coefficient D. The considered load-shedding scheme includes four steps depending on the frequency deviation [37]. The model in Fig. 4 can be described in state-space form as:

$$\begin{cases} \dot{\boldsymbol{x}}=\boldsymbol{Ax}+\boldsymbol{Bv} \\ \boldsymbol{y}=\boldsymbol{Cx}=[\Delta f] \\ \boldsymbol{v}=[\Delta P_{dist}] \end{cases} \tag{2}$$

III. Attack Modeling

Communication networks and cyber components, which belong to the bottom, middle, and top levels, can be targeted by different types of cyber attacks. However, in this paper, FDI attacks against WFAPC that exploit the top level of WF are investigated.

A. FDI Attack Against WFAPC

The attackers can manipulate measurement signals (e. g., grid frequency) by launching different types of FDI attacks.

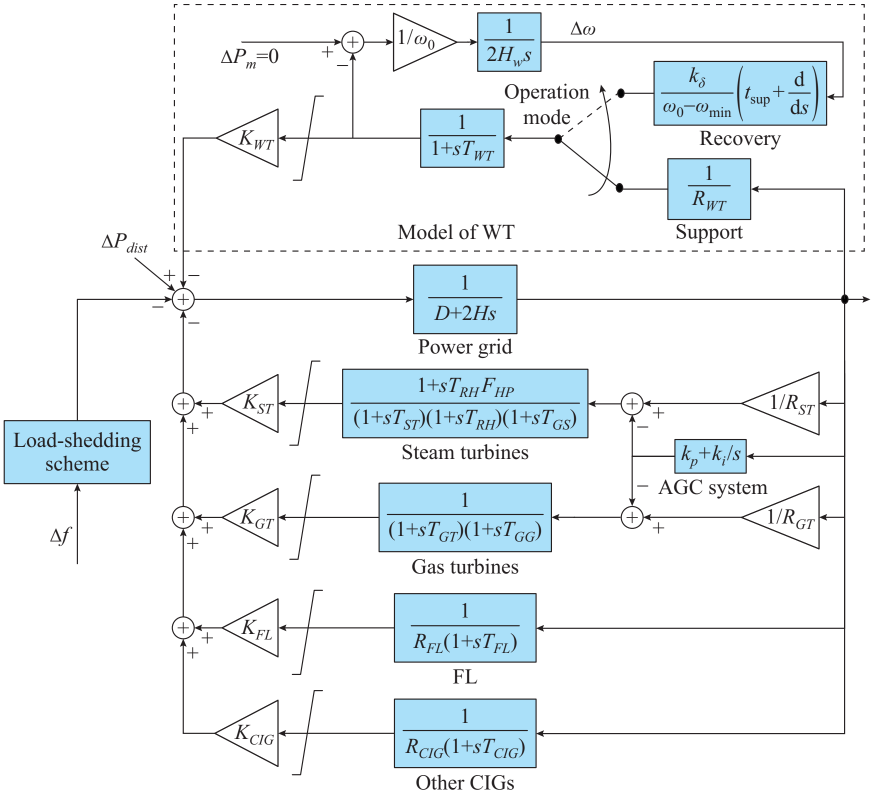


Fig. 4. Approximate model for frequency stability study in single-area power grid following active power disturbance.

Figure 5 illustrates an exemplary FDI attack path. The attacker may take the following steps [38], [39].

Step 1: bypass the firewall of the web server control center.

This step aims to establish initial access to the internal network. Attackers may exploit known vulnerabilities in web server softwares (e. g., Apache, Nginx) or web applications (e.g., SCADA monitoring interfaces).

Step 2: bypass the firewall of the ICCP server and access LAN of the WF control center. This step focuses on lateral movement within the network to gain access to the SCADA system. Techniques such as port scanning, exploiting trust relationship, creating tunnels, or leveraging stolen credentials and pass-the-hash attacks may be employed.

Step 3: log into SCADA server. Adversaries may utilize the methods such as exploiting SCADA software vulnerabilities and credential theft (e.g., through phishing emails, social engineering, and malware installation). Persistent access may also be achieved by installing backdoors on compromised systems [38].

Step 4: Manipulate targeted data. Manipulating targeted data requires understanding the SCADA protocols used by the GDA module (e.g., Modbus and DNP3) to map data objects (DOs) to their corresponding physical parameters within the WF [38]. By analyzing communication traffic between the SCADA server and the GDA module, attackers can identify and manipulate targeted data attributes (DAs).

In this paper, the objective of the attacker is to drive the actual grid frequency beyond the thresholds f_{thr} defined by grid codes [40], e.g., 59.1 Hz.

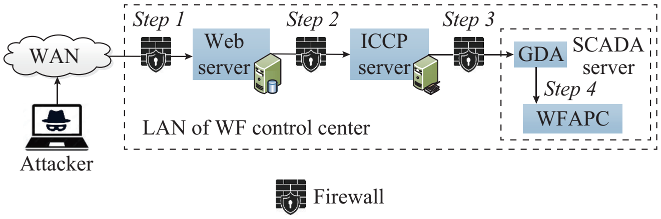


Fig. 5. Exemplary FDI attack path.

Under such a condition, load-shedding schemes may be activated to avoid frequency instability. If the frequency drop is not captured by activation of load-shedding steps, rotating generators may begin to disconnect to avoid physical damage, which results in system collapse. For instance, in the FDI attack model proposed in this paper, the attackers can force the WFs to decline their power output by sending a fake over-frequency signal to WFAPC. This may immediately cause a power generation deficit in the power grid. Thus, the target of the attacker is WFAPC, and the attack vector includes the frequency measurement received by the control center of the WF.

In a non-stealthy FDI attack, $\tilde{f}(k)$ can be chosen arbitrarily by the attacker at any time step k, but there is a high chance of being ineffective or even being detected easily. Note that, k is in milliseconds and $k \in \{0, t_{sample}, 2t_{sample}, ..., T_n\}$. To launch a stealthy FDI attack, the attack variables should at least satisfy the following constraints.

$$\begin{cases} f_{\min} \le \tilde{f}(k) \le f_{nom} - \Delta f_{DB} \\ f_{nom} + \Delta f_{DB} \le \tilde{f}(k) \le f_{\max} \end{cases} \quad \forall k \tag{3}$$

$$\widetilde{RoCoF}(k) \leq RoCoF_{\max} \quad \forall k \tag{4}$$

$$\tilde{\tau} \leq k \leq t_{sup} + \tilde{\tau} \tag{5}$$

In the rest of this paper, only stealthy FDI attacks are considered due to their severity.

Equation (3) indicates that $\tilde{f}(k)$ should be out of the frequency deadband Δf_{DB} to keep the WFAPC in frequency support mode (FSM). Moreover, if the frequency goes lower or higher than $f_{\min}$ and $f_{\max}$, respectively, the corresponding relays, which are much faster than WFAPC, may isolate the WF. Therefore, $\tilde{f}(k)$ out of the frequency deadband is disregarded as bad data by WFAPC. Using the same logic, (4) puts constraint on the RoCoF of $\tilde{f}(k)$ (e.g., ±2.00 Hz/s for 500 ms moving-average window, ±1.50 Hz/s for 1000 ms moving-average window, and ±1.25 Hz/s for 2000 ms moving-average window [41]). In addition, the time constraint of frequency support is imposed by (5). After passing t_{sup}, the WF may again be in MPPT mode even if there are any frequency deviations. The trajectory of $\tilde{f}(k)$ is depicted in Fig. 6. When $\tilde{f}(k)$ is sent to WFAPC, its output may deviate from the accurate value by $\widetilde{\Delta P}$, as shown in (6). The aim of the attacker is to create enough $\widetilde{\Delta P}$ to cross f_{thr} and ultimately trigger load-shedding schemes.

$$\widetilde{\Delta P}(z) = T_{WT} \frac{K_{WT}}{R_{WT}} \frac{z}{z - \mathrm{e}^{-\frac{t_{sample}}{T_{WT}}}} \left(\tilde{f}(z) - f_{nom}\right) \tag{6}$$

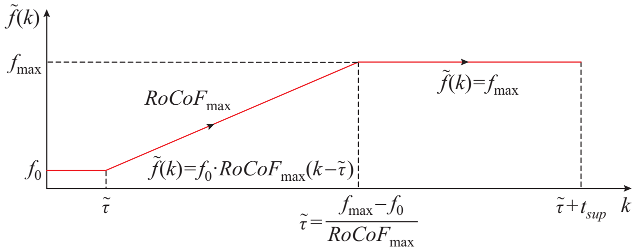


Fig. 6. Trajectory of $\tilde{f}(k)$.

Two different stealthy FDI attack models based on the level of the attacker's knowledge about the grid parameters are discussed in the next subsections. The attackers may know the exact values of parameters mentioned in Fig. 4, i.e., the white-box FDI attack, or they have to estimate the parameters based on historical data, i.e., the black-box FDI attack.

B. White-box FDI Attack Model

In the white-box FDI attack model, we assume that the attackers have enough knowledge about the grid parameters, i.e., H, D, F_{HP}, K_X, R_X, and T_X. Thus, using model (2) and applying $\tilde{f}(k)$, they can predict the frequency changes following $\widetilde{\Delta P}$.

Figure 7 shows the flow chart of the algorithm for the attacker to launch white-box FDI attack considering the load and generation profile. This algorithm can be time-wise divided into three stages:

Stage 1: long-term stage. The adversaries predict $P_L(d)$ and the maximum $\bar{P}_W(d)$ using historical data. Then, they find the day on which the maximum dependency of the power grid on wind energy occurs (day d^*). We measure this dependency using index $LDWI(d)$ as:

$$LDWI(d) = \frac{\bar{P}_W(d)}{P_L(d) - \bar{P}_W(d)} \quad \forall d \in [1, 365] \tag{7}$$

The higher the value of $LDWI$, the higher the chance of a successful attack. Additionally, employing this index can reduce the risk of the attacker being detected during unauthorized access to the LAN of WF control center. Attack detection and critical vulnerability patching can take organizations tens to hundreds of days, with longer detection time for more sophisticated threats [42], [43].

Stage 2: short-term stage. On day d^*, the required grid parameters should be updated every minute. Then, the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, and $P_W(t)$ are computed accordingly. Next, attackers need to calculate the value of the needed active power disturbance ΔP^*_{dist} to achieve $\Delta f = \Delta f_{thr}$ assuming that the initial frequency f_0 is the nominal frequency. Moreover, they should calculate the value of the needed frequency deviation Δf^* after the decline of $P_W(t)$. In other words, attackers aim to answer two key questions: what is the frequency drop caused by the WFs shutdown, and what active power shortage is needed to violate the frequency threshold?

Stage 3: real-time stage. The attackers have access to the measured $f(k)$ and $P_W(k)$. The attack will be successful if one of the two following conditions is met: first, $\Delta P^*_{dist} \leq P_W(k)$; second, $f(k) - \Delta f^* \leq f_{thr}$. Otherwise, go to the next time step in stage 2.

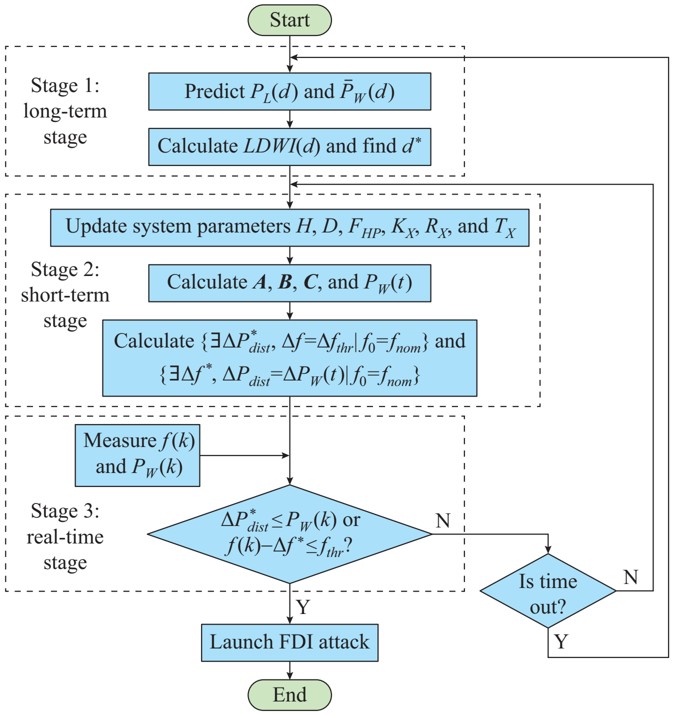

Start
Stage 1:
long-term
stage
Predict $P_L(d)$ and $\bar{P}_W(d)$
Calculate $LDWI(d)$ and find d^*
Update system parameters H, D, F_{HP}, K_X, R_X, and T_X
Stage 2:
short-term
stage
Calculate $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $P_W(t)$
Calculate $\{\exists \Delta P^*_{dist}, \Delta f=\Delta f_{thr} | f_0=f_{nom}\}$ and $\{\exists \Delta f^*, \Delta P_{dist}=\Delta P_W(t) | f_0=f_{nom}\}$
Measure $f(k)$ and $P_W(k)$
Stage 3:
real-time
stage
$\Delta P^*_{dist} \leq P_W(k)$ or $f(k)-\Delta f^* \leq f_{thr}$?
N
Y
Is time out?
N
Y
Launch FDI attack
End

Fig. 7. Flow chart of algorithm for attacker to launch white-box FDI attack.

C. Black-box FDI Attack Model

Contrary to the white-box FDI attack, in this subsection, the attackers do not have access to the accurate system parameters. Accordingly, stages 1 and 3 are the same as those shown in Fig. 7, but they have to employ a system identification method to estimate the parameters in stage 2, which are described in Fig. 8. In the first step, the attackers use the historical discrete form of WF power changes $\Delta P_W(z)$ and the grid frequency variations $\Delta f(z)$, which can be obtained by compromising the historian server. Then, a Z-transform technique is applied to find the amount of disturbance in power grid at time step k $e[k]$. The procedure to calculate ΔP^*_{dist} and Δf^* is detailed in Supplementary Material C.

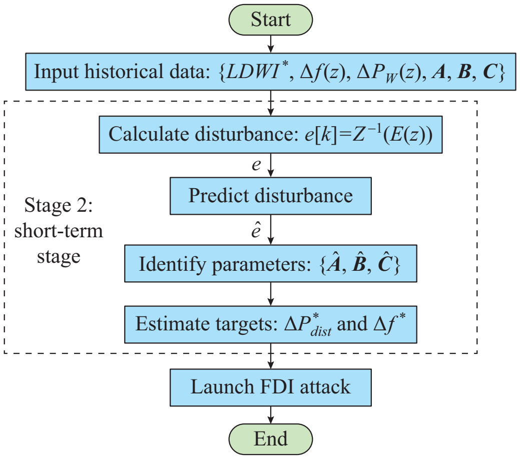

Start
Input historical data: $\{LDWI^*, \Delta f(z), \Delta P_W(z), \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$
Calculate disturbance: $e[k]=Z^{-1}(E(z))$
e
Predict disturbance
$\hat{e}$
Stage 2: short-term stage
Identify parameters: $\{\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}\}$
Estimate targets: ΔP^*_{dist} and Δf^*
Launch FDI attack
End

Fig. 8. Model of black-box FDI.

IV. Attack Detection and Mitigation

The standard intrusion detection systems (IDSs) are ill-equipped to detect the designed FDI attacks against frequency measurements of WFs due to several key factors: the stealthy nature of the attacks, the lack of power grid context of the IDS, the complexity of relevant communication protocols, the time-sensitive nature of frequency data, potential weaknesses in encryption, and the absence of specialized attack signatures. Therefore, it is preferable to detect and mitigate any cyber attacks where and when they occur before their propagation. Therefore, WFs should be outfitted with suitable cybersecurity measures, i.e., the primary detector. It must be able to operate based on an approximate model since the accurate model is not accessible to the operator/owner of the WF.

However, a secondary detector at the dispatch center of the power grid is warranted due to ① the potential for undeniable errors in the output of the primary detector as a result of the constantly evolving grid model, and ② the recognition that the operator of the grid retains responsibility for the security of the grid and has access to redundant measurements. The secondary detector serves to defend against potential undetected attacks bypassing the primary detector. Considering the limitations of sampling rates and communication delays in wide-area power grids, and to maintain coordination between the primary and secondary security layers, the secondary detector is designed to operate at a slower pace compared with the primary detector.

A generic diagram of the proposed bi-level detection and mitigation technique is depicted in Fig 9. Machine learning (ML)-based techniques are deployed as part of the proposed bi-level technique due to their capability to effectively capture uncertain patterns in noisy data and flexibility and adaptability in handling different wind conditions and grid operation scenarios. Accordingly, ML-based techniques are widely employed as the core of anomaly and attack detection in the context of frequency stability of the power grid [44]. At the control center of WF (primary level), the primary detector employs an observer to estimate the grid frequency, as shown in Fig 9. This estimation is then compared with the value received from the SCADA system using an SVM-based technique. At the dispatch center of the power grid (secondary level), a well-tailored LSTM-based technique is deployed to classify the data streams. Subsequently, a new power setpoint is commanded to the affected WFs.

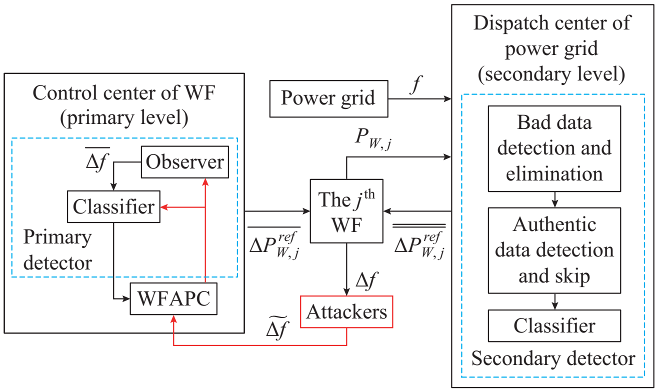

Control center of WF (primary level)
Power grid
f
Dispatch center of power grid (secondary level)
$\overline{\Delta f}$
Observer
Classifier
Primary detector
WFAPC
$P_{W,j}$
The j^{th} WF
$\overline{\Delta P^{ref}_{W,j}}$
$\overline{\overline{\Delta P^{ref}_{W,j}}}$
Δf
Attackers
$\widetilde{\Delta f}$
Bad data detection and elimination
Authentic data detection and skip
Classifier
Secondary detector

Fig. 9. Generic diagram of proposed bi-level detection and mitigation technique.

A. Algorithm of Primary Level

The model of the proposed bi-level technique is shown in Fig. 10. The primary detector works based on an approximate model of the power grid to estimate the frequency (corresponds to $\overline{\Delta f}$ in Fig. 10). Hence, the estimated frequency might contain an error. Thus, ML-based techniques are employed to recognize error behavior under various operation conditions in the training stage. Therefore, the trained ML-based techniques are utilized instead of applying a fixed threshold. To this end, the primary detector (SVM-based technique) is employed at the control center of the WF to compare the acquired measured value $\widetilde{\Delta f}$ with the estimated value $\overline{\Delta f}$.

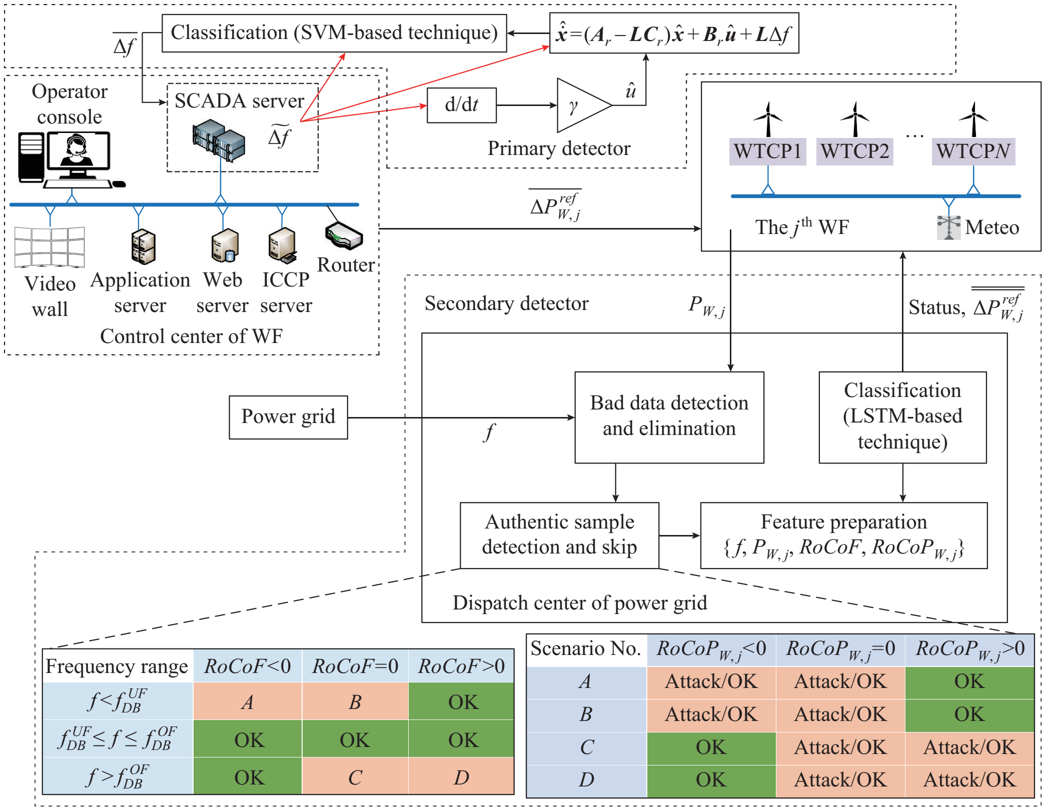
1) Observer Design

As shown in Fig. 10, the power disturbance $\hat{u}$ is estimated by a predefined constant coefficient γ due to the linear relationship between frequency derivation and the power disturbance. Then, $\overline{\Delta f}$ is calculated using a reduced-order model and a linear quadratic regulator (LQR) observer applying the following steps.

Step 1: a full-order linearized model of the power grid is extracted.

Step 2: the Hankel singular values of the power grid are calculated.

Step 3: the reduced-order model is obtained by keeping the appropriate number of the largest Hankel singular values and neglecting the rest.


$\overline{\Delta f}$
Classification (SVM-based technique)
$\dot{\hat{x}}=(A_r-LC_r)\hat{x}+B_r\hat{u}+L\Delta f$
Operator console
SCADA server
$\widetilde{\Delta f}$
d/dt
γ
$\hat{u}$
Primary detector
WTCP1
WTCP2
…
WTCPN
The jth WF
Meteo
$\overline{\Delta P_{W,j}^{ref}}$
Video wall
Application server
Web server
ICCP server
Router
Control center of WF
Secondary detector
$P_{W,j}$
Status, $\overline{\overline{\Delta P_{W,j}^{ref}}}$
Power grid
f
Bad data detection and elimination
Classification (LSTM-based technique)
Authentic sample detection and skip
Feature preparation $\{f, P_{W,j}, RoCoF, RoCoP_{W,j}\}$
Dispatch center of power grid

Frequency range	$RoCoF<0$	$RoCoF=0$	$RoCoF>0$
$f<f_{DB}^{UF}$	A	B	OK
$f_{DB}^{UF}\leq f\leq f_{DB}^{OF}$	OK	OK	OK
$f>f_{DB}^{OF}$	OK	C	D

Scenario No.	$RoCoP_{W,j}<0$	$RoCoP_{W,j}=0$	$RoCoP_{W,j}>0$
A	Attack/OK	Attack/OK	OK
B	Attack/OK	Attack/OK	OK
C	OK	Attack/OK	Attack/OK
D	OK	Attack/OK	Attack/OK

Fig. 10. Model of proposed bi-level technique.

Step 4: to cover the operation range of interest, three operation points determined by different WF instantaneous levels of integration (10%, 25%, and 50% of the power grid load) are considered.

Step 5: the Bode diagrams of the full-order and the approximated models are compared in these three operation points. The preferred approximated model exhibits minimal error, particularly within the frequency range of less than 10^3 rad/s in the Bode diagrams.

Step 6: the reduced-order model is considered to represent the system dynamics, whose matrices are indicated by $\boldsymbol{A}_r$, $\boldsymbol{B}_r$, and $\boldsymbol{C}_r$.

Step 7: Bryson's method [45] is applied to adjust symmetric weighting matrices ($\boldsymbol{Q}_r\geq\boldsymbol{0}$ and $\boldsymbol{R}_r>\boldsymbol{0}$) in the quadratic cost function.

Step 8: the observer gain vector $\boldsymbol{L}$ is calculated as $\boldsymbol{L}=\boldsymbol{R}_r^{-1}\boldsymbol{B}_r^{\mathrm{T}}\boldsymbol{S}$, where $\boldsymbol{S}$ is the solution to (8).

$$\boldsymbol{A}_r^{\mathrm{T}}\boldsymbol{S}+\boldsymbol{S}\boldsymbol{A}_r-\boldsymbol{S}\boldsymbol{B}_r\boldsymbol{R}^{-1}\boldsymbol{B}_r^{\mathrm{T}}\boldsymbol{S}+\boldsymbol{Q}_r=0 \tag{8}$$

Step 9: grid frequency is estimated as:

$$\begin{cases}\overline{\Delta f}=\boldsymbol{C}_r\hat{\boldsymbol{x}}\\ \dot{\hat{\boldsymbol{x}}}=(\boldsymbol{A}_r-\boldsymbol{L}\boldsymbol{C}_r)\hat{\boldsymbol{x}}+\boldsymbol{B}_r\hat{u}+\boldsymbol{L}\Delta f\end{cases} \tag{9}$$

2) Design of ML-based Technique at Primary Level

The steps taken to train an efficient ML-based technique for classification at the primary level are explained in Supplementary Material D.

3) Mitigation at Primary Level

The primary detector replaces the measured value $\widetilde{\Delta f}$ with the estimated value $\overline{\Delta f}$.

B. Algorithm of Secondary Level

As shown in Fig. 10, the dispatch center of the power grid takes over from the control center of the WF when the primary detector fails.

1) Design of ML-based Technique at Secondary Level

The LSTM-based technique can extract information from past data streams, forget less informative ones, and update learnable weights, i.e., the input weight matrix $\boldsymbol{W}$, the recurrent weight matrix $\boldsymbol{R}$, and the bias matrix $\boldsymbol{b}$. Therefore, LSTM-based technique is employed for attack detection at the secondary level due to the time-dependency of the data in the underlying system. The design of ML-based technique at the secondary level is explained in the Supplementary Material E.

2) Mitigation at Secondary Level

At first, the dispatch center of the power grid receives secure $P_{W,j}$ and f from the jth WF and from the power grid, respectively. Then, the data go through a filtering process to detect and eliminate bad data (outliers). Also, in order to lessen the computational burden, authentic sample detection and skip are done based on the logic illustrated in Fig. 10 (bottom tables). Based on the healthy frequency dynamic, the following cases are recognized as authentic samples: ① frequency is within the deadband ($f_{DB}^{UF}\leq f\leq f_{DB}^{OF}$); ② under the over-frequency condition ($f>f_{DB}^{OF}$), the $RoCoF$ is negative; ③ under the under-frequency condition ($f<f_{DB}^{UF}$), the $RoCoF$ is positive; ④ under the over-frequency condition ($f>f_{DB}^{OF}$), the $RoCoF$ is non-negative but the rate of change of active power of the jth WF $RoCoP_{W,j}$ is negative; and ⑤ under the under-frequency condition ($f<f_{DB}^{UF}$), the *RoCoF* is non-positive but $RoCoP_{W,j}$ is positive, which means the WF is supporting the grid frequency.

3) Time Coordination

As the measured values (signals) are continuous data streams, due to the high sampling rate of IEDs, a moving window is adopted to segregate the input data into smaller segments. Using this technique, the maximum allowable response time of LSTM-based technique can be extended, which may result in achieving higher accuracy. Once the LSTM-based technique detects a cyber attack against the j^{th} WF, the dispatch center of the power grid sends an FSM turn-on command (status is 1) to the j^{th} WF and replaces the corrupted reference value (signal) $\widetilde{\Delta P_{W,j}^{ref}}$ with the alternative value (signal) $\overline{\overline{\Delta P_{W,j}^{ref}}}=R_{WT}^{-1}(f_{nom}-f)$. Note that the compromised WFAPC is bypassed by the dispatch center of the power grid to avoid malfunctions. That means the secondary detector command is superior to the primary one in case of conflict.

It is worth noting that the grid operators should have an estimation of the maximum response time. Thus, the attack should be mitigated before activation of the load-shedding scheme. Equation (10) shows the minimum time of first-step load-shedding activation $t_{LSh1}^{\min}$. In addition, we set f_{DB}^{UF} and f_{LSh1} to be 59.8 Hz and 59.1 Hz, respectively.

$$\begin{cases} t_{LSh1}^{\min}=\dfrac{\Delta f_{LSh1}}{RoCoF_{\max}} \\ \Delta f_{LSh1}=f_{DB}^{UF}-f_{LSh1} \end{cases} \tag{10}$$

A constraint that should not be violated to have a timely response is given as:

$$\begin{cases} t_{LSh1}^{\min}-t_{delay}-T_{WT}\geq t_{ML}+t_n \\ t_n=nt_{sample}=\dfrac{n}{f_{sample}} \end{cases} \tag{11}$$

where $t_{LSh1}^{\min}\in[400,500]$ms; $t_{delay}\in[10,100]$ms; $T_{WT}\in[50,200]$ms; $t_{ML}\in[0,10]$ms; and $t_n\in[90,440]$ms [46].

Hence, it is recommended to consider the sampling rate, the number of samples, the communication delay, and the minimum time to trigger the load-shedding scheme in parameter tuning.

V. Performance Evaluation

In this section, numerical results are carried out to assess the consequences of cyber attacks against WFAPC in FSM, as well as the effectiveness of the proposed bi-level technique. For this purpose, the modified 39-bus New England system is studied, where generators connected to buses 30, 37, and 38 are replaced by three WFs, respectively, each including 330, 670, and 670 WTs of GE1.5xle [47]. In order to model the wind speed, the Weibull distribution of Penascal Wind Power Site, Texas, USA is utilized and extracted based on 7 years of measured data (2007-2013) [48]. Moreover, it is assumed that 70% of the power generation including WFs, are providing frequency support, i.e., total participation factor $K=0.7$ [35]. The generators connected to bus 39 are equipped with gas turbine, and the generators connected to buses 33 and 34 are equipped with steam turbine. Also, the load connected to bus 7 is considered to be an FL. Additionally, IEEE RTS load profile is adopted in this study, with an annual peak load of 8300 MW [49]. The AGC system controls the load reference of the conventional generators. Additionally, the load-shedding scheme is designed based on Western Electric Coordinating Council (WECC) grid code [40]. MATLAB simulations for the case studies are conducted on a computational platform featuring an Intel Xeon W-1370 CPU operating at 2.90 GHz and 32 GB of memory.

A. FDI Attack Consequence

It is assumed that the FDI attack is launched at $t=1$ s. The frequency and load-shedding caused by FDI attack are depicted in Fig. 11, where "WF: 0.100 p.u.→Load shedding: 0.053 p.u." denotes the WF capacity is 0.100 p.u. in the power grid and the cyber attack against it leads to 0.053 p.u. load-shedding in the power grid. Figure 11 shows that the FDI attack causes frequency instability when the instantaneous WF capacity is 0.450 p.u. or greater, even though full steps of load-shedding are done. In such circumstances, generators begin to disconnect from the grid, and islanding schemes may even occur. Moreover, even a relatively low share of compromised WFs, e.g. 0.150 p.u., can trigger the load-shedding scheme.

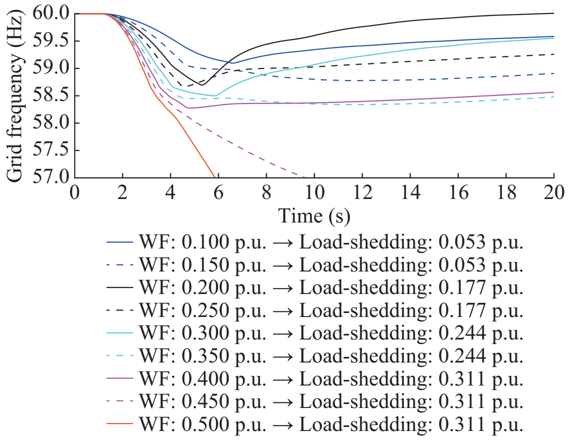


Fig. 11. Frequency and load-shedding caused by FDI attack.

In Fig. 12, the frequency nadir and load-shedding for different values of power grid inertia H, power grid stiffness $D+R^{-1}$, and participation factor K in FDI attack against WFAPC are depicted. Results show that power grid stiffness changes have the greatest impact on frequency nadir and load-shedding value. A decrease in stiffness from 10.0 p.u. to 6.5 p.u. doubles the frequency nadir. In realistic power grids, stiffness may change from 6.00 p.u. to 16.50 p.u. [35]. Therefore, the FDI attack may result in system collapse when $P_W>0.35$ p.u. and stiffness is less than 7.50 p.u.. Additionally, though H plays an undeniable role in the RoCoF right after the event, it has a limited impact on the frequency nadir. Moreover, frequency stability is improved significantly at $K=77\%$.

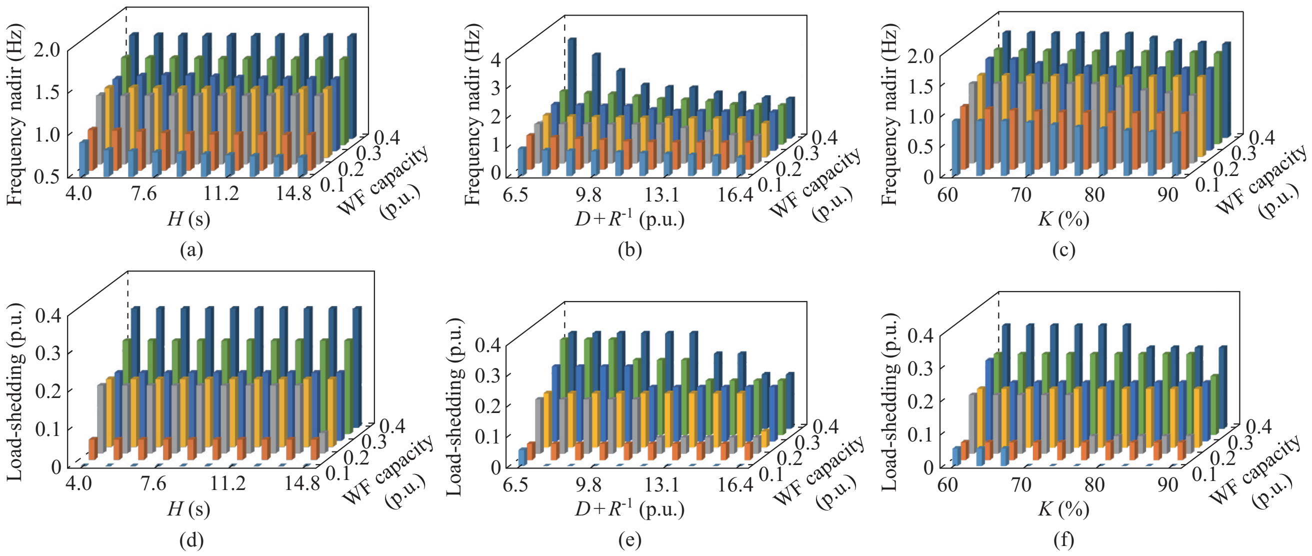


Fig. 12. Sensitivity analysis of frequency nadir and load-shedding for different values of power grid inertia, stiffness, and participation factor in FDI attack against WFAPC. (a) Frequency nadir v.s. H. (b) Frequency nadir v.s. $D+R^{-1}$. (c) Frequency nadir v.s. K. (d) Load-shedding v.s. H. (e) Load-shedding v.s. $D+R^{-1}$. (f) Load-shedding v.s. K.

1) White-box FDI Attack

In white-box FDI attacks, adversaries can go through the algorithm shown in Fig. 7 and select the best attack time. The results of the modified IEEE 39-bus system is illustrated in Fig. 13.

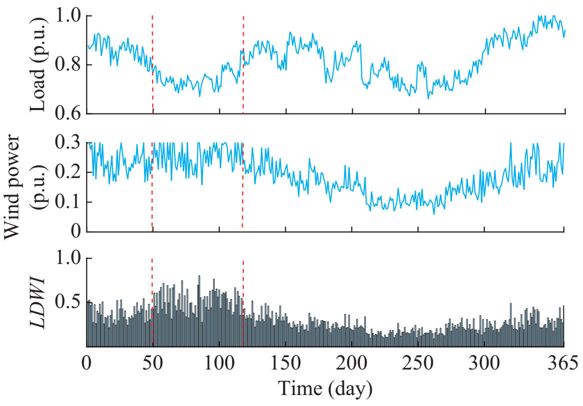


Fig. 13. Results of modified IEEE 39-bus system.

Inspired by [42] and [43], if an intruder remains undetected within the LAN of the control center of the WF, there is a high probability of launching a maximally effective attack during the identified 80-day period (dashed red line), leading to power grid collapse. Figure 13 shows that *LDWI* can even reach 80% at $d^*=69$ although the installed capacity of WFs accounts for approximately 15% of the total system generation capacity and the rated WF generation is about 30% of the annual peak load.

The results of FDI attacks considering different WF capacities are shown in Table I. In Table I, the best day and the best moment are found by the "long-term stage" and "short-term and real-time stages", respectively. In addition, random day/moment refers to skipping the corresponding stage. A Monte Carlo simulation is employed and the average result is reported for a random day/moment. For instance, if no capacity is applied, choosing the best day by the attackers may result in system collapse. However, if a random day is picked, the power grid may survive, but it suffers from 0.244 p.u. load-shedding for the best moment or 0.177 p.u. load-shedding for a random moment. Even if a WF capacity of 40% is applied, the attack on the best day may result in system collapse. In case the WF capacity is 30%, the attack at the best day and moment may result in 0.311 p.u. load-shedding while a random attack results in only 0.053 p.u. load-shedding.

Comparing Fig. 11, Fig. 13, and Table I, we observe that an FDI attack compromising approximately 3% of the WF capacity can trigger load-shedding, while exceeding 15% of the WF capacity leads to a rapid collapse of the power grid.

2) Black-box FDI Attack

As detailed in Section III-C, black-box FDI attacks, which rely on historical data, are susceptible to estimation errors. Here, the effects of the attacker estimation error on the success of attackers are investigated. To this end, a sensitivity analysis is conducted to determine the relationship between the successful attack and variations in relevant parameters. Specifically, the analysis aims to identify the parameters whose estimation accuracy significantly influences the successful attack and to quantify the tolerable error margins.

The partial correlation matrix of the black-box FDI attack is shown in Table II, where system and generator parameters are set as inputs; and frequency nadir and load-shedding are set as outputs.

Output power of WF varies from 10% to 60% of the power grid demand, with the corresponding adjustments to K_{WT}. In addition, the attack is designed based on the known parameters. Then, parameters D, H, K_{WT}, F_{HP}, T_{GT}, T_{RH}, and R^{-1} are changed from −10% to 10% with the interval of 0.1% one by one. Finally, the attack is applied to calculate frequency nadir and load-shedding. The obtained results in Table II show that K_{WT} and R_{WT} have the most considerable effect on the outputs.

TABLE I
RESULTS OF FDI ATTACKS

WF capacity (%)	Day	*LDWI*	Moment	Load-shedding (p.u.)	Δf_{nadir} (Hz)	Result
0	Best	0.744	Best	0.311	Null	Collapse
			Random	0.311	8.418	Collapse
	Random	0.361	Best	0.240	1.530	Load-shedding
			Random	0.170	1.354	Load-shedding
40	Best	0.667	Best	0.311	Null	Collapse
			Random	0.311	4.066	Collapse
	Random	0.357	Best	0.244	1.524	Load-shedding
			Random	0.177	1.353	Load-shedding
30	Best	0.429	Best	0.311	1.726	Load-shedding
			Random	0.244	1.512	Load-shedding
	Random	0.293	Best	0.177	1.364	Load-shedding
			Random	0.053	1.313	Load-shedding

TABLE II
PARTIAL CORRELATION MATRIX OF BLACK-BOX FDI ATTACK

Parameter	Frequency nadir (Hz)	Load-shedding (p.u.)
K_{WT}	0.985	0.980
D	–0.154	–0.228
H	–0.148	–0.156
R^{-1}	0.802	0.775
T_{RH}	0.042	0.028
F_{HP}	–0.062	–0.038
T_{ST}	0.018	0.004
T_{SG}	0.006	0.008
T_{GT}	0.058	0.021
T_{GG}	0.022	0.014

Note that R^{-1} is set based on grid code requirements [3], and attackers do not face troubles in estimating it. Additionally, among all turbine-governor parameters, F_{HP}, T_{GT}, and T_{RH} have relatively bigger partial correlation values but may be lower than those of D and H. Hence, the attackers can use typical values for turbine-governor parameters. It should be noted that a proper estimation of H and D is required to design a successful attack.

However, they are contingent upon the dynamic behavior of aggregated generators and loads, which are subject to inherent uncertainties stemming from load variations and generator dispatch schemes.

In Fig. 14, the impact of estimation error on the correct prediction of the power grid collapse is depicted from the perspective of the attacker for different WF participation factor K_{WT}. For instance, when $K_{WT}=0.491$, which means the WFs are providing 49% of the frequency support, the estimation error may not affect the attack result if estimation errors of D and H are less than 10.5% and 36%, respectively. In addition, within a specific narrow bound for K_{WT}, from 0.484 to 0.487, the accurate values of D and H are needed. As a result, in the black-box FDI attack, perfect knowledge about K_{WT} is required, and D and H need to be estimated carefully depending on K_{WT}.

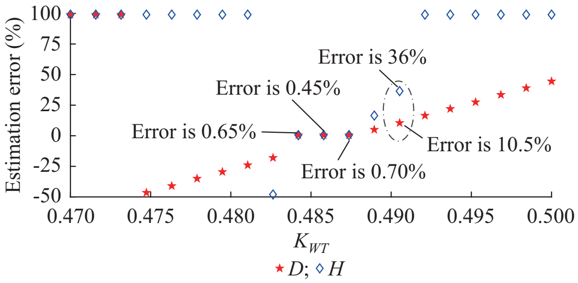


Fig. 14. Impact of estimation error on correct prediction of power grid collapse.

B. Evaluation of Proposed Bi-level Technique

1) Primary Detector

To construct the required dataset, the following operation ranges are considered: wind speed range is [3.5, 11.5]m/s and load deviation range is ±0.3 p.u., which encompass the specified period depicted in Fig. 13. To simulate the disturbances, a three-phase short circuit lasting three cycles is randomly applied to nearby buses in 50% of the benign samples. Consequently, the attack initiation time is randomly selected within this highlighted period. A total of 120000 samples are generated, comprising 60000 benign samples (with and without short circuits) and 60000 FDI attack samples, equally divided into 50% black-box and 50% white-box FDI attacks. The dataset is divided into training, validation, and test sets with ratios of 70%, 15%, and 15%, respectively. The training process is conducted offline using a reliable dataset. The load-shedding sensitivity to communication latency is shown in Table III, where the WF capacities are chosen as 0.35, 0.40, 0.45, 0.50, and 0.55 p.u.. Based on the results, SVM-based technique shows the highest performance as the primary detector in the proposed bi-level technique. SVM-based technique employs a radial basis function kernel, which is selected due to its effectiveness in capturing nonlinear relationships inherent in cyber-physical measurements. The key hyperparameters are determined via cross-validation: the regularization parameter is set as $C=50$, and the kernel coefficient is set as $\gamma=0.15$. To compensate for any potential class imbalance in the dataset, the class weight parameter is set to be balanced. Note that the grid parameters are the same in all samples and are equal to the system parameters used in the observer design. The impacts of errors in turbine-governor parameters and power grid parameters within range of ±50% are investigated in Fig. 15. As shown in Fig. 15, the classification accuracy of the primary detector may downgrade following the grid parameter changes. Therefore, the secondary detector is proposed in this paper to back up the primary detector.

TABLE III
LOAD-SHEDDING SENSITIVITY TO COMMUNICATION LATENCY

Communication delay (s)	Load-shedding (p.u.)				
	0.35 p.u.	0.40 p.u.	0.45 p.u.	0.50 p.u.	0.55 p.u.
0.01	0	0	0	0	0
0.05	0	0	0	0	0.053
0.10	0	0	0.053	0.053	0.053
0.15	0	0	0.053	0.053	0.053
0.20	0	0	0.053	0.053	0.117
0.25	0	0.053	0.053	0.117	0.117

As the primary detector uses data from IEC 61400-25 compatible protocols, it is inherently prone to communication delays.

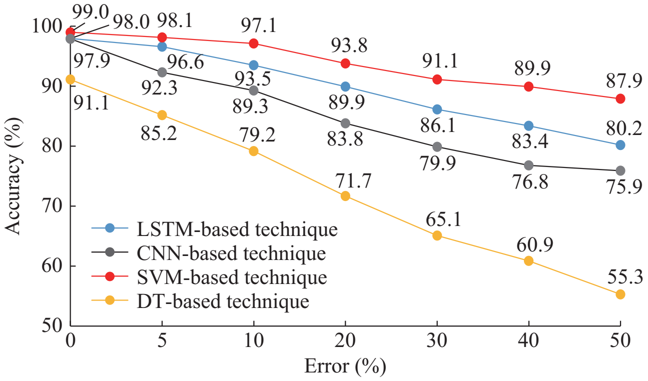


Fig. 15. Classification accuracy of primary detector.

The classification performance of the SVM-based technique is compared with CNN-based and DT-based techniques. Table IV demonstrates the detector classification performances. The time constant of WT and the moving window length are 0.15 s and 0.5 s, respectively. Specifically, with a WF capacity of 0.45 p.u., a 100-ms latency results in a one-step load-shedding response (0.053 p.u.). This highlights that as WF capacity increases, the influence of communication latency on the attack mitigation technique becomes a critical concern, necessitating careful consideration in control system design.

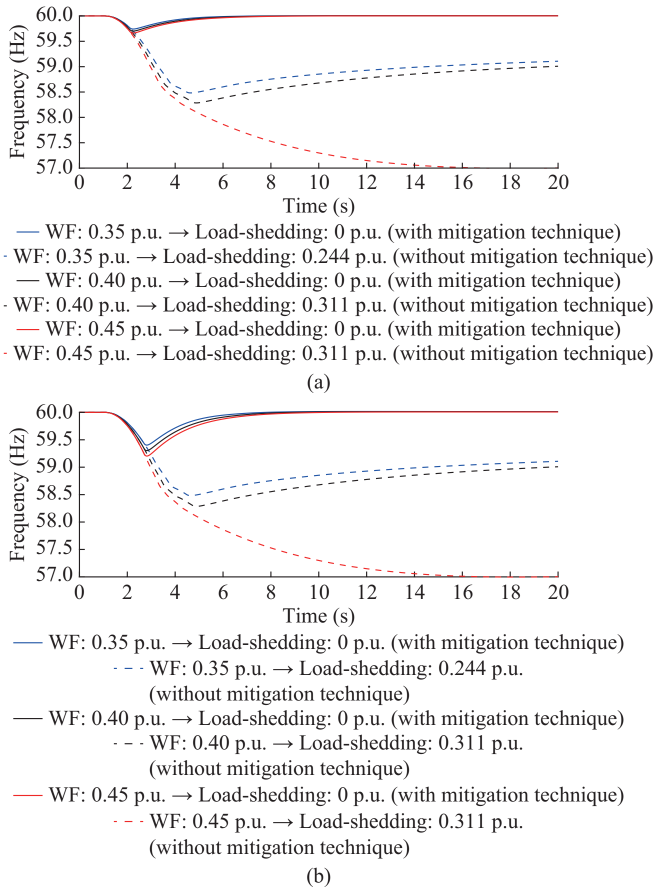
TABLE IV
DETECTOR CLASSIFICATION PERFORMANCES

Detector	Indicator	LSTM-based technique	SVM-based technique	CNN-based technique	DT-based technique
Only primary	Accuracy (%)	97.9500	98.9900	97.9000	91.1000
	Frequency nadir ratio (FNR)	0.0210	0.0016	0.0120	0.0800
	FPR	0.0200	0.0186	0.0300	0.0980
	$F1$-score	0.9795	0.9900	0.9792	0.9118
Only secondary	Accuracy (%)	99.1100	93.8800	96.8000	89.1100
	FNR	0.0016	0.0424	0.0240	0.0598
	FPR	0.0162	0.0800	0.0400	0.1580
	$F1$-score	0.9911	0.9399	0.9683	0.8962

Figure 16 depicts the grid frequency after FDI attack with and without the mitigation technique at the best day and best moment for varying levels of WF capacity. Figure 16(a) shows the isolated performance of the primary detector, while Fig. 16(b) demonstrates the performance of the secondary detector when the primary detector is deactivated. Figure 16(a) shows that the mitigation technique effectively maintains frequency without any load-shedding in all cases with different levels of WF capacity. However, without the mitigation, the FDI attack can lead to load-shedding or even power grid collapse. If WF capacity is 0.350 p.u. and 0.400 p.u., the LFC maintains the grid frequency after load-shedding of 0.244 p.u. and 0.311 p.u., respectively. Yet, when WF capacity is 0.450 p.u., the grid frequency becomes unstable following the attack, and even four steps of load-shedding (0.311 p.u.) fail to stabilize it. However, the mitigation technique prevents power grid collapse in this scenario and maintains frequency without any load-shedding.

2) Secondary Detector

The dataset for training and testing of the secondary detector is generated considering the following operation ranges: ① wind speed: [3.5,11.5]m/s; ② load deviations: ±0.3 p.u.; ③ turbine-governor parameters: ±10%; and ④ power grid parameters: ±10%, which are the same as the primary detector. The assumed values of parameters in hyperparameter tuning process of the LSTM-based technique are given as follows: $\alpha=0.8$, $\Delta f_{LSh1}=0.7$ Hz, $\tau_{tr}=5$ min, $t_{LSh1}^{\min}=470$ ms, $\tau_{ts}=3.5$ s, $t_{delay}=80$ ms, $\mu=10$, $t_{sample}=20$ ms, $RoCoF_{\max}=1.5$ Hz/s (over 1 s window), $t_n=10$ s, $T_{WT}=150$ ms, and $n_{sample}=10$. The optimization model (20) in Supplementary Material D is applied to tune hyperparameters of the LSTM-based technique as follows: ① the number of hidden units is 100; ② the initial learning rate is 0.001; ③ the rate of learning rate dropout is 0.2; ④ the learning rate drop period is 10 s; ⑤ the maximum value of epochs is 180; ⑥ the minimum batch size is 80; ⑦ the validation frequency is 20 Hz; and ⑧ the solver is root mean square propagation.



(a)

(b)

Fig. 16. Grid frequency after FDI attack with and without mitigation technique. (a) Primary detector. (b) Secondary detector.

Based on the simulation results, T_{ML} falls within 4-7 ms. Based on the results, LSTM-based technique shows the highest classification performance, since it is particularly effective at capturing and recognizing patterns in sequential data.

The grid frequency and load-shedding after the FDI attack applying the secondary detector are illustrated in Fig. 16(b). It shows that the mitigation technique maintains frequency without any load-shedding in all cases. The slower response of the secondary detector compared with that of the primary detector causes a slightly larger frequency nadir.

The classification performance of the mitigation technique at the secondary level is assessed for different time window lengths t_n, and the results are presented in Table V, where the WF capacities are chosen as 0.35, 0.40, and 0.45 p.u.. It shows that the FDI attack is mitigated efficiently when $t_n \leq$ 0.5 s. However, the bigger t_n is, the larger the load-shedding required to maintain the frequency. For instance, in the case that t_n=1 s, e.g., $n=20$ and $t_{sample}=50$ ms, the secondary detector can not prevent load-shedding at any levels of WF capacity.

3) *Proposed Bi-level Technique*

The performance of the proposed bi-level technique against FDI attacks is investigated. Using the same dataset as Section V-A and applying the trained SVM-based technique and LSTM-based technique as the primary and secondary detectors, respectively, the obtained classification performances include accuracy of 99.88%, *FNR* of 0.0004, *FPR* of 0.002, and *F*1-score of 0.998.

TABLE V
LOAD-SHEDDING IN DIFFERENT TIME WINDOW LENGTHS

Time window length (s)	Load shedding (p.u.)		
	0.35 p.u.	0.40 p.u.	0.45 p.u.
0.2	0	0	0
0.4	0	0	0
0.5	0	0	0
0.6	0	0	0.053
0.8	0	0.053	0.053
0.9	0.053	0.053	0.053
1.0	0.053	0.053	0.117
1.4	0.117	0.117	0.244
2.0	0.117	0.244	0.311
2.4	0.244	0.311	0.311
Without mitigation	0.244	0.311	0.311

Thus, the proposed bi-level technique can significantly improve the security of WFs. The detection accuracy of the proposed bi-level technique is compared against existing techniques in Table VI. The results demonstrate that the proposed bi-level technique achieves superior performance, and a lower rate of false alarms.

TABLE VI
DETECTION ACCURACY COMPARISON BETWEEN PROPOSED BI-LEVEL TECHNIQUE AND EXISTING TECHNIQUES

Technique	Accuracy (%)	*FNR*	*FPR*	*F*1-score
Technique in [50]	95.26	0.0984	0.058	0.957
Technique in [51]	76.38	0.1947	0.291	0.795
Technique in [52]	87.86	0.1071	0.138	0.889
Technique in [53]	94.24	0.0588	0.057	0.948
Proposed bi-level technique	99.88	0.0004	0.002	0.998

VI. CONCLUSION

The development and integration of converter-based and renewable energy resources in modern power grids cause a lack of inertia and may threaten the frequency stability. Additionally, dispersed WTs are monitored and controlled from a control center via widespread cyber networks, which introduces vulnerabilities to cyber threats. In this paper, the vulnerabilities of WF to cyber threats in PFSM are investigated. The FDI attacks are designed based on the WF communication architecture, protocols, and grid code specifications. It is shown that FDI attacks on LAN of the WF control center can destroy the frequency stability of the entire power grid even if the installed WF capacity in the power grid is relatively low (e.g., 15%). Moreover, the impacts of knowledge of adversaries about the power grid parameters on the severity of the FDI attacks are discussed. In addition, a bi-level detection and mitigation technique is proposed, which is evaluated using a modified New England 39-bus system. The results show that attackers can shut down WTs without accessing the HMI and sending direct commands, which is more detectable. The investigated cyber attack in this paper is stealthy, indirect, and capable of causing severe damage, potentially leading to system collapse. At the same time, the proposed bi-level technique can ensure the power grid frequency remains stable within the valid range even under the condition of high wind energy integration.

References

[1] International Renewable Energy Agency. (2020, Jan.). Wind energy. [Online]. Available: https://www.irena.org/Energy-Transition/Technology/Wind-energy

[2] U.S. Energy Information Administration. (2020, Jun.). The central united states set several wind power records this spring. [Online]. Available: https://www.eia.gov/todayinenergy/ detail. php?id=44075

[3] Q. Wu and Y. Sun, *Modeling and Modern Control of Wind Power.* New York: John Wiley & Sons, 2018.

[4] A. Fernández-Guillamón, E. Gómez-Lázaro, E. Muljadi *et al*., "Power systems with high renewable energy sources: a review of inertia and frequency control strategies over time," *Renewable and Sustainable Energy Reviews*, vol. 115, p. 109369, Nov. 2019.

[5] National-Gride-SO. (2025, Mar.). The grid code-revision 14. [Online]. Available: https://cms.eirgrid.ie/sites/default/files/publications/GridCode-Version14.3.pdf

[6] W. Bao, Q. Wu, L. Ding *et al*., "A hierarchical inertial control scheme for multiple wind farms with BESSs based on ADMM," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 751-760, Apr. 2021.

[7] M. Ansari, M. Zadsar, S. Sebtahmadi *et al*., "Optimal sizing of supporting facilities for a wind farm considering natural gas and electricity networks and markets constraints," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105816, Jun. 2020.

[8] M. Ansari, M. Latify, and G. Yousefi, "GenCo's mid-term optimal operation analysis: interaction of wind farm, gas turbine, and energy storage systems in electricity and natural gas markets," *IET Generation, Transmission & Distribution*, vol. 13, no. 12, pp. 2328-2338, Jun. 2019.

[9] Hydro-Quebec. (2009, Jan.). Transmission provider technical requirements for the connection of power plants to the hydro quebec transmission system. [Online]. Available: https://www.hydroquebec.com/

[10] M. Garmroodi, G. Verbič, and D. Hill, "Frequency support from wind turbine generators with a time-variable droop characteristic," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 676-684, Apr. 2018.

[11] Z. Guo and W. Wu, "Data-driven model predictive control method for wind farms to provide frequency support," *IEEE Transactions on Energy Conversion*, vol. 37, no. 2, pp. 1304-1313, Jun. 2022.

[12] H. Xu, C. Wang, Z. Wang *et al*., "Stability analysis and enhanced virtual synchronous control for brushless doubly-fed induction generator based wind turbines," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 5, pp. 1445-1458, Sept. 2024.

[13] X. Lyu, Y. Jia, and Z. Dong, "Adaptive frequency responsive control for wind farm considering wake interaction," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1066-1075, Sept. 2021.

[14] D. Sun, H. Liu, S. Gao *et al*., "Comparison of different virtual inertia control methods for inverter-based generators," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 768-777, Jul. 2020.

[15] International Electrotechnical Commission. (2025, Jan.). IEC international standards. [Online]. Available: https://iec.ch/publications/international-standards

[16] U.S. Department of Energy. (2020, Jul.). Road-map for wind cybersecurity. [Online]. Available: https://www. energy. gov/eere/wind/articles/roadmap-wind-cybersecurity

[17] M. McCarty, J. Johnson, B. Richardson *et al*., "Cybersecurity resilience demonstration for wind energy sites in co-simulation environment," *IEEE Access*, vol. 11, pp. 15297-15313, Feb. 2023.

[18] A. Zabetian-Hosseini, A. Mehrizi-Sani, and C. Liu, "Cyberattack to cyber-physical model of wind farm SCADA," in *Proceedings of IECON 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, USA, Oct. 2018, pp. 4929-4934.

[19] M. Ansari, M. Ghafouri, and A. Ameli, "Cyber-security vulnerabilities of the active power control scheme in large-scale wind-integrated power systems," in *Proceedings of 2022 IEEE Electrical Power and Energy Conference*, Victoria, Canada, Dec. 2022, pp. 79-84.

[20] M. Ghafouri, U. Karaagac, A. Ameli *et al*., "A cyber attack mitigation scheme for series compensated DFIG-based wind parks," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5221-5232, Nov. 2021.

[21] Y. Zhang, Y. Xiang, and L. Wang, "Power system reliability assessment incorporating cyber attacks against wind farm energy management systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2343-2357, Sept. 2017.

[22] M. Ghafouri, U. Karaagac, I. Kocar *et al*., "Analysis and mitigation of the communication delay impacts on wind farm central SSI damping controller," *IEEE Access*, vol. 9, pp. 105641-105650, Jul. 2021.

[23] H. Du, J. Yan, M. Ghafouri *et al*., "Modeling and assessment of cyber attacks targeting converter-driven stability of power grids with PMSG-based wind farms," *IEEE Transactions on Power Systems*, vol. 39, no. 5, pp. 6716-6728, Sept. 2024.

[24] A. Ameli, A. Hooshyar, E. El-Saadany *et al*., "Attack detection and identification for automatic generation control systems," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4760-4774, Sept. 2018.

[25] H. Badihi, S. Jadidi, Z. Yu *et al*., "Smart cyber-attack diagnosis and mitigation in a wind farm network operator," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9468-9478, Sept. 2023.

[26] P. Yang, X. Dong, Y. Li *et al*., "Research on primary frequency regulation control strategy of wind-thermal power coordination," *IEEE Access*, vol. 7, pp. 144766-144776, Oct. 2019.

[27] R. Azizipanah-Abarghooee, M. Malekpour, T. Dragičević *et al*., "A linear inertial response emulation for variable speed wind turbines," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1198-1208, Mar. 2020.

[28] M. Hwang, E. Muljadi, G. Jang *et al*., "Disturbance-adaptive short-term frequency support of a DFIG associated with the variable gain based on the ROCOF and rotor speed," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1873-1881, May 2017.

[29] G. Tu, Y. Li, and J. Xiang, "Coordinated rotor speed and pitch angle control of wind turbines for accurate and efficient frequency response," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3566-3576, Sept. 2022.

[30] I. Sardou and M. Ansari, "Risk-constrained self-scheduling of a generation company considering natural gas flexibilities for wind energy integration," *Journal of Renewable and Sustainable Energy*, vol. 12, no. 1, p. 013301, Jan. 2020.

[31] K. Doenges, L. Sigrist, I. Egido *et al*., "Wind farms in AGC: modelling, simulation and validation," *IET Renewable Power Generation*, vol. 16, no. 1, pp. 139-147, Jan. 2022.

[32] J. Huang, Z. Yang, J. Yu *et al*., "Optimization for DFIG fast frequency response with small-signal stability constraint," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 2452-2462, Sept. 2021.

[33] Z. Zhang, J. Hu, J. Lu *et al*., "Detection and defense method against false data injection attacks for distributed load frequency control system in microgrid," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 3, pp. 913-924, May 2024.

[34] S. Heier. (2014, Apr.). Grid integration of wind energy: onshore and offshore conversion systems. [Online]. Available: https://www.ndls.org.cn/standard/detail/

[35] A. Gorbunov, J. Peng, J. Bialek *et al*., "Can center-of-inertia model be identified from ambient frequency measurements?" *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2459-2462, May 2022.

[36] O. Stanojev, U. Markovic, P. Aristidou *et al*., "MPC-based fast frequency control of voltage source converters in low-inertia power systems," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 3209-3220, Jul. 2022.

[37] A. Tummala and R. Inapakurthi, "A two-stage Kalman filter for cyber-attack detection in automatic generation control system," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 1, pp. 50-59, Jan. 2022.

[38] E. Knapp and J. Langill. (2011, Sept.). Industrial network security: securing critical infrastructure networks for smart grid, SCADA, and other industrial control systems. [Online]. Available: https://dl.acm.org/doi/10.5555/2597834

[39] R. Sangkhro and A. Agrawal, "Cybersecurity in industrial control systems: a review of the current trends and challenges," in *Proceedings of 10th International Conference on Computing for Sustainable Global Development*, New Delhi, India, Mar. 2023, pp. 355-359.

[40] *IEEE Guide for Abnormal Frequency Protection for Power Generating Plants*, IEEE Standard C37.106-2003, 2003.

[41] European Network of Tranmission System Operators. (2018, Jan.). Rate of change of frequency withstand capability. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents

[42] Mandiant Consulting. (2024, Jan.). M-trends 2024 special report. [Online]. Available: https://www.defenseone.com/assets/m-trends-2024-special-report/portal/

[43] L. Tidjon, M. Frappier, and A. Mammar, "Intrusion detection systems:

a cross-domain overview," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3639-3681, Nov. 2019.

[44] L. Xi, L. Zhou, Y. Xu *et al*., "A multi-step unified reinforcement learning method for automatic generation control in multi-area interconnected power grid," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1406-1415, Apr. 2021.

[45] X. Deng, X. Sun, R. Liu *et al*., "Optimal analysis of the weighted matrices in LQR based on the differential evolution algorithm," in *Proceedings of 29th Chinese Control and Decision Conference*, Chongqing, China, May 2017, pp. 832-836.

[46] ABB Group. (2019, Feb.). 650 series IEC 61850 communication protocol manual. [Online]. Available: https://search.abb.com/library/

[47] General Electric Company. (2023, Jan.). Capacity factor leadership in class II winds. [Online]. Available: https://kipdf.com/ge-energy-renewable-energy-ge-s-capacity-factor-leadership-in-class-ii-winds5ac8604c1723ddc6d69a1576.html

[48] National Renewable Energy Laboratory. (2023, Jan.). Wind prospector, National Renewable Energy Laboratory. [Online]. Available: https://www.nrel.gov

[49] C. Barrows, A. Bloom, A. Ehlen *et al*., "The IEEE reliability test system: a proposed 2019 update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119-127, Jan. 2020.

[50] M. Sun, I. Konstantelos, and G. Strbac, "A deep learning-based feature extraction framework for system security assessment," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5007-5020, Sept. 2019.

[51] S. Liu, S. You, H. Yin *et al*., "Model-free data authentication for cyber security in power systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4565-4568, Sept. 2020.

[52] S. Ahmed, Y. Lee, S. Hyun *et al*., "Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2765-2777, Oct. 2019.

[53] V. Singh and M. Govindarasu, "A cyber-physical anomaly detection for wide-area protection using machine learning," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3514-3526, Jul. 2021.

Mostafa Ansari received the B. Sc. degree in electrical engineering from Shahid Beheshti University, Tehran, Iran, in 2014, and the M. Sc. degree in electrical engineering of power systems from Isfahan University of Technology, Isfahan, Iran, in 2017. He is current pursuing the Ph.D. degree in Concordia University, Montreal, Canada, since 2022. His research interests include power system dynamic control, power system economy, integration of large-scale renewable resources into power grid, and modeling of power systems with a focus on resilient operation of wind energy.

Mohsen Ghafouri received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2009 and 2011, respectively, and the Ph.D. degree in electrical engineering from Polytechnique Montreal, Montreal, Canada, in 2018. In 2018, he was a Researcher with CYME International, Eaton Power System Solutions, Montreal, Canada. In August 2018, he joined as the Horizon Postdoctoral Fellow with Security Research Group, Concordia University, Montreal, Canada, where he is currently an Associate Professor. His research interests include smart grid, power system modeling, microgrid, wind energy, and control of industrial process.

Amir Ameli received the B. Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2011, the M. Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2013, and the Ph. D. degree in electrical engineering from the University of Waterloo, Waterloo, Canada, in 2019. He was a Postdoctoral Fellow with the Electrical and Computer Engineering Department, University of Waterloo, from August 2019 to July 2020. Currently, he is an Assistant Professor with the Electrical Engineering Department, Lakehead University, Thunder Bay, Canada. He is a registered Professional Engineer in Thunder Bay, Canada. His current research interests include power system cybersecurity and protection.

Ulas Karaagac received the B. Sc. and M. Sc. degrees from Middle East Technical University, Ankara, Turkey, in 1999 and 2002, respectively, and the Ph.D. degree from École Polytechnique de Montreal (affiliated with Université de Montreal), Montreal, Canada, in 2011. He has 25 years of diverse experience in power engineering, spanning industry, and academia. He has been involved in numerous research and industry projects across Europe, North America, and Asia. In 2025, he joined the Electrical-Electronics Engineering Department at Middle East Technical University. His research interests include integration of large-scale renewable energy into power grid, modeling and simulation of large-scale power system, and power system dynamic control.

Ilhan Kocar received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Orta Doğu Teknik Üniversitesi, Ankara, Turkey, in 1998 and 2003, respectively, and the Ph.D. degree in electronics engineering from Polytechnique/Université de Montreal, Montreal, Canada, in 2009. He has 25 years of diverse experience in the power engineering field across industry, academia, and major regions including North America, Asia, and Europe. He is a Full Professor at Polytechnique Montreal, Montreal, Canada, and President of DIgSILENT North America Inc. His research aims to address critical challenge in integrating renewable energy source into power system.