An Extended False Data Injection Attack Via Deep Reinforcement Learning: Attack Model and Countermeasures in Cyber-Physical Power Systems

Xiaohong Ran, *Member, IEEE*, and Lei Ma, *Member, IEEE*

***Abstract*—False data injection attacks are commonly used to evade the bad data detector in cyber-physical power systems. This paper proposes an extended attack strategy and a deep reinforcement learning-based detection method. Traditional false data injection attacks aim to remain stealthy and avoid detection by conventional detection mechanisms. An extended load attack is introduced to increase the potential for damage. Directly adding an extended component directly to the measurement makes it easily detectable by bad data detector. Accordingly, the extended attack integrates the added component into the state variables to improve stealth. An optimization model for the extended components of the proposed attack is developed, along with a homologous matrix. Additionally, an online attack detection scheme is formulated as a partially observable Markov decision process problem. A deep reinforcement learning-based detection framework is proposed, featuring a compound reward designed to minimize false alarms and time delays. The proposed online detector extracts state features under varying operating conditions and generates a policy to determine whether the power grid is under attack. An extended Euclidean distance indicator and an adaptive weight matrix are also proposed in the dynamic state estimation to improve estimation or detection accuracy. Numerical experiments validate the effectiveness and robustness of the proposed deep reinforcement learning-based detection scheme in power systems.**

Note to Practitioners

This paper is motivated by the lack of research on modeling the destructive capabilities of cyber-attacks and the inaccuracy of anomaly detection methods for data integrity attacks. Existing approaches to modeling false data injection attacks primarily focus on the hidden features bypassing detection of bad data detection in cyber-physical power systems. This paper proposes a novel false data injection modeling approach, triggered by the spinning reserves of power grids. The proposed false data injection strategy mathematically characterizes extended stealth attack mechanisms and introduces an extended load attack to explore greater destructive potential. Considering the uncertain environments in perspective attackers and defenders, this study formulates the attack detection problem as a partially observable Markov decision process. This study then characterizes how such metrics of detectors can be efficiently computed, this can allow a defender to automatically learn or generate a detection threshold policy using the deep reinforcement learning, and distinguishes real false data injection attacks from system noises. To improve the detection performance, a novel indicator and an adaptive weight matrix are proposed to enhance learning efficiency of detectors. Numerical simulations suggest that the proposed detection scheme is feasible to both traditional and extended false data injection attacks, though it does not yet account for incomplete measurements. Future work will focus on designing attack detection strategies under conditions of incomplete knowledge of network topology and measurements.

***Index Terms*—Deep Reinforcement Learning; Initial False Data Injection Attacks; Extended False Data Injection Attacks.**

Xiaohong Ran is with the department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong; Lei Ma is with the Department of Computer Science, The University of Tokyo, Tokyo 113-8654, Japan, and also with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mails: xiaohong.ran@polyu.edu.hk, lma7@ualberta.ca).

NOMENCLATURE

A. Matrices

$\mathbf{H}$, $\mathbf{A}$	Measurement Jacobian and system matrixes.
$\boldsymbol{I}_M$, $\boldsymbol{I}_N$	Identity matrix with dimension $M(N)\times 1$.

B. Parameters

M, N	Number of measurements and system bus.
σ_v, σ_w	Standard deviations of process and measurement noises.
b_{ij}	Susceptance of the line ij.
P_{ij}^{L}	Power flow of line ij under normal condition.
$P_{ij}^{L,o}$	Power flow of line ij under I-FDIA condition.
θ_i, θ_j	Phase angle at bus i and j under normal condition.
θ_i^o, θ_j^o	Phase angle at bus i and j under I-FDIA condition.
P_i^D, P_i^G	Load and generation at bus i under normal condition.
$P_i^{D,o}$	Load at bus i under I-FDIA condition.
$P_i^{G,o}$	Generation at bus i under I-FDIA condition.
Γ, τ	Stopping time and attack launching time.
α, ϵ	Learning rate and probability.
θ, θ^-	Current and target networks parameters.
r, r_o, r_e	Residuals under normal, I-FDIA, E-FDIA conditions.
τ_s, c_β	A detection threshold and relative cost.
T	Maximum length of an episode.
N_m, N_d	Number of transition samples and bus loads.

C. Vectors

$\mathbf{z}_t$, $\mathbf{z}_t^o$	Measurements under normal and I-FDIA at time t.
$\mathbf{z}_t^e$, $\hat{\mathbf{y}}(\cdot)$	Measurements under E-FDIA and output action.
$\mathbf{w}_t$, $\mathbf{v}_t$	Measurement and process noise vectors.
$\mathbf{x}_t$, $\hat{\mathbf{x}}_t$	A state variable vector and its estimation.
$\mathbf{a}_o$, $\mathbf{c}_o$	I-FDIA and initial nonzero vector.
$\mathbf{a}_e$, $\boldsymbol{\delta}_e$	E-FDIA and extended attack component.
s_t, a_t	A state variable and action at time t.
G_t, r_t	A return and immediate reward.
φ_t, η_t	Cosine and Euclidean similarities.
$\boldsymbol{K}_t$	Kalman gain.
$\mathbf{F}_{t\|t}$	State covariance matrix.

D. Abbreviations

BDD	Bad data detector.
CKF	Cubature Kalman filter.
CPS	Cyber-physical system.
CUSUM	Cumulative sum.
DQN	Deep Q-network.
DRL	Deep reinforcement learning.

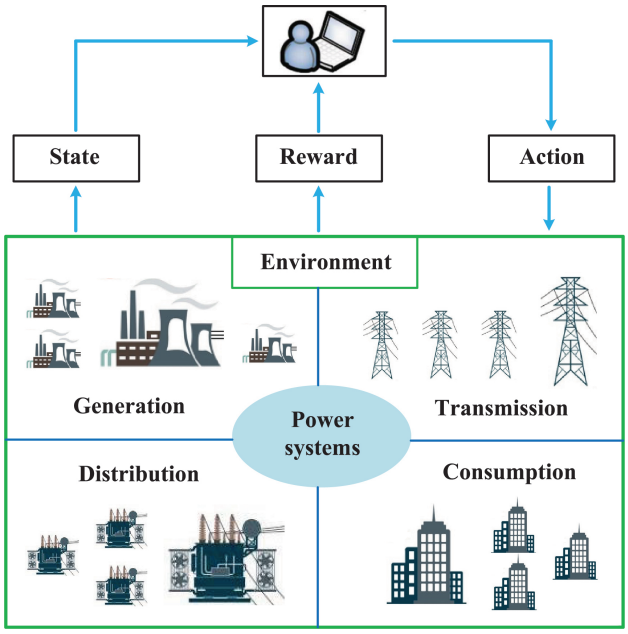


Fig. 1. Flowchart of Markov reward process in a power system.

EDSR Euclidian distance similarity ratio.
E-FDIA Extended false data injection attack.
FDIA False data injection attack.
I-FDIA Initial false data injection attack.
KF Kalman filter.
NAD Neural attack detection.
POMDP Partially observable Markov decision process.
RL Reinforcement learning.
SE State estimation.
SR Spinning reserve.
SRCKF Square-root cubature Kalman filter.
UKF Unscented Kalman filter.

I. Introduction

THE smart grid is a typical Cyber-Physical System (CPS), and its rapid development largely depends on advancements in information and communication technology. However, the reliance on information and communication technology [1] makes smart grids more vulnerable to various cyber-attacks, such as load redistribution attacks [2] and false data injection attacks (FDIAs) [3, 4]. Although different attacks have varied effects on power grids [5], FDIAs are particularly difficult to detect due to strong stealthiness. Moreover, deliberate FDIAs can cause significant economic losses to power grids and even lead to the collapse of power systems. Therefore, it is crucial to propose and design effective detection methods against FDIAs.

Recently, numerous studies have focused on detecting FDIAs in power systems. For example, a resilience-enhanced scheme in [6] has been presented to detect both conventional and collusive FDIAs in power systems. Additionally, a semi-supervised deep learning-based detection method was designed in [7], requiring only a limited number of labeled data in the training dataset. Other FDIA detection algorithms have also been designed, such as adversarial machine learning-based detection methods [8] and recurrent neural network-based attack detection [9]. Therefore, current detection methods can be categorized into two types: model-based methods [6] and data-driven algorithms [7 - 9]. For the latter method mentioned above, a flowchart illustrating the Markov reward process is shown in Fig. 1.

Cumulative Sum (CUSUM)-based detector is a well-known method for detecting FDIAs and other cyber-attacks in smart grids. A real-time CUSUM-based detection algorithm was proposed for FDIA and jamming attacks in [10], where the probability density functions of measurements are modeled using Gaussian distribution before and after the attack, and the unknown attack parameters were obtained using maximum likelihood estimation. Due to the exponential increase in calculation burden with the number of measurements, a relaxed generalized CUSUM was presented to detect FDIAs in power systems [11]. The computational complexity of this method scales linearly with the measurement dimension. To address the large calculation burden of the centralized CUSUM algorithm, a consensus-based distributed implementation of the generalized CUSUM was designed, requiring only local communications [12]. In addition to conventional FDIA testing, a coordinated cyber-physical attack was constructed in power systems, and an adaptive nonparametric CUSUM was designed to detect both coordinated attacks and FDIAs simultaneously [13]. Furthermore, a novel normalized Rao-CUSUM detection scheme was developed [14], enabling the proposed algorithm to distinguish FDIAs from sudden changes in power systems. Finally, other cyber-attack detections, such as replay attacks, have also been conducted using the CUSUM test [15].

Compared with the CUSUM-based detector, the machine learning-based detector for FDIAs is a promising technique in smart grids. A Reinforcement Learning (RL) - based online cyber-attack detection algorithm was formulated in [16]. The detection problem is modeled as a partially observable Markov decision process (POMDP), with an optimization objective to minimize detection delay and false alarm rate. Attack strategies including FDIA, jamming, denial-of-service, and network topology attacks were modeled. Although the method in [16] is relatively simple, it provides an important foundation for RL-based cyber-attack detection research. Inspired by [16], a Deep Reinforcement Learning (DRL)-based FDIA detection method was proposed [17], where both the process and measurement covariance matrices are unknown. The proposed detection algorithm also explained how to guarantee the effectiveness of the detection scheme by selecting appropriate DRL parameters. Based on the data integrity attack in [16], FDIAs in unbalanced distribution networks were further studied in [18, 19]. Various other machine learning-based detection methods have also been developed, including federated deep learning [20], Kalman filter and recurrent neural networks [9], and wavelet transform combined with deep neural networks [21]. The aforementioned FDIAs primarily focus on adding deviation vectors to remain undetected, without deeper design considerations for attack capacity construction.

Inspired by the concept of FDIAs, blind FDIAs have been developed in power systems [22], utilizing subspace estimation and matrix reconstruction. However, a corresponding detection algorithm was not proposed. Furthermore, due to the vulnerability of data-driven detection technologies, adversarial examples of FDIAs have emerged, prompting the development of DRL-based robust detectors [23, 24]. Both Kalman Filter (KF) and Cubature KF (CKF) were used in the designed detection scheme. The proposed detector can robustly detect both FDIAs and their adversarial examples simultaneously. However, the aforementioned attack vectors did not evaluate their destructive potential and focused solely on success rate and stealthiness. Additionally, a comparative study of various FDIAs is summarized in Table I. As illustrated in Table I, existing FDIA studies either overlook attack stealth or does not propose the load attack range and efficient detection methods.

TABLE I
COMPARATIVE STUDY OF DIFFERENT FALSE DATA INJECTION ATTACKS

Methods	Problem	Algorithm	Contributions	Limitations
[16]	An online attack detection using the RL	RL-based algorithm	The detection is modeled as a POMDP, and a robust detector is designed	The stealthy of cyber-attack is not considered
[17]	A DRL-based discontinuous attack detection	DRL-based algorithm	Design the parameters of RL to guarantee the validity of the detection	The stealthy of cyber-attack is not ensured
[18]	FDIA detection for unbalanced distribution networks	Generalized likelihood ratio test	Design a square-root unscented Kalman filter based state estimator	Parameter adjustment is more complicated for the detector
[20]	Detection of FDIA in smart grids	Federated deep learning approach	Propose a secure federated learning scheme by combing Paillier cryptosystem	No further design for the attack capacity
[22]	Blind FDIA approach against the state estimation	Matrix reconstruction	Perform high successful rate of FDIA when measurement data is very limited	No study or design on attack detection method
[23,24]	Robust data-driven attack detection algorithm	DRL-based algorithm	The FDIA and its adversarial example are simultaneously and robustly detected	No report on the potential attack capability for attackers

Accordingly, the motivation of this paper aims to construct extended attack vectors based on traditional FDIAs, not only to further exploit attack capability but also to maintain high stealth against bad data detection (BDD). Second, due to the limitations of traditional residual-based detectors, a novel DRL-based detector is designed to improve the detection performance against highly stealthy cyber-attacks. The key contributions of this paper are threefold: (1) attack range estimation of the extended FDIA (E-FDIA), (2) construction of the stealthy E-FDIA model, and (3) development of a secure defense and a new FDIA detector. The detailed innovation points are discussed as follows.

(1) Attack range estimation of the E-FDIA. Based on the DC power systems model, the initial FDIA (I-FDIA) aims to remain stealthy while avoiding detection by the BDD. This study proposes an E-FDIA to enhance destructive capabilities through load attacks. The attackable load range is analyzed and estimated under four scenarios, considering load spinning reserve.

(2) Construction of stealthy E-FDIA model. Directly adding the extended component to measurements is easily detected by the BDD. To avoid this, E-FDIA integrates the component into state variables, enhancing stealth. Accordingly, an optimization model for E-FDIA is developed to enhance stealth against both neural attack detection (NAD) schemes and BDD. The proposed E-FDIA model incorporates constraints derived from system measurements (I-FDIA and E-FDIA), extended attack components, and the output actions of the DRL-based approach.

(3) Secure defense framework and new FDIA detector. Given the limitations of conventional detectors [16], [23], [25]-[27], a Deep Q-Network (DQN)-based safe defense framework and its detection algorithm implementation are proposed, including training phase and online detection phase. A compound reward is designed to minimize false alarm rate and detection delays, incorporating a relative cost factor c_β. Furthermore, an extended Euclidean distance indicator and an adaptive weight matrix are introduced in the dynamic state estimation to enhance detection accuracy within the DRL framework. Experimental results on IEEE-14, 30, 39, and 118 bus systems confirm the effectiveness and robustness of the proposed detector.

The remainder of the paper is organized as follows. Section II describes the novel stealthy attack strategy. Section III and IV establish detection metrics and DRL-based detection scheme, and simulation results demonstrate its effectiveness in Section V. Finally, the paper concludes in Section VI.

II. Novel Stealthy Attack Model

This section first introduces state estimation and I-FDIA, and then the strategy of E-FDIA is proposed, respectively.

A. State Estimation

The control center uses collected measurements to estimate the state variables, which is carried out by the state estimation (SE). Accordingly, the measurement based on AC power systems is represented by [24]

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{w}_t. \tag{1}$$

where $h(.)$ represents a non-linear function from state variables to measurements. $\mathbf{w}_t = [w_{1,t}, \ldots, w_{M,t}]^{\mathrm{T}}$ is the measurement error, which follows a Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{R}_t)$, and $\boldsymbol{R}_t = \boldsymbol{\sigma}_w^2 \boldsymbol{I}_M$.

Since AC state estimation is time consuming in large-scale power grids, accordingly, the measurement based on DC model of power grids can be formulated as follows [24]

$$\begin{cases} \mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{v}_t, \\ \mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t, \quad \forall t \end{cases} \tag{2}$$

Here $\mathbf{v}_t = [v_{1,t}, \ldots, v_{N,t}]^{\mathrm{T}}$ denotes the process noise, which is assumed to follow a Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{Q}_t)$, and $\boldsymbol{Q}_t = \boldsymbol{\sigma}_v^2 \boldsymbol{I}_N$. The optimal solution of state variable is denoted by: $\hat{\mathbf{x}}_t = (\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{z}_t$, and the matrix $\mathbf{W} = \mathrm{diag}(\boldsymbol{\sigma}_{\mathrm{w}}^{-2})$ is a diagonal matrix. In DC power flow, since the bus voltage magnitude is set to 1.0 p.u, the state variable and measurement include phase angles and active power flows.

B. Initial FDIA and Objective

When an FDIA is launched, false data is injected to manipulate measurements while satisfying Kirchhoff's law. However, load changes generally remain within a predefined range. If this safe range is exceeded, defenders might suspect anomalies in the measurement data and trigger power flow recalibration.

Under normal conditions, the power flow P_{ij}^L is descried as [28]: $P_{ij}^L = b_{ij}(\theta_i - \theta_j)$. The injection power of the bus i is represented by: $P_i^{\mathrm{Inject}} = P_i^{\mathrm{D}} - P_i^{\mathrm{G}}$. Once the I-FDIA occurs, the power flow $P_{ij}^{L,o}$ is given by: $P_{ij}^{L,o} = b_{ij}(\theta_i^o - \theta_j^o)$. The power flow variation before and after an attack is: $\Delta P_{ij}^{L,o} = b_{ij}(\Delta\theta_i^o - \Delta\theta_j^o)$. $\Delta\theta_i^o = \theta_i^o - \theta_i$, $\Delta\theta_j^o = \theta_j^o - \theta_j$, and $\Delta P_{ij}^{L,o} = P_{ij}^{L,o} - P_{ij}^L$. Moreover, the variation of bus injection power before and after an attack is described as $\Delta P_i^{\mathrm{Inject}} = \Delta P_i^D - \Delta P_i^G$, and $\Delta P_i^D = P_i^{D,o} - P_i^D$,

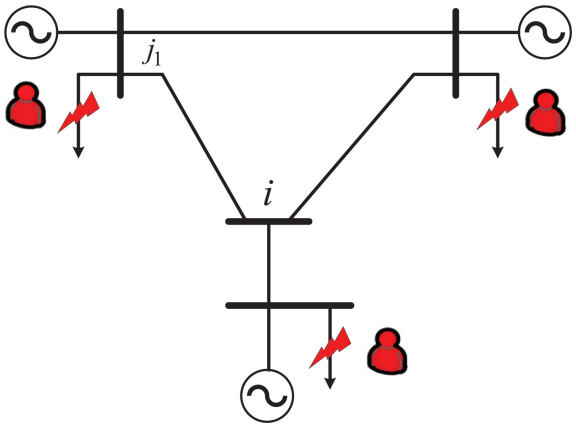


Fig. 2. FDIA launched in a four-bus power system.

and $\Delta P_i^G = P_i^{G,o} - P_i^G$. The power generation is generally well defined by the operator, which is changed only if the grid operator finds it necessary. It means $\Delta P_i^G = 0$. Accordingly, the variation of bus injection power is rewritten as $\Delta P_i^{\text{Inject}} = \Delta P_i^D$.

The topology of a four-bus power system is shown in Fig. 2, taking an attacked bus j_1 as an example (bus i is not attacked), the variation of power flow for line ij_1 is represented by.

$$\begin{aligned} &\Delta P_{ij_1}^{L,o} = -b_{ij_1}\Delta\theta_{j_1}^o \\ &\Rightarrow \Delta\theta_{j_1}^a = -b_{ij_1}^{-1}\Delta P_{ij_1}^{L,o} \end{aligned} \tag{3}$$

The initial load attack $\Delta P_{j_1}^{D,o}$ is defined as follows.

$$\Delta P_{j_1}^{D,o} = \begin{cases} \Delta P_{j_1}^{D,o}, \text{I} - \text{FDIA}, \\ 0, \ \text{Otherwise}. \end{cases} \tag{4}$$

According to the definition of I-FDIA, the initial attack vector

$$\boldsymbol{c}_o = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\Delta\boldsymbol{P}^{D,o} \tag{5}$$

Accordingly, the I-FDIA vector is described as follows

$$\begin{cases} \boldsymbol{a}_o = \mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\Delta\boldsymbol{P}^{D,o} \\ \Delta\boldsymbol{P}^{D,o} = [\Delta P_1^{D,o}, \cdots, \Delta P_{N_d}^{D,o}]^{\mathrm{T}} \end{cases} \tag{6}$$

C. Proposed Model of Extended FDIA and Objective

The I-FDIA aims to maintain the likelihood of a successful attack, though its destructive capability may be limited. Accordingly, E-FDIA is introduced to increase destructive impact while improving stealth. Given the initial false measurement $\mathbf{z}_t^o = \mathbf{z}_t + \boldsymbol{a}_o$, the attacker seeks to launch an E-FDIA that introduces a load attack within the allowable safe range. The final compromised measurement is expressed as: $\mathbf{z}_t^e = \mathbf{z}_t^o + \boldsymbol{a}_e = \mathbf{z}_t + \boldsymbol{a}_o + \boldsymbol{a}_e$. The total injected false data is given by: $\boldsymbol{a}_t = \boldsymbol{a}_o + \boldsymbol{a}_e$. The corresponding optimization model is formulated as follows

$$\begin{aligned} \mathbf{P1}: \max_{\boldsymbol{a}_e} \ & \rho(\mathbf{z}_t^o, \mathbf{z}_t^e) \\ s.t. \ & \mathbf{z}_t^e = \mathbf{z}_t^o + \boldsymbol{a}_e, \ \mathbf{z}_t^o = \mathbf{z}_t + \boldsymbol{a}_o \\ & \hat{\mathbf{y}}(\mathbf{z}_t + \boldsymbol{a}_o) = \hat{\mathbf{y}}(\mathbf{z}_t + \boldsymbol{a}_o + \boldsymbol{a}_e), \quad \forall t \end{aligned} \tag{7}$$

The second constraint ensures that the residuals of I-FDIA and E-FDIA are equal, indicating that both remain stealthy. However, this optimization problem is incomplete, as the load attack range has not yet been explicitly defined.

The initial attack vector is calculated as $\boldsymbol{a}_o = \mathbf{H}\boldsymbol{c}_o$. To remain undetectable by both BDD and NAD, the E-FDIA embeds the extended component into the state variables, thereby expanding

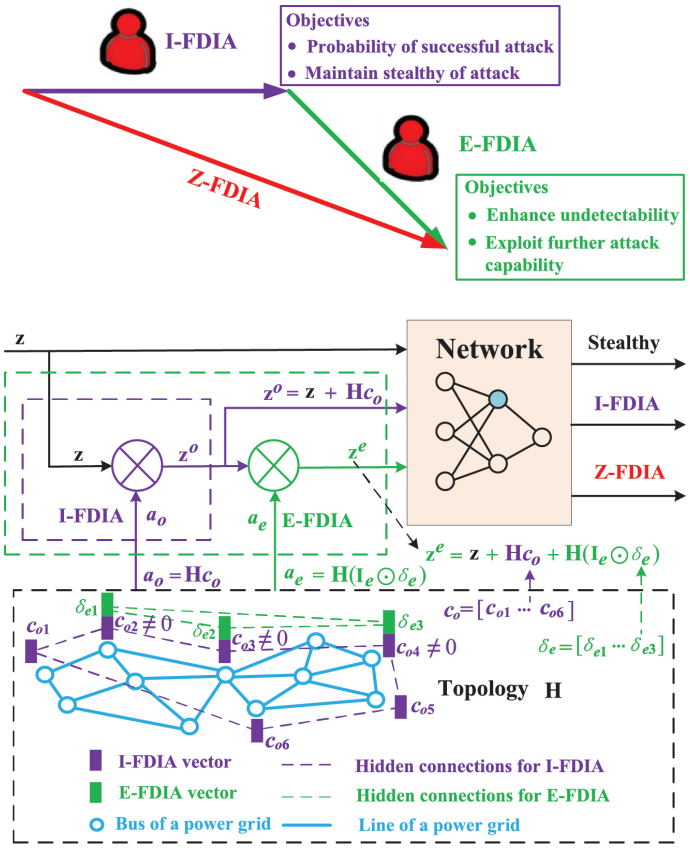


Fig. 3. Relationship between I-FDIA and E-FDIA.

the load attack range. First, a column vector $\mathbf{I}_e$ is defined as follows

$$(\mathbf{I}_e)_i = \begin{cases} 1, \ \boldsymbol{c}_{o,i} \neq 0, \\ 0, \ \boldsymbol{c}_{o,i} = 0, \quad \forall i \end{cases} \tag{8}$$

The optimization problem in Eq. (7) can be represented by

$$\begin{aligned} \mathbf{P2}: \max_{\boldsymbol{\delta}_e} \ & ||\mathbf{I}_e \odot \boldsymbol{\delta}_e||_2 \\ s.t. \ & \mathbf{z}_t^e = \mathbf{z}_t^o + \boldsymbol{a}_e, \ \mathbf{z}_t^o = \mathbf{z}_t + \boldsymbol{a}_o \\ & \boldsymbol{a}_e = \mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e) \\ & \hat{\mathbf{y}}(\mathbf{z}_t + \boldsymbol{a}_o) = \hat{\mathbf{y}}(\mathbf{z}_t + \boldsymbol{a}_o + \boldsymbol{a}_e), \quad \forall t \end{aligned} \tag{9}$$

where $\odot$ denotes the Hadamard product operation. The total injected false data $\boldsymbol{a}_t = \boldsymbol{a}_o + \boldsymbol{a}_e = \mathbf{H}\boldsymbol{c}_o + \mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e) = \mathbf{H}(c_o + \mathbf{I}_e \odot \boldsymbol{\delta}_e)$ does not generate new residuals, ensuring the final attack remains undetectable by the BDD. Furthermore, according to [24], the stealthiness of the final injected data with respect to NAD is guaranteed by the feasible solutions of Eq. (9). As a result, a deeply stealthy FDIA capable of causing greater load loss in power systems is constructed.

Generally, the variation range of the load is defined as follows

$$\begin{cases} \hat{P}_{j_1}^D = P_{j_1}^D + \Delta P_{j_1}^D \\ P_{j_1}^{D,\min} \leq \hat{P}_{j_1}^D \leq P_{j_1}^{D,\max} \end{cases} \tag{10}$$

Similarly, the extended load attack $\Delta P_{j_1}^{D,e}$ is represented by

$$\Delta P_{j_1}^{D,e} = \begin{cases} \Delta P_{j_1}^{D,e}, \text{E} - \text{FDIA}, \\ 0, \ \text{Otherwise}. \end{cases} \tag{11}$$

Fig. 3 illustrates the relationship between E-FDIA and I-FDIA. Z-FDIA denotes the combined attack composed of both I-FDIA and E-FDIA, resulting in the total destructive impact on the power grid. The choice of attack direction and magnitude is also critical.

General speaking, if the actual load exceeds the predicted load $P_{j_1}^{D,f}$, positive spinning reserves (SRs) are dispatched. Otherwise, negative SRs are used. From the attacker's perspective, the ideal strategy is to maximize damage while minimizing the cost. The E-FDIA follows the same attack direction as the I-FDIA, with the attack magnitude described in the following four cases.

(1) The initial load attack exceeds the predicted load: $P_{j_1}^{D,o} \geq P_{j_1}^{D,f}$. Positive SRs of power grids are required, and $\Delta P_{j_1}^{D,o} = P_{j_1}^{D,o} - P_{j_1}^{D,f}$. To further widen the load gap or increase positive SR demand, the attacker continues the attack in the same direction, with $\Delta P_{j_1}^{D,e} = P_{j_1}^{D,\max} - P_{j_1}^{D,o}$.

(2) The initial load attack is less than the predicted load $0 \leq P_{j_1}^{D,o} \leq P_{j_1}^{D,f}$, and negative SRs are required. The initial attack vector is described as: $\Delta P_{j_1}^{D,o} = P_{j_1}^{D,f} - P_{j_1}^{D,o}$. To increase the power gap and demand for negative SRs, the attacker continues the attack in the same direction, resulting in $\Delta P_{j_1}^{D,e} = P_{j_1}^{D,o} - P_{j_1}^{D,\min}$.

(3) When the initial attacks satisfy: $P_{j_1}^{D,o} \leq P_{j_1}^{D,f} \leq 0$, the positive SRs occur. The initial attack vector is given by: $\Delta P_{j_1}^{D,o} = P_{j_1}^{D,o} - P_{j_1}^{D,f}$, then $\Delta P_{j_1}^{D,e} = P_{j_1}^{D,\min} - P_{j_1}^{D,o}$.

(4) When the initial load attack satisfies $P_{j_1}^{D,f} \leq P_{j_1}^{D,o} \leq 0$, the negative SRs of power grids are required. The initial attack vector $\Delta P_{j_1}^{D,o} = P_{j_1}^{D,f} - P_{j_1}^{D,o}$, then $\Delta P_{j_1}^{D,e} = P_{j_1}^{D,o} - P_{j_1}^{D,\max}$.

Accordingly, the stealthy vectors of I-FDIA and E-FDIA can be described as follows

$$\begin{cases} \boldsymbol{c}_o = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\Delta\boldsymbol{P}^{D,o},\ \boldsymbol{a}_o = \mathbf{H}\boldsymbol{c}_o \\ \Delta\boldsymbol{P}^{D,o} = \left[\Delta P_1^{D,o}, \cdots, \Delta P_{N_d}^{D,o}\right] \\ \boldsymbol{c}_e = (\mathbf{I}_e \odot \delta_e), \boldsymbol{a}_e = \mathbf{H}(\mathbf{I}_e \odot \delta_e) \\ ||\mathbf{I}_e \odot \delta_e||_2 \leq ||\Delta\boldsymbol{P}^{D,e}||_2 \\ \Delta\boldsymbol{P}^{D,e} = \left[\Delta P_1^{D,e}, \cdots, \Delta P_{N_d}^{D,e}\right] \end{cases} \tag{12}$$

The residual between the measurement and estimated data under normal conditions is denoted by

$$r = ||\mathbf{z}_t - \mathbf{H}\mathbf{x}_t||_2 < \tau_s \tag{13}$$

where τ_s is a predefined threshold determined using a chi-square distribution $\chi^2_{1-\varrho}$, where ϱ represents the significance level.

Moreover, the residual under I-FDIA condition is represented by

$$\begin{aligned} r_o &= ||\mathbf{z}_t^o - \mathbf{H}\hat{\mathbf{x}}_t^o||_2 = ||\mathbf{z}_t + \boldsymbol{a}_o - \mathbf{H}(\hat{\mathbf{x}}_t + \boldsymbol{c}_o)||_2 \\ &= ||\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_t + \boldsymbol{a}_o - \mathbf{H}\boldsymbol{c}_o)||_2 = r < \tau_s \end{aligned} \tag{14}$$

Due to the residual $r_o = r$, accordingly, the I-FDIA can remain stealthy to the BDD.

E-FDIA is launched through injecting a vector $\mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e)$ into the state variable, and the residual is described as

$$\begin{aligned} r_e &= ||\mathbf{z}_t^e - \mathbf{H}\hat{\mathbf{x}}_t^e||_2 = ||\mathbf{z}_t^o + \boldsymbol{a}_e - \mathbf{H}(\hat{\mathbf{x}}_t^o + \mathbf{I}_e \odot \boldsymbol{\delta}_e)||_2 \\ &= ||\mathbf{z}_t + \boldsymbol{a}_o - \mathbf{H}(\hat{\mathbf{x}}_t + \boldsymbol{c}_o)||_2 = r_o \\ &= ||\mathbf{z}_t - \mathbf{H}\mathbf{x}_t||_2 = r < \tau_s \end{aligned} \tag{15}$$

According to Eq. (15), the attack vector $\boldsymbol{a}_e$ of E-FDIA can deceive the BDD. E-FDIA also qualifies as a highly stealthy attack. In E-FDIA modeling, attackers must first determine load attack range estimation through I-FDIA and subsequently solve the extended component optimization model. Compared with I-FDIA, E-FDIA exhibits enhanced stealth properties, thereby demanding more resource allocation.

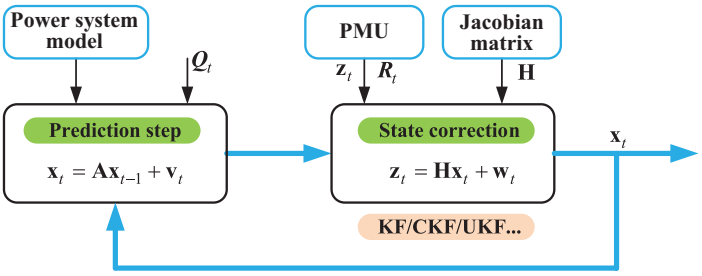


Fig. 4. Implementation flow chart of dynamic SE.

D. Dynamic State Estimation

The KF functions as an online estimator in SE, incorporating both prediction and correction steps for improved accuracy. Various KF variants are employed in this study for measurement estimation, including the CKF, and UKF. The flowchart illustrating the implementation of dynamic SE is presented in Fig. 4. Here, the KF equations at slot t are presented as follows

1) Prediction Step

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1} \\ \mathbf{F}_{t|t-1} &= \mathbf{A}\mathbf{F}_{t-1|t-1}\mathbf{A}^{\mathrm{T}} + \boldsymbol{Q}_t \end{aligned} \tag{16}$$

2) Update Step

$$\begin{aligned} \boldsymbol{K}_t &= \mathbf{F}_{t|t-1}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{F}_{t|t-1}\mathbf{H}^{\mathrm{T}} + \boldsymbol{R}_t)^{-1} \\ \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \boldsymbol{K}_t(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \\ \mathbf{F}_{t|t} &= \mathbf{F}_{t|t-1} - \boldsymbol{K}_t\mathbf{H}\mathbf{F}_{t|t-1} \end{aligned} \tag{17}$$

Here, the basic KF underpins all other mentioned KF methods. In this study, CKF, SRCKF and UKF, are applied to realize the measurement estimation.

III. DETECTION METRICS OF FDIA

The conventional metrics of the detection algorithm is first introduced. Then the novel metric is proposed.

A. Conventional Metrics

Conventional RL-based detection methods characterize the operating state of power systems using the residual between measured and estimated data obtained via the KF, as proposed in [16, 23]. In [16], the residual-based estimation metric denoted by Υ_t is described as: $\Upsilon_t = (\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t})^{\mathrm{T}}(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t})$. A smaller value of Υ_t indicates that the power system is operating under normal conditions. Conversely, a larger Υ_t value suggests the presence of a cyber-attack. However, stealthy FDIAs cannot be detected solely based on the residual between measured and estimated data. This is because attackers deliberately design the false data to bypass the residual threshold of the BDD. Therefore, to more accurately capture the system's operating state, a cosine similarity-based estimation metric was introduced in [25].

The predicted state variables at time t are modeled in Eq. (16) and (17), and the cosine similarity estimation metric is given by

$$\varphi_t = \frac{\mathbf{z}_{t|t-1} \cdot \hat{\mathbf{z}}_{t|t-1}}{\|\mathbf{z}_{t|t-1}\| \cdot \|\hat{\mathbf{z}}_{t|t-1}\|}, t = 1, 2, \ldots T \tag{18}$$

where $0 \leq \varphi_t \leq 1$. If there is no attack in the power systems, the metric φ_t is nearly equal to 1.

B. Proposed Metrics

In this paper, an extended version of the Euclidean similarity metric is established in Eq. (19).

$$\eta_t = 1 - \frac{||\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2},\ t = 1, 2, ...T \tag{19}$$

Here η_t represents the ratio of the Euclidean similarity at time t. If there is no attack, the metric η_t is nearly equal to zero.

The proposed Euclidean similarity at initial moment can be described as follows: $\eta_1 = 1 - \frac{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2} = 0$. Accordingly, if there is no FDIA, then the actual and estimated measurements exactly match. The value of η_1 is equal to zero denoting there is no FDIA. If an FDIA occurs in the power grids at time t, the actual measurement $\mathbf{z}_t$ is immediately changed due to the injection of FDIA and described as follows: $\mathbf{z}_t^a = \mathbf{z}_t + \boldsymbol{a} = \mathbf{z}_t + \boldsymbol{a}_o + \boldsymbol{a}_e$. The extended euclidean similarity is given by

$$\eta_t^a = 1 - \frac{||\mathbf{z}_t^a - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2} = 1 - \frac{||(\mathbf{z}_t + \boldsymbol{a}) - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2}. \tag{20}$$

Before time t, the power grids operate under normal conditions, and the estimated state vector and measurement $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{z}}_{t|t-1}$ are the normal state and prediction, thereby there is no large spike for Euclidean similarity. However, the value of η_t^a sharply deviates from 0, then the stealthy FDIA is detected.

According to the dynamic state estimator shown in Eq. (17), the estimated state variable is described as follows [29].

$$\hat{\mathbf{x}}_t = \mathbf{A}\hat{\mathbf{x}}_{t-1} + \boldsymbol{K}_t(\mathbf{z}_t + \boldsymbol{a} - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{21}$$

The estimated state after FDIAs at time t is described as $\hat{\mathbf{x}}_t^a = \hat{\mathbf{x}}_t + \boldsymbol{K}_t\boldsymbol{a}_t$. The estimated state at the time $t+1$ can be represented by

$$\begin{aligned}\hat{\mathbf{x}}_{t+1}^a &= \mathbf{A}\hat{\mathbf{x}}_t^a + \boldsymbol{K}_{t+1}(\mathbf{z}_{t+1}^a - \mathbf{H}\hat{\mathbf{x}}_{t+1}^a) \\ &= \hat{\mathbf{x}}_{t+1} + \mathbf{A}\Delta\hat{\mathbf{x}}_t^a - \boldsymbol{K}_{t+1}\mathbf{H}\mathbf{A}\Delta\hat{\mathbf{x}}_t^a \\ &\quad + \boldsymbol{K}_{t+1}\boldsymbol{a}_{t+1}\end{aligned} \tag{22}$$

where $\Delta\hat{\mathbf{x}}_t^a = \hat{\mathbf{x}}_t^a - \hat{\mathbf{x}}_t$.

Accordingly, the injected bias of $\hat{\mathbf{x}}_t^a$ is given by

$$\begin{aligned}&\hat{\mathbf{x}}_{t+1}^a - \hat{\mathbf{x}}_{t+1} = \mathbf{A}\Delta\hat{\mathbf{x}}_t^a - \boldsymbol{K}_{t+1}\mathbf{H}\mathbf{A}\Delta\hat{\mathbf{x}}_t^a + \boldsymbol{K}_{t+1}\boldsymbol{a}_{t+1} \\ &\Longrightarrow \Delta\hat{\mathbf{x}}_{t+1}^a = (\mathbf{A} - \boldsymbol{K}_{t+1}\mathbf{H}\mathbf{A})\Delta\hat{\mathbf{x}}_t^a \\ &\qquad\qquad + \boldsymbol{K}_{t+1}\boldsymbol{a}_{t+1}\end{aligned} \tag{23}$$

As shown in Eq. (23), the injected bias at time $t+1$ consists of both the accumulated deviations from previous steps and the newly injected false data. During an FDIA, the historical estimated data is adjusted during the dynamic estimation of predicted measurements. To mitigate the influence of injected deviations during dynamic estimation, the Kalman gain is adaptively adjusted. When considering the injected deviation, the noise covariance $\boldsymbol{R}_t$ can be adaptively reduced to adjust the previously accumulated deviations. Consequently, a higher weighting factor is applied in the measurement prediction step. Initially, the adaptive weight coefficient matrix C_{nov} is defined as follows

$$\mathbf{C}_{\text{nov}} = \begin{bmatrix} e^{-|\mathbf{z}_{1t} - \hat{\mathbf{z}}_{1t|t-1}|} & & \\ & \ddots & \\ & & e^{-|\mathbf{z}_{Mt} - \hat{\mathbf{z}}_{Mt|t-1}|} \end{bmatrix} \tag{24}$$

Here, in original dynamic estimator, the error covariance matrix $\boldsymbol{R}_t = \text{diag}(\sigma_{w,11}^2, \ldots, \sigma_{w,MM}^2)$. Accordingly, the improved noise covariance $\boldsymbol{R}_{\text{nov}}$ is given by

$$\begin{aligned}\boldsymbol{R}_{\text{nov}} &= \boldsymbol{R}_t \cdot \mathbf{C}_{\text{nov}} \\ &= \begin{bmatrix} \sigma_{w,11}^2 e^{-|\mathbf{z}_{1t} - \hat{\mathbf{z}}_{1t|t-1}|} & & \\ & \ddots & \\ & & \sigma_{w,MM}^2 e^{-|\mathbf{z}_{Mt} - \hat{\mathbf{z}}_{Mt|t-1}|} \end{bmatrix}\end{aligned} \tag{25}$$

Therefore, in the dynamic state estimation, the Kalman gain $\boldsymbol{K}_t$ is updated using $\boldsymbol{R}_{\text{nov}}$: $\boldsymbol{K}_t = \mathbf{F}_{t|t-1}\mathbf{H}^{\text{T}}(\mathbf{H}\mathbf{F}_{t|t-1}\mathbf{H}^{\text{T}} + \boldsymbol{R}_{\text{nov}})^{-1}$, then the updated measurements based on the dynamic estimator: $\hat{\mathbf{z}}_{t|t-1} = \mathbf{H}\mathbf{A}\hat{\mathbf{x}}_{t-1} + \mathbf{H}\boldsymbol{K}_t(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})$.

IV. DRL-BASED EXTENDED FDIA DETECTION

The state and action space of detection scheme are defined, and an DQN-based online detection scheme is presented.

A. State and Action Space

Online FDIA detection is formulated as a POMDP problem because the change point from normal to attack state, as well as the transition probabilities, are unknown. Generally, a POMDP is represented by the quintuple ($\mathcal{S}$,$\mathcal{A}$,$\mathcal{F}$,$\mathcal{R}$, $\mathcal{G}$,γ), defined as follows.

1) $\mathcal{S}$ represents the set of all hidden states. Given a state variable s_t of the power systems, then the $s_t \in \mathcal{S}$.

2) $\mathcal{A}$ consists of all possible actions. Given a particular active a_t, and the active $a_t \in \mathcal{A}$.

3) $\mathcal{F}$ represents the set of conditional transition probabilities between all hidden states. $\mathcal{F}(s_{t+1|s_t,a_t})$ describes the transition from s_t to the s_{t+1}.

4) $\mathcal{R}$ represents the set of all rewards. Based on the state s_t and action a_t at the time t, the reward is expressed as $r_t \in \mathcal{R}(s_t, a_t)$.

5) $\mathcal{G}$ represents the set of conditional observation probability.

6) γ represents a discount factor.

The exact launch time of an FDIA, represented by τ, is unknown. Additionally, both the FDIA strategy and the state transition probabilities remain unknown. Accordingly, the pre-attack and post-attack are modeled as two hidden states in the online detection process. The former represents normal system operation, while the latter corresponds to abnormal (attacked) conditions. Upon receiving power system measurements $\mathbf{z}_t$, the defender chooses between two actions: *stop* and *continue*. If the *stop* action is taken, an attack is declared. Otherwise, the detection process continues.

B. Online FDIA Detection Based on Reinforcement Learning

Given the current state of the power, a state $s_t \in \mathcal{S}$ is observed. Upon taking an action $a_t \in \mathcal{A}$, the defender receives a reward $r_t \in \mathcal{R}(s_t, a_t)$ from the power system. The next state s_{t+1} is then predicted with transition probability $f \in \mathcal{G}$. This process continues until a predefined termination condition is met. The cumulative rewards can be described as the sum of all discounted rewards.

$$G_t = r_{t+1} + \gamma r_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}, \tag{26}$$

where $\gamma \in [0, 1]$ is a discount factor, indicating the relative valuation of immediate versus future rewards.

Based on [30], the POMDP problem of online FDIA detection in power grids is characterized as follows.

$$\min_{\pi:\mathcal{S}\to\mathcal{A}} \ \mathbb{E}\left[G(t)\right] \tag{27}$$

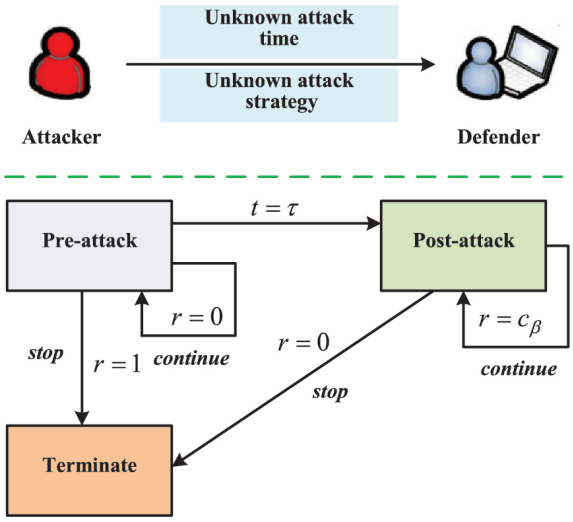


Fig. 5. Transitions between the hidden states.

In the proposed stealthy attack detection framework, the objective is to detect FDIAs while minimizing both the false alarm rate and detection delay. Because the attack launch time τ is unknown, two hidden states are defined: pre-attack and post-attack. A false alarm occurs when the system is in a normal (pre-attack) state but is incorrectly classified as being under attack (post-attack). Conversely, the system may also incorrectly indicate a pre-attack state during an actual attack. Detection delay arises when the defender identifies the FDIA only after it has occurred, rather than at the exact moment of the attack. The state transition process during detection is illustrated in Fig. 5. As illustrated in Fig. 5, an FDIA is launched at $t = \tau$, causing the system state to shift from pre-attack to post-attack. If the defender chooses the action *continue*, the corresponding rewards are 0 in the pre-attack state and c_β in the post-attack state. If the defender selects the action *stop*, the respective received rewards are 1.0 and 0, respectively. The detection objective function minimizes a weighted sum of the false alarm rate and detection delay. If the detector prioritizes minimizing detection delay, a higher value of c_β (e.g., c_β=1) can be assigned. Otherwise, a smaller c_β is used to reduce the penalty on delay.

$$\begin{aligned}\mathbf{P3}: \mathbb{E}\{G_t\} &= \mathbb{E}\left\{\psi(\Gamma<\tau) + \sum\nolimits_{t=\tau}^{\Gamma} c_\beta\right\} \\ &= \mathbb{E}\{\psi(\Gamma<\tau)\} + c_\beta\, \mathbb{E}\left\{(\Gamma-\tau)^+\right\}\end{aligned} \tag{28}$$

where the first item is a false alarm event, and the second item is the time delay. Note that $(q)^+ = \max(q, 0)$.

The optimization model for online FDIA detection problem is formulated as follows.

$$\mathbf{P4}: \min_{\Gamma} \ \mathbb{E}\{\psi(\Gamma<\tau)\} + c_\beta\, \mathbb{E}\left\{(\Gamma-\tau)^+\right\} \tag{29}$$

The attack detection method described applies universally to any FDIA, including I-FDIA and E-FDIA.

C. Reinforcement Learning

The RL-based algorithm is a powerful tool for addressing attack detection in cyber physical systems, even with large-scale action spaces and state spaces. Accordingly, our proposed method for stealthy attack detection can be formulated as an RL problem aimed at minimizing the cumulative cost function. In Eq. (26) and (27), calculating the return G_t is complex. The return for all trajectories starting from the current state s is calculated, followed by the calculation of corresponding expectations. The state value function is represented as follows.

$$\begin{aligned}V(s) &= \mathbb{E}_{s,a}\left[G(t)\right] \\ &= \sum_{(s_t,a_t,\ldots)} \pi(a_t|s_t)p(s_{t+1}|s_t,a_t)G(t) \\ &= \mathbb{E}_\pi\left[r_{t+1} + \gamma V(s_{t+1})|s_{t+1}=s\right]\end{aligned} \tag{30}$$

The action-state function, $Q(s_t, a_t)$, given action a and state s, is defined in [31] as follows.

$$Q(s_t,a_t) = \mathbb{E}_\pi\left\{r_{t+1} + \gamma Q(s_{t+1},a_{t+1})|s_t=s, a_t=a\right\} \tag{31}$$

By maximizing action-state function, the optimal $Q^*(s_t, a_t)$ is determined as follows.

$$Q^*(s_t,a_t) = \mathbb{E}_\pi\left\{r_{t+1} + \gamma \max_{a'} Q(s_{t+1},a')|s_t=s, a_t=a\right\}. \tag{32}$$

Consequently, the optimal policy, based on the optimal action-state value function, is described below.

$$\pi^*(a|s) = \begin{cases}1, & a = \arg\max_{a\in\mathcal{A}} Q^*(s_t,a), \\ 0, & \text{Otherwise},\end{cases} \tag{33}$$

The $Q(s_t, a_t)$ is updated according to Eq. (31).

$$\begin{aligned}Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \\ \gamma \max_{a'\in\mathcal{A}} Q(s_{t+1},a') - Q(s_t,a_t)]\end{aligned} \tag{34}$$

D. Deep Q-Network

During DQN training, the loss function is utilized to update the parameters of the evaluation network. The loss function is generally defined as follows.

$$Loss(\theta) = \frac{1}{N_m}\sum_i \left[(Q_i^{\text{target}} - Q(s_i,a_i;\theta))^2\right] \tag{35}$$

and

$$Q_i^{\text{target}} = \begin{cases} r_i + \gamma \max_{a_i'} Q(s_i', a_i'; \theta^-), & a_i \neq 1 \\ r_i, & \text{otherwise}. \end{cases} \tag{36}$$

where (s_i, a_i, r_i, s_i') represents the ith experience.

Therefore, with respect to θ, the gradient is given by

$$v = [Q_i^{\text{target}} - Q(s_i,a_i;\theta)]\nabla_\theta Q(s,a;\theta). \tag{37}$$

where $\nabla_\theta Q(s,a;\theta)$ is the gradient descent of $Q(s,a;\theta)$.

E. Detection Algorithm Implementation

This paper proposes an E-FDIA and its corresponding detection countermeasures based on a DRL framework. As both the exact attack launch time and attack strategy are unknown, the key challenge lies in designing a DRL-based detector that can effectively detect both I-FDIA and E-FDIA. The proposed framework is detailed in Algorithms 1 and 2. The DQN-based FDIA detection algorithm generally consists of two phases: a training phase and an online detection phase, each covering both I-FDIA and E-FDIA. The available actions are *continue* and *stop*. As shown in the optimization objective Eq. (29), the defender aims to take appropriate actions in both *pre-attack* and *post-attack* underlying states, minimizing false alarm rate and detection delay. During training, a simulation environment based on the DC power system model is created to generate measurement data. The defender applies dynamic state estimation to estimate state variables and compute the

Algorithm 1: DQN-based FDIA learning phase

Input: Initialize network $Q(s,a)$ and target network $Q^{'}(s,a)$. The matrix $\mathbf{A}$ and $\mathbf{H}$.
Output: Learned policy of Q network.

for i=1 : T **do**
  while $t \leq$ max_time **do**
    Based on dynamic SE, obtain the input state $\boldsymbol{s}_t$;
    With probability ϵ select a random action a_t
    Otherwise execute $a_t = \text{argmax}_a Q(s_t, a; \theta)$;
    Sample random minibatch of $\left\{s_i, a_i, r_i, s_i^{'}\right\}$;
    if t<attack_time **then**
      At time t collect measurement vector $\mathbf{z}_t$;
      Obtain the estimated values $\hat{\mathbf{z}}_{t|t-1}$ and $\hat{\mathbf{z}}_{1|0}$;
      Calculate $\eta_t = 1 - \frac{||\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2}$;
    else
      Collect the measurements t $\mathbf{z}_t^o = \mathbf{z}_t + \boldsymbol{a}_o$;
      Calcuate extended attack vector $\mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e)$;
      Obtain $\mathbf{z}_t^e = \mathbf{z}_t + \boldsymbol{a}_o + \mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e)$;
      Calcualte the metrics of E-FDIA $\eta_t = 1 - \frac{||\mathbf{z}_t^e - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1^e - \hat{\mathbf{z}}_{1|0}||_2}$;
      Achieve the measurement noise matrix $\boldsymbol{R}_{\text{nov}}$;
      Update the Kalman gain $\boldsymbol{K}_t$: $\boldsymbol{K}_t = \mathbf{F}_{t|t-1}\mathbf{H}^{\text{T}}(\mathbf{H}\mathbf{F}_{t|t-1}\mathbf{H}^{\text{T}} + \boldsymbol{R}_{\text{nov}})^{-1}$;
      Calculate the updated measurements: $\hat{\mathbf{z}}_{t|t-1} = \mathbf{H}\mathbf{A}\hat{\mathbf{x}}_{t-1} + \mathbf{H}\boldsymbol{K}_t(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})$.
    Non-terminal s_i: $y_i = r_i + \gamma \max_{a'} Q^{'}(s_i^{'}, a^{'}; \theta)$;
    Terminal s_i: $y_i = r_i$;
    Compute the loss function as: $Loss(\theta) = \frac{1}{N_m}\sum_i \left[(Q_i^{\text{target}} - Q(s_i, a_i; \theta))^2\right]$;
    Parameter $\theta^- = \theta$ is set.

Algorithm 2: DQN-based FDIA online detection

Input: Learned Q ntwork under FDIA conditions.
Output: False alarm, detection delay, etc.

for i=1 : T **do**
  while $t \leq$ max_time **do**
    Obtain the state s_t based on SE;
    if t<attack_time **then**
      Collect the measurement $\mathbf{z}_t$;
      Obtain the estimated values $\hat{\mathbf{z}}_{t|t-1}$ and $\hat{\mathbf{z}}_{1|0}$;
      Calculate $\eta_t = 1 - \frac{||\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1 - \hat{\mathbf{z}}_{1|0}||_2}$;
    else
      Collect $\mathbf{z}_t^e = \mathbf{z}_t + \boldsymbol{a}_o + \mathbf{H}(\mathbf{I}_e \odot \boldsymbol{\delta}_e)$;
      According to the updated the measurement noise matrix $\boldsymbol{R}_{\text{nov}}$ and Kalman gain $\boldsymbol{K}_t$, calcualte the metrics $\eta_t = 1 - \frac{||\mathbf{z}_t^e - \hat{\mathbf{z}}_{t|t-1}||_2}{||\mathbf{z}_1^e - \hat{\mathbf{z}}_{1|0}||_2}$;
    Based on learned Q network, obtain an action;
    if action $\neq$ *stop* **then**
      $t \leftarrow t + 1$;
      Re-collect measurements and perform SE;
    else
      Declare an FDIA and end this cycle.

evaluation metric η_t. For t>attack_time, the measurements $\mathbf{z}_t^o$ and $\mathbf{z}_t^e$ corresponding to I-FDIA and E-FDIA, respectively, are obtained. The defender then calculates η_t and evaluates the system's current operating status. Note that η_t for I-FDIA is calculated using the measured $\mathbf{z}_t^o$. During each training iteration, the defender selects an action and receives a reward or penalty based on the current system state. After M training episodes, the defender learns a policy for detecting FDIAs. As described in Algorithm 1, $\mathbf{F}$ denotes the state covariance matrix, and $\boldsymbol{K}_t$ represents the Kalman gain. The detailed procedure for online detection is presented in Algorithm 2. The optimal action is determined using the policy learned in Algorithm 1. If the output action is *continue* based on η_t, detection continues. Otherwise, the defender declares an attack, and the detection process terminates. After completing all online detection experiments, detection performance metrics, such as false alarm rate and detection delay, are statistically evaluated.

V. Simulation Studies

The performance of the proposed DRL-based algorithm is verified against existing works. Furthermore, the impacts of model parameters and various attack scenarios on the detection are evaluated.

A. System Setup and Parameters

A comprehensive set of case studies is performed to validate the effectiveness of the proposed FDIAs and their detection schemes. The experiments are conducted on the IEEE-14, 30, 39, and 118 bus power systems, based on the DC model of power systems, the state variables denoting the phase angles, can be obtained for case-14, 30, 39, and 118 in MATPOWER, respectively. As illustrated in the SE, the system matrix $\mathbf{A}$ and measurement matrix $\mathbf{H}$ are selected to be an identity matrix and derived from the power flow calculations. Moreover, the parameters of noise variances are chosen as follows: $\sigma_v^2 = 10^{-4}$, and $\sigma_w^2 = 2 \times 10^{-4}$, and $c_\beta = 0.02$ as noted in [16, 23]. For the variation range of load, it can calculated as: $\hat{P}_{j_1,\min}^D = \lambda_{\min} P_{j_1}^D$, and $\hat{P}_{j_1,\max}^D = \lambda_{\max} P_{j_1}^D$, where $\lambda_{\max} = 1.15$ and $\lambda_{\min} = 0.85$. The episodes of trained and tested phases are 1.0×10^4. The launch time of attack follows a geometric random variable with the parameter ρ, and ρ represents an uniform random variable in $[10^{-2}, 10^{-1}]$. The hyperparameters of stealthy FDIA detection algorithm are detailed in [23]. All simulations have been conducted in MATLAB 2021b on a system configured with an Intel® Core™ i9-11900K processor (11th Gen, 3.5 GHz).

B. Performance Comparisons of Different Methods

The performance of the presented DRL-based detector is validated and compared with the other detection schemes. The original metrics, such as precision, recall, F-score, false alarm, and detection delay, have been proposed in [16, 23]. Here, another metric, called Ratio-at-time (RAT), presents that the defender can immediately detect the attack/ anomaly events once an FDIA is launched.

$$\text{Ratio-at-time} = \frac{\#trials[(\Gamma == \tau)]}{\#Total\text{-}trials} \tag{38}$$

where $\#trials[(\Gamma == \tau)]$ represents the number that the defenders immediately detect attacks when the FDIAs are just launched. $\#Total\text{-}trials$ represents the total number of trials. This metric does not consider data transmission time in the simulation.

The DRL-based attack detection is also proposed. A random noise attack is performed to verify the availability of proposed scheme (trained and tested under c_β=0.02). A small amplitude vector is assumed and initialized as ϕ = (0.001+ 0.001*rand(M,1)). Subsequently, the attack vector is modeled

TABLE II
DETECTION RESULTS OF I-FDIA USING DIFFERENT METHODS

Methods	Time delay (s)	Rate-at-time	False alarm rate	Precision	Recall	F-score
RBA-based detector [16, 23]	0.00	0.00%	n.a.	n.a.	0.00	n.a.
EDBD-based detector	0.0818	95.19%	0.25%	0.9975	0.9998	0.9986
CSBD-based detector [25]	0.2358	79.28%	0.10%	0.9990	0.9999	0.9994
SRA-based detector [26]	0.1116	93.34%	0.00%	1.000	0.9999	0.9999
Proposed method	0.0152	98.56%	0.00%	1.000	1.000	1.000

TABLE III
DETECTION RESULTS OF E-FDIA USING DIFFERENT METHODS

Methods	Time delay (s)	Rate-at-time	False alarm rate	Precision	Recall	F-score
RBA-based detector [16, 23]	0.00	0.00%	n.a.	n.a.	0.00	n.a.
EDBD-based detector	0.0576	96.20%	0.11%	0.9989	0.9998	0.9993
CSBD-based detector [25]	0.1078	90.66%	0.07%	0.9993	0.9999	0.9995
SRA-based detector [26]	0.1085	93.70%	0.00%	1.000	1.000	1.000
Proposed method	0.0141	98.59%	0.00%	1.000	1.000	1.000

as a multivariate Gaussian distribution $\mathcal{N}(\text{zeros}(M,1), \text{diag}(\phi))$. Here, diag$(\cdot)$ denotes a diagonal matrix. The achieved RAT and detection delay are 99.36% and 0.0064. There is no false alarm rate. Focusing on the detectors, the proposed detection method is compared with the existing benchmark algorithms as follows.

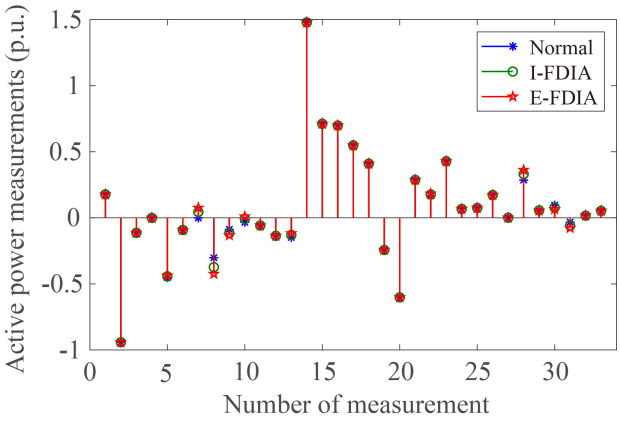
1) Residual-based attack (RBA) detectors [16, 23]. Both the SARSA and DRL were used for detecting the random attack.

2) Cosine similarity-based detection (CSBD) scheme [25]. The cosine similarity metric and Chi-square detector are used for the detection of smart grids, and the KF was adopted.

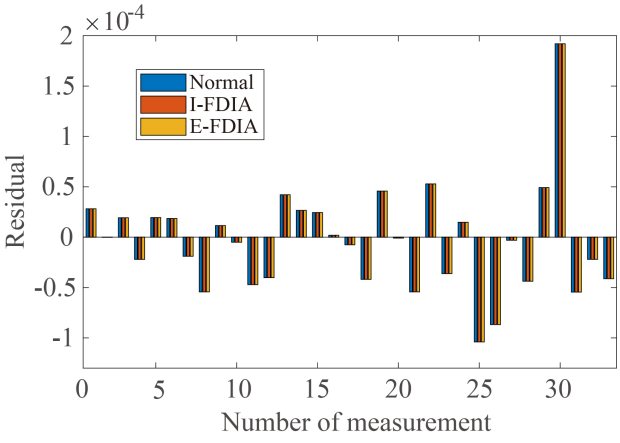
3) State residual analysis (SRA) method is designed to detect FDIA in power systems [26].

4) Euclidean distance-based detection (EDBD) indicator, and the KF is used in dynamic estimation.

During training and testing phases, it consists of 10,000 trials each, false alarm rates and average detection delays are calculated for the existing detectors and the proposed detection scheme on the IEEE-14 bus system. Detection results for I-FDIA and E-FDIA, including metrics such as precision, recall, F-score, false alarm rate, and time delay, are reported in Tables II and III. Here, the attacked buses are randomly selected from all buses, excluding slack and generator buses. Fig. 6 illustrates the measurement and residual results under normal, I-FDIA and E-FDIA conditions ($\lambda_{\max} = 1.05$ and $\lambda_{\min} = 0.95$). According to Tables II and III, traditional residual-based attack detectors fail to identify any attack vectors because the stealthy attacks are designed to maintain the same residuals before and after attack, which are shown in Fig. 6. Furthermore, the proposed DRL-based detector outperforms existing detection schemes for I-FDIA and E-FDIA. This obtained superior performance is attributed to the proposed detector's ability to distinguish between noise and persistent attacks in smart grids, facilitated by enhanced noise covariance. Additionally, the proposed similarity metrics effectively highlight the distinctions between high-level noise and actual stealthy attacks.



(a)



(b)

Fig. 6. Results for normal, I-FDIA and E-FDIA: (a) measurement, (b) residual. Among all 33 measurements, the first 13 correspond to bus real power injections, while the remaining 20 represent branch real power flows.

C. Results and Analysis

1) Performance using Different Filters

Initially, KF, CKF, SR-CKF, and UKF are employed in the detectors tested on the IEEE-118 bus power system, with results

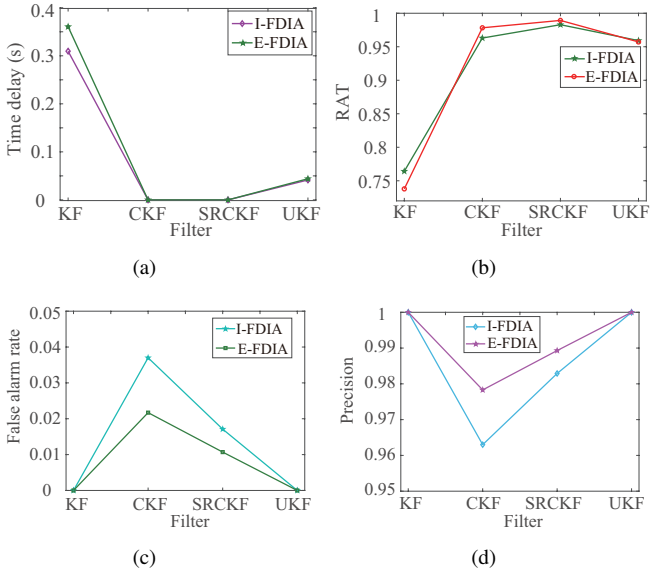


Fig. 7. Detection results using different filter: (a) detection delay, (b) RAT, (c) false alarm rate, (d) precision.

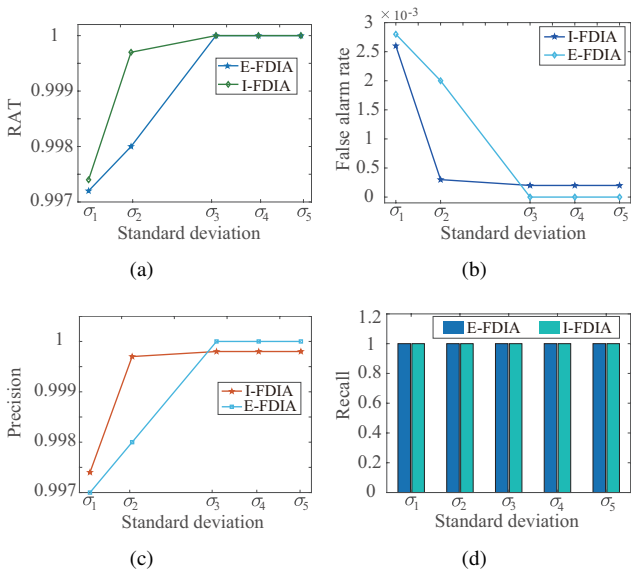


Fig. 9. Detection results using different standard deviation: (a) RAT, (b) false alarm rate, (c) precision, (d) recall.

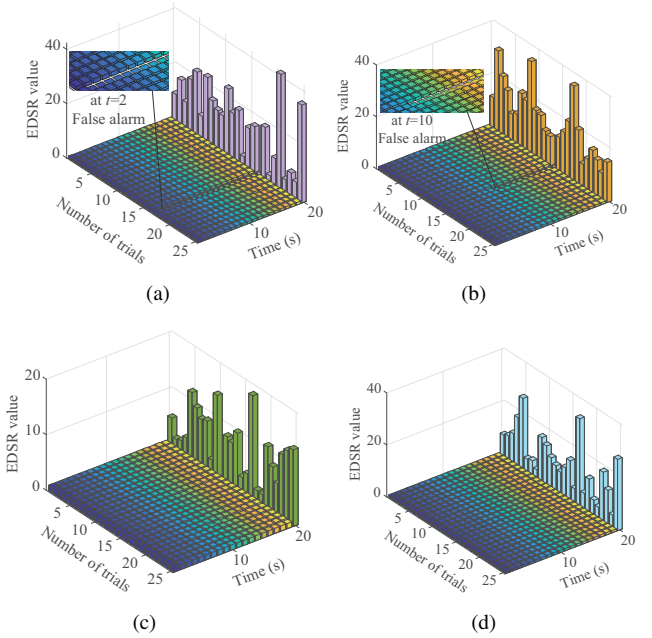


Fig. 8. EDSR results using different standard deviation: (a) $\sigma_1^2 = 0.2 \times 10^{-4}$, (b) $\sigma_2^2 = 0.6 \times 10^{-4}$, (c) $\sigma_3^2 = 1.4 \times 10^{-4}$, (d) $\sigma_4^2 = 1.8 \times 10^{-4}$.

displayed in Fig. 7. According to Fig. 7, the KF-based detector exhibits the poorest performance, with detection delays and false alarm rates of 0.3096 and 0.00% for I-FDIA, and 0.3606 and 0.00% for E-FDIA, respectively. However, false alarms were observed with CKF and SR-CKF- based detectors, registering false alarm rates of 3.7% and 1.71% for I-FDIA, respectively. The reason is that the defender cannot well differentiate high-level noises from real attacks. In our proposed detectors, the time delays and false alarm rates were 0.0415 and 0.00% for I-FDIA, and 0.044 and 0.00% for E-FDIA, respectively. Therefore, filter selection varies based on the power system's requirements for real-time performance and detection accuracy.

2) Performance of standard deviation of covariance matrix

According to Eq. (25), the proposed detector's performance in dynamic estimation correlates with the standard deviation of the measurements. How the measurement standard deviation affects the detection performance under E-FDIA condition is analyzed on the IEEE-39 bus system. Fig. 8 illustrates the detection indicator from the initial moment up to the launch of E-FDIA, with four set standard deviations, $\sigma_1^2 = 0.2 \times 10^{-4}$, $\sigma_2^2 = 0.6 \times 10^{-4}$, $\sigma_3^2 = 1.4 \times 10^{-4}$ and $\sigma_4^2 = 1.8 \times 10^{-4}$, respectively. Since the Euclidian distance similarity ratio (EDSR) value is negative, whose value at time $t = 20$ is presented as its absolute. In Fig. 8(a), the EDSR value abruptly drops to 0, generating a false alarm at $t = 2$, no stealth E-FDIA occurs, yet the defenders mistakenly believe a FDIA have been detected. Increasing the measurement's standard deviation to $\sigma_2^2 = 0.6 \times 10^{-4}$ results in a similar outcome, with a false alarm occurring at $t = 6$. At standard deviations $\sigma_3^2 = 1.4 \times 10^{-4}$ and $\sigma_4^2 = 1.8 \times 10^{-4}$, no false alarms are recorded.

Furthermore, Fig. 9 details the detection results under five different standard deviation settings. As the measurement standard deviation increases from σ_1 to σ_4, the false alarm rate gradually decreases, and the detection accuracy improves to 100%. This improvement occurs because an increase in the measurement's standard deviation leads to a higher coefficient in the covariance matrix, preventing the detector from mistaking measurement noise for an attack.

3) Performance of Different Power Systems

In this subsection, the performances on the IEEE-14, 30, and 118 systems are compared. In Fig. 10, the proposed detector consistently achieves lower time delays without any false alarms. When the network scale is expanded to the 118 bus system, the time delays and false alarm rates for I-FDIAs are 0.0415 and 0.00%, respectively, which is consistent with E-FDIAs as well. As the network scale increases from 14 to 118 buses, the operational complexity of the power grids and the substantial increase in measurements and control variables add complexity to detection algorithm. The calculations indicate that a larger network scale slightly increases the detection delay. Therefore, these results validate the scalability of the proposed detection algorithm in large-scale power grids.

4) Performance of Different Weight Coefficient c_β

To evaluate the impact of weight coefficient c_β on detection performance, c_β is varied in the proposed detector on the 14 bus system for I-FDIAs. The results for false alarm, detection delay, and precision are presented in Table IV. In all scenarios, an increase in the relative cost coefficient c_β correlates with a decrease in detection delay. For instance, the minimal detection

TABLE IV
DETECTION RESULTS USING DIFFERENT c_β FOR I-FDIA

c_β	Time delay (s)	RAT	False alarm	Precision
$c_\beta = 0.02$	0.0252	97.56%	0.00%	1.000
$c_\beta = 0.20$	0.0035	99.66%	0.00%	1.000
$c_\beta = 0.40$	0.0031	99.69%	0.00%	1.000
$c_\beta = 0.80$	0.0024	99.76%	0.00%	1.000

delay observed is 0.0024 when $c_\beta = 0.80$, with the corresponding RAT increasing to 99.76%. According to Eq. (29), this occurs because a larger the relative cost coefficient means the defender prioritizes reducing detection delays in smart grids. Consequently, based on the operational requirements of power systems, schedulers may opt for a smaller c_β to enhance attack detection precision and reduce time delay, or conversely, a larger c_β to decrease sensitivity.

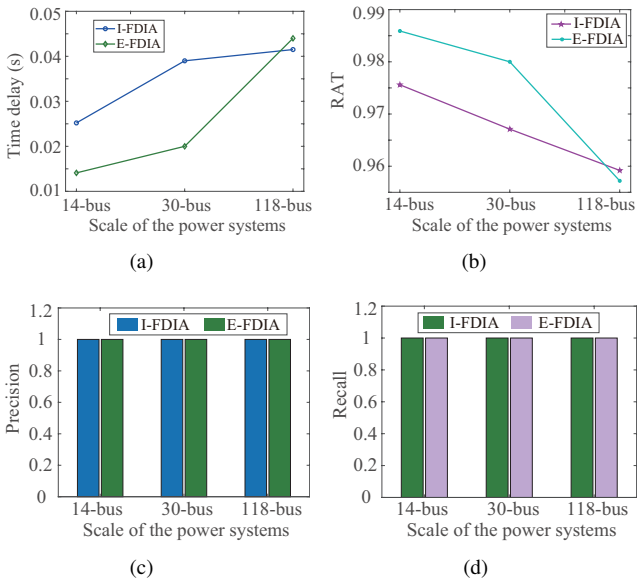


Fig. 10. Results obtained on different network scale: (a) detection delay, (b) RAT, (c) precision, (d) recall.

VI. Conclusion

This paper proposes an extended false data injection attack (E-FDIA) strategy and deep reinforcement learning (DRL)-based detection method. First, building on initial false data injection attack (I-FDIA), the optimization model about the extended components of E-FDIA is established, and a homologous matrix is constructed to extended attack vector avoiding detection by neural attack detection. In our proposed optimization model, the extended injection components are minimized, which are stealthy to neural attack detection scheme and bad data detectors. Moreover, avoiding residual-based detection, this paper proposes an extended Euclidean distance indicator to distinguish the measurements and dynamic state estimations. To improve detection accuracy, an adaptive weight matrix is proposed in dynamic state estimate, which is integrated into the DRL approach. Experimental results have validated that, compared to state-of-the-art detection methods, the proposed detectors have better performance, and the minimal detection delay can be obtained with strong robustness of I-FDIA and E-FDIA. As for future work, when the grid topology is partially unknown, a DRL-based detection strategy will be proposed under the condition that the grid state observations are subjected to adversarial perturbations.

VII. References

[1] X. Su, C. Deng, J. Yang, *et al*, "DAMGAT based interpretable detection of false data injection attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 15, no. 4, pp. 4182 - 4195, Jul. 2024.

[2] R. Kaviani and K. W. Hedman, "An enhanced energy management system including a real-time load-redistribution threat analysis tool and cyber-physical SCED," *IEEE Trans. Power Syst.*, vol. 37, no. 5, pp. 3346 - 3358, Sep. 2022.

[3] H. Y. Tran, J. Hu, X. Yin, and H. R. Pota, "An efficient privacy-enhancing cross-silo federated learning and applications for false data injection attack detection in smart grids," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2538 - 2552, Apr. 2023.

[4] A. S. Musleh, G. Chen, Z. Dong, *et al*, "Attack detection in automatic generation control systems using LSTM-based stacked autoencoders," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 153 - 165, Jan. 2023.

[5] J. Yan, G. Yang, and Y. Wang, "Dynamic reduced-order observer-based detection of false data injection attacks with application to smart grid systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6712 - 6722, Oct. 2022.

[6] B. Li, R. Lu, G. Xiao, *et al*, "Detection of false data injection attacks on smart grids: a resilience-enhanced scheme," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2679 - 2692, Jul. 2022.

[7] Y. Zhang, J. Wang, and B. Chen, "Detecting false data injection attacks in smart grids: a semi-supervised deep learning approach," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 623 - 634, Jan. 2021.

[8] J. Tian, B. Wang, J. Li, *et al*, "Exploring targeted and stealthy false data injection attacks via adversarial machine learning," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 14116 - 14125, Aug. 2022.

[9] Y. Wang, Z. Zhang, J. Ma, and Q. Jin, "KFRNN: An effective false data injection attack detection in smart grid based on Kalman filter and recurrent neural network," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6893 - 6904, May 2022.

[10] M. N. Kurt, Y. Yilmaz, and X. Wang, "Real-time detection of hybrid and stealthy cyber-attacks in smart grid," *IEEE Trans. Inf. Forensics Security* , vol. 14, no. 2, pp. 498 - 513, Feb. 2019.

[11] J. Zhang and X. Wang, "Low-complexity quickest change detection in linear systems with unknown time-varying pre- and post-change distributions," *IEEE Trans. Inf. Forensics Security*, vol. 67, no. 3, pp. 1804 - 1824, Mar. 2021.

[12] J. Zhang and X. Wang, "Consensus-based distributed quickest detection of attacks with unknown parameters," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1864 - 1885, Mar. 2021.

[13] T. Zhou, K. Xiahou, L. L. Zhang, and Q. H. Wu, "Real-time detection of cyber-physical false data injection attacks on power systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6810 - 6819, Oct. 2021.

[14] S. Nath, I. Akingeneye, J. Wu, and Z. Han, "Quickest detection of false data injection attacks in smart grid with dynamic models," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 1, pp. 1292 - 1302, Feb. 2022.

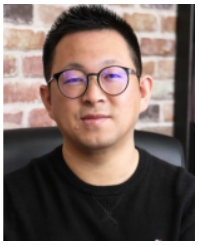
[15] A. Naha, A. Teixeira, A. Ahlén, and S. Dey, "Sequential detection of replay attacks," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1941 - 1948, Mar. 2023.

[16] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: a reinforcement learning approach," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5174 - 5185, Sep. 2019.

[17] K. Liu, H. Zhang, Y. Zhang, and C. Sun, "False data-injection attack detection in cyber-physical systems with unknown parameters: a deep reinforcement learning approach," *IEEE Trans. Cybern.*, vol. 53, no. 11, pp. 7115 - 7125, Nov. 2023.
[18] S. Wei, J. Xu, Z. Wu, *et al*, "A false data injection attack detection strategy for unbalanced distribution networks state estimation," *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 3992 - 4006, Sep. 2023.
[19] M. Kesici, B. Pal, and G. Yang, "Detection of false data injection attacks in distribution networks: a vertical federated learning approach," *IEEE Trans. Smart Grid*, vol. 15, no. 6, pp. 5952 - 5964, Nov. 2024.
[20] Y. Li, X. Wei, Y. Li, *et al*, "Detection of false data injection attacks in smart grid: a secure federated deep learning approach," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4862 - 4872, Nov. 2022.
[21] J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3271 - 3280, Jul. 2018.
[22] H. Yang, X. He, Z. Wang, *et al*, "Blind false data injection attacks against state estimation based on matrix reconstruction," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 3174 - 3187, Jul. 2022.
[23] X. Ran, W. P. Tay, and C. H. T. Lee, "A robust deep Q-network based attack detection approach in power systems," *in Proc. Int. Conf. Smart Power Internet Energy Syst.*, Dec. 2022, pp. 995 - 1000.
[24] X. Ran, W. P. Tay, and C. H. T. Lee, "Robust data-driven adversarial false data injection attack detection method with deep Q-network in power systems," *IEEE Trans. Ind. Informat.*, vol. 20, no. 8, pp. 10405 - 10418, Aug. 2024.
[25] Z. Wang, Q. Zhang, H. Sun and J. Hu, "Detection of false data injection attacks in smart grids based on cubature kalman filtering," *in Proc. Chin. Control and Decis. Conf.*, May 2021, pp. 2526 - 2532.
[26] R. Zhang, Q. Zhang, Z. Wang, and H. Sun, "Detection of false data injection attack in smart grid based on iterative Kalman filter," *in Proc. China Autom. Congr.*, Oct. 2021, pp. 6083 - 6088.
[27] D. An, F. Zhang, Q. Yang, and C. Zhang "Data integrity attack in dynamic state estimation of smart grid: attack model and countermeasures," *IEEE Trans. Autom. Sci. Eng.* , vol. 19, no. 3, pp. 1631 - 1644, Jul., 2022.
[28] M. A. Rahman and A. Datta, "Impact of stealthy attacks on optimal power flow: a simulink-driven formal analysis," *IEEE Trans. Depend. Sec. Comput.*, vol. 17, no. 3, pp. 451 - 464, May-Jun., 2020.
[29] C. Pei, Y. Xiao, W. Liang, and X. Han, "A deviation-based detection method against false data injection attacks in smart grid," *IEEE Access*, vol. 9, pp. 15499 - 15509, Jan., 2021.
[30] C. Wang, D. Deng, L. Xu, and W. Wang, "Resource scheduling based on deep reinforcement learning in UAV assisted emergency communication networks," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3834 - 3848, Jun., 2022.
[31] S. Xu, Y. Li, S. Guo, *et al*, "Cloud edge collaborative SFC mapping for industrial IoT using deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4158 - 4168, Jun., 2022.

Xiaohong Ran received the Ph.D. degree in the electrical power engineer from the Huazhong University of Science and Technology, Wuhan, China, in 2015.

Since 2015, he has been with the school of electrical engineering and Automation in Wuhan University, Wuhan, China. Now he is with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, HongKong. From 2018 to 2021, he was a Visiting Researcher with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA. From 2021 to 2023, he was a Senior Research Fellow at the Energy Research Institute, Nanyang Technological University, Singapore. His research interests include the modeling, stability and control of converter-interfaced power grid, learning based detection, and security assessment and control in cyber-physical power systems.



Lei Ma received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, and the M.E. and Ph.D. degrees from The University of Tokyo, Tokyo, Japan, in 2011 and 2014, respectively.

He is currently an Associate Professor with The University of Tokyo, and the University of Alberta, Edmonton, AB, Canada. He was honorably selected as Canada CIFAR AI Chair and a fellow with Alberta Machine Intelligence Institute (Amii), Edmonton. His research interests include the interdisciplinary felds of software engineering (SE) and trustworthy artifcial intelligence, with a special focus on the quality, reliability, safety, and security aspects of AI systems.