Defense of Massive False Data Injection Attack via Sparse Attack Points Considering Uncertain Topological Changes

Xiaoge Huang, Zhijun Qin, Ming Xie, Hui Liu, and Liang Meng

***Abstract*—False data injection attack (FDIA) is a typical cyber-attack aiming at falsifying measurement data for state estimation (SE), which may incur catastrophic consequences on cyber-physical system operation. In this paper, we develop a deep learning based methodology for detection, localization, and data recovery of FDIA on power systems in a coherent and holistic manner. However, the multi-modal probability distributions of both measurements and state variables in SE due to ever-changing operating points and structural/topological changes pose great challenges in detecting and localizing FDIA. To address this challenge, we first propose an enhanced attack model to launch massive FDIA on limited access points. Second, we train an auto-encoder (AE) with a Bayesian change verification (BCV) classifier using $N-1$ contingencies to detect FDIA with unseen $N-k$ operational topologies. Third, to avoid model collapse caused by multi-modal measurement distribution, an AE-based generative adversarial network (GAN) is derived to generate a diverse candidate set of normal measurement vectors with various operational topologies. Finally, we develop a pattern match algorithm to localize and recover the falsified measurements and state variables by comparing the falsified measurement vectors with the normal measurement vectors in the candidate set. Case studies with IEEE benchmark systems and a modified 415-bus China Southern Grid system are provided to validate the proposed methodology. It shows that the proposed methodology achieves an average 95% accuracy for detection, over 80% accuracy for localization of FDIA, and recovers the measurement and state variables close to their true values.**

***Index Terms*—False data injection attack, auto-encoder, generative adversarial network, state estimation, cyber security.**

I. Introduction

UBIQUITOUS applications of information technologies and tele-communications pose great challenges to the security and resilience of power grid operation. Cyber-attacks have been identified as major threats for power grids and associated stakeholders. False data injection attack (FDIA) is a typical cyber-attack aiming at falsifying measurement data for static state estimation (SE), which may incur catastrophic consequences on the power grid operation [1]. For example, on December 23, 2015, FDIA was launched against Ukraine, resulting in a massive blackout covering seven 110 kV substations and twenty-three 35 kV substations. The power supplies of the three control zones, over 80000 users, were interrupted by this attack [2]. Therefore, many research efforts have been put into real-time intrusion and FDIA detection to enhance information security and integrity.

Even though FDIA was first proposed and realized in power grid operation, cyber-physical systems (including the new generation of smart grids) built upon communications and publicly accessible sensor networks are also vulnerable to FDIA, as these access points are exposed to cyber-attackers. Despite that FDIA can cause adverse effect on dynamical system, e.g., [3]-[5], we focus on the defense of static FDIA targeting on a single snapshot.

To address FDIA challenges, we take effort in developing a systematic defense methodology for online detection, localization, and recovery of both measurements and state variables from FDIA. Different from the existing work focusing on one single technical problem, we place the emphasis on the entire chain from issuing alarms of FDIA, identifying compromised measurements, and mitigating adverse impacts by the recovery of both measurements and states. However, it is difficult to identify the characteristics of the probability density function (PDF) of measurement and state vectors of SE, as these distributions are intrinsically multi-modal. Therefore, the FDIA detector should be generalized for uncertain/unseen PDFs. To achieve this goal, we have applied generative deep-learning models to learn the deep structure of the multi-modal PDF and falsified measurements and state variables.

The challenges of FDIA were first identified by [6] and aroused widespread concerns. Much research work has been carried out in the past decade, focusing on the attack model [7], [8], impact assessment [9], and detection of FDIA [10]-[16].

For the detection of FDIA, both model-based algorithms [10]-[12] and data-driven approaches have been applied. For

Manuscript received: September 30, 2020; revised: January 1, 2021; accepted: June 10, 2021. Date of CrossCheck: June 10, 2021. Date of online publication: September 2, 2021.

This work was supported in part by the National Natural Science Foundation of China (No. 51767001).

X. Huang, Z. Qin (corresponding author), and H. Liu are with the School of Electrical Engineering, Guangxi University, Nanning, China (e-mail: nnhuangxg@163.com; zjqin@gxu.edu.cn; huglhl@126.com).

M. Xie and L. Meng are with the Guangxi Power Grid Co. Ltd., China Southern Grid (CSG), Guilin, China (e-mail: xie_m@gx.csg.cn; meng_l.xt@gx.csg.cn).

DOI: 10.35833/MPCE.2020.000686

model-based algorithms, unscented Kalman filter [10], graph theory based method [11], and temporal correlation analysis technique [12] are applied. For the machine learning approaches, conditional deep belief network [13], wavelet transform and recurrent neural network [14], improved ensemble learning [15], and multivariant Gaussian anomaly detection [16] have been applied. The general breakthrough shared by machine learning based methods is to distinguish camouflaged false data from normal measurements based on various statistical properties. While most of the above research has demonstrated effective performance, the common assumption in their works, i.e., implicitly assuming a fixed distributional mode on the measurement vectors in SE, is not generally valid. A comprehensive survey of the detection algorithms for FDIA is given in [17]. Besides, the localization of FDIA still needs further investigation. References [18] and [19] localize and isolate the attacked points of FDIA using logical judgment matrices of attack signature. A convolutional neural network based multi-label classifier is applied to localize the FDIA, which relies on a massive training dataset [20]. Mixed outlier detection [21] and deep learning based interval SE [22] are used to locate the errors on estimated state variables caused by FDIA with small-scale cases. The basic assumption of FDIA is that both the attacker and the defender have complete information on the topology and parameters of the power grid. References [23] and [24] propose two novel attack models with incomplete information, demonstrating the significance and complexity of defending FDIA. We would refer the readers to our recent survey [25] on major advancements and limitations of the existing FDIA research. Besides the detection of FDIA, identifying the locations of falsified measurements, rather than the locations of falsified state variables, is critical for the prevention of FDIA, as the system operators can consolidate identified vulnerable components or communication channels.

The above research works have missed a key aspect to improve the overall effectiveness in defending FDIA. We notice that the distribution of the measurement has multi-modal properties due to the diversified and ever-changing system structure, i.e., operating points and topologies, of power systems. We hereby make mild assumptions on the distribution of measurements, rather than assume a particular family of distributions. In this paper, we assume that the defender and the attacker possess the complete information of the power grid. In designing our defense methodology based on deep-learning techniques, we resolve the generalization difficulties caused by multi-modal distribution to a wider range of uncertain operating points of power systems. As a result, the proposed detection/localization model trained offline can be generalized to online ever-changing operating points without retraining.

To summarize, we make the following contributions.

1) We develop an enhanced attack model to launch FDIA with a limited number of targeted access points. The key feature of this model, as compared with [6], is that this model can launch massive FDIA by attacking only smaller numbers of substations. From the defender's standpoint, the proposed FDIA model leads to a better understanding of the impact of FDIA and better planning for systematic defense methodology.

2) We develop an auto-encoder (AE) feature extractor together with Bayesian change verification (AE-BCV) classifier to detect FDIA. The AE is trained to learn lossless mapping from multi-modal joint probability density distribution of state variables along with normal/falsified measurement vectors into a lower-dimensional distribution. Then, the BCV classifier is applied to detect FDIA with adaptiveness for unseen topological changes of power grids.

3) We derive an AE-based generative adversarial network (AE-GAN) for the offline generation of various multi-modal probability distributions of normal measurements under unseen power system topologies, which constructs a candidate set for localization and recovery of falsified measurement data. Compared with [26], AE-GAN can avoid the model collapse problem. Therefore, the candidate set consists of measurement samples under various unseen system structures, rather than being stuck into pre-selected topologies.

4) We develop a pattern match algorithm for the online recovery of falsified measurements/states to their normal values: ① clean the measurement vectors by locating and removing suspicious falsified measurements; ② compare the similarities of the cleaned measurement with the candidate set and choose the most similar measurement vector from the candidate set as a recovered measurement vector; ③ recover state variables with the cleaned measurements.

For sake of clarification, the major differences between this research work and closely related research works are summarized below.

1) Although the proposed attack models are extended based on [6] and [27], the fundamental differences between our research work and [9], [27] are that the proposed attack models falsify much fewer measurements and cause substantial changes in more state variables. Additionally, in the targeted mode, we restrict the changes of state variables in non-targeted substations, leading to a less detectable massive FDIA.

2) The proposed methodology places an emphasis on the topological changes in the detection/localization of falsified measurements and recovery of corresponding state variables, as compared with [21]. We construct our training/validation dataset with unseen contingencies.

3) We aim to localize and recover the falsified measurements rather than state variables, to provide more insights for the prevention/consolidation against FDIA. Compared with [21] which localizes falsified state variables using a predefined candidate set of normal state variables, we construct a comprehensive candidate set with AE, which is, in theory, able to generate an infinite number of samples for the candidate set. Reference [28] provides a novel SE method which can recover the state variables. Different from that work, we recover both the measurements and state variable considering the topology uncertainty, and validate our methods using a real-world larger power system.

The rest of this paper is structured as follows. In Section II, the enhanced FDIA model is formulated. From Section III to Section V, an overview of the proposed defense meth-

odology, methods for detection, and localization and data recovery of FDIA are described, respectively. Case studies are provided in Section VI. Finally, we conclude our work in Section VII.

II. Enhanced FDIA Model

In this section, we propose an enhanced FDIA model that can launch massive FDIA by falsifying a larger number of measurements with fewer attack points, as compared with the existing models without triggering bad data detection (BDD). For ease of verification and a moderate research scope, we focus on the DC SE model with complete information, and formulate our proposed model into a convex optimization problem. Note that the proposed model can also be extended to full AC SE, which will be applied to generate FDIA samples for training deep learning models in subsequent sections.

BDD had been regarded as a strong guarantee for measurement reliability in power system SE [29]. However, FDIA, known as a topology-knowledge-based cyber-attack, is able to inject malicious measurement data bypassing BDD [30]. We begin with the DC SE model and BDD principle, followed by the proposed model.

A. DC SE Model and BDD Principle

SE aims to estimate the states of the power system via the measurement model. Considering m measurements in an n-bus power system, the measurement model using DC power flow is:

$$\boldsymbol{z}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{e} \tag{1}$$

where $\boldsymbol{z}\in\mathbb{R}^m$ is the measurement vector, normally including phase angles, transmission line power flows, etc.; $\boldsymbol{x}=(\theta_i)^{\mathrm{T}}$, $\boldsymbol{x}\in\mathbb{R}^{n-1}$ denotes the state variables, namely the phase angles θ_i; $\boldsymbol{e}\in\mathbb{R}^m$ is the irreducible measurement noise; and $\boldsymbol{H}\in\mathbb{R}^{m\times(n-1)}$ is the Jacobian matrix containing the information about the power grid topology. Therefore, $\boldsymbol{H}$ is time-varying due to the changes of system structures or topologies.

The DC SE model can be solved by minimizing the weighted least square (WLS) as (2), and the estimation result $\hat{\boldsymbol{x}}$ has a closed form given by (3).

$$\min \boldsymbol{J}(\boldsymbol{x})=(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x}) \tag{2}$$

$$\hat{\boldsymbol{x}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z} \tag{3}$$

where $\boldsymbol{R}$ is a diagonal matrix with diagonal elements equal to σ_i^2, and σ_i is the measurement error of the i^{th} bus.

Based on BDD theory, bad data can be detected if the following condition holds:

$$\left\|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}\right\|_2>\tau_{\mathrm{BDD}} \tag{4}$$

where τ_{BDD} is the pre-defined threshold for BDD.

B. FDIA Principle and Proposed FDIA Model

Let $\boldsymbol{a}$ denote the injection data for $\boldsymbol{z}$, and $\boldsymbol{z}'=\boldsymbol{z}+\boldsymbol{a}$ is the falsified measurement vector, $\boldsymbol{c}$ denote the introduced error in the estimation, and $\boldsymbol{x}'=\hat{\boldsymbol{x}}+\boldsymbol{c}$ denote the estimated state vector after FDIA attack.

If condition (5) holds [6], BDD will fail to detect abnormal measurements, as the residual of measurement model is less than the pre-defined threshold τ_1, as in (6).

$$\boldsymbol{a}=\boldsymbol{H}\boldsymbol{c} \tag{5}$$

$$\left\|\boldsymbol{z}'-\boldsymbol{H}\boldsymbol{x}'\right\|_2=\left\|\boldsymbol{z}+\boldsymbol{a}-\boldsymbol{H}(\hat{\boldsymbol{x}}+\boldsymbol{c})\right\|_2=\left\|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}\right\|_2\le\tau_1 \tag{6}$$

Given the complete information on $\boldsymbol{z}$, $\hat{\boldsymbol{x}}$, and $\boldsymbol{H}$, cyber attackers are able to launch the most effective and concealed FDIA by solving a convex FDIA attack model [27] given by (7), which poses great challenges to power grid operation.

$$\min_{\boldsymbol{c}}\left\|\boldsymbol{H}\boldsymbol{c}\right\|_1 \tag{7}$$

However, cyber-attackers need to compromise meters by using the above model. As shown in [6], to create an impact on 10 variables in the error vector $\boldsymbol{c}$, 60-140 variables of injected measurement data $\boldsymbol{a}$ for the IEEE 118-bus system are falsified.

We propose an enhanced FDIA model by adding several constraints to (7) based on two realistic attack preferences. These expansions of the model enable the attackers to launch massive FDIA with fewer compromised meters. The FDIA can be raised to a specified intensity on a preselected set of buses.

1) Model I: Untargeted FDIA

In this model, the attacker aims to cause valid errors on the system without specific targeting. The optimal strategy of this model is to: ① ensure that the total impact on estimation reaches a given level; and ② minimize the number of compromised meters. This strategy enables the attacker to launch a massive FDIA with restricted accesses to meters. The model is given by:

$$\begin{cases}\min\limits_{\boldsymbol{c}}\left\|\boldsymbol{H}\boldsymbol{c}\right\|_1\\ \text{s.t.}\ \ \sum\boldsymbol{c}\ge k\sum|\hat{\boldsymbol{x}}|\end{cases} \tag{8}$$

where k entitled attack intensity is a given value ensuring that the total caused estimation error is k times larger than the sum of the absolute values of real estimation.

2) Model II: Targeted FDIA

Targeted FDIA aims to cause valid errors on a selected set of estimated measurements. The optimal strategy of this model is to: ① cause the impacts on given variables to reach a given level; ② minimize the number of compromised meters and the total error of the estimate state. This model is formulated as:

$$\begin{cases}\min\limits_{\boldsymbol{c}}\left(\left\|\boldsymbol{H}\boldsymbol{c}\right\|_1+\left\|\boldsymbol{c}\right\|_1\right)\\ \text{s.t.}\ \ |\boldsymbol{c}|\ge|\boldsymbol{c}_e|\\ \quad\ \boldsymbol{c}_e=k\boldsymbol{x}_e\end{cases} \tag{9}$$

where $\boldsymbol{c}_e\in\mathbb{R}^n$ is a given vector, and $\boldsymbol{c}_e$ represents the expected errors on the estimated state variables; and $\boldsymbol{x}_e$ is the affected state vector. If the i_e^{th} variable is not expected to be falsified, $\boldsymbol{x}_e(i_e)=0$, otherwise $\boldsymbol{x}_e(i_e)=\hat{\boldsymbol{x}}(i_e)$. The constraint $\boldsymbol{c}_e=k\boldsymbol{x}_e$ ensures that valid errors are injected into pre-selected state variables.

k specifies the attack strategies of the attackers with a limited budget of accessible attack points. By increasing k, the attackers cause a larger deviation of many state variables

(untargeted FDIA), or a selected set of state variables (targeted FDIA). System operators have priori knowledge to perceive abnormal state deviation, it is still difficult to assert an FDIA event in case of extreme operational conditions or in junction with physical attacks [31]. More importantly, the proposed attack model poses great challenges in the localization of the sparse falsified measurements and the system-wide affected state variables by launching intensive FDIA.

III. Defense Methodology Against FDIA

We propose a comprehensive methodology as in Fig. 1 to systematically address the challenges of FDIA by detecting and locating falsified measurement by FDIA, and subsequently recovering these measurements to their correct values.

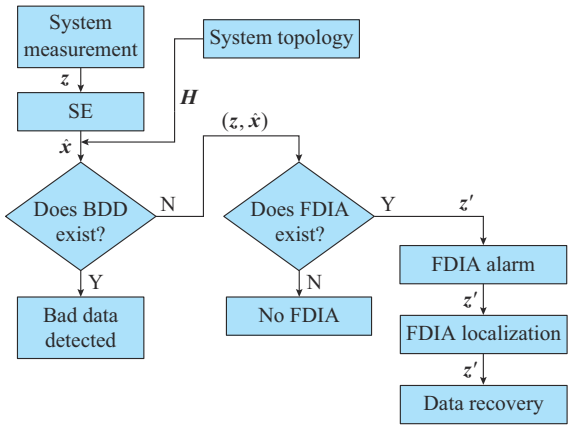

Fig. 1. Proposed methodology for defensing FDIA.

This methodology includes three co-related tasks as follows.

1) Detection. $\boldsymbol{z}$, $\boldsymbol{H}$, $\hat{\boldsymbol{x}}$ are first sent to the BDD module. If no bad data are detected, $\boldsymbol{z}$ and $\hat{\boldsymbol{x}}$ will be sent to the AE-BCV classifier to detect FDIA.

2) Localization. Once an FDIA event is detected, $\boldsymbol{z}$ will be labelled as $\boldsymbol{z}'$ that will be sent to the localization module, aiming to identify the set of measurements having been falsified.

3) Recovery. Once the injections on measurements are located, $\boldsymbol{z}'$ will then be sent to the recovery module. The falsified variables on $\boldsymbol{z}'$ will be recovered to approximate its original values. Finally, the true value of $\hat{\boldsymbol{x}}$ is re-estimated based on the recovered $\boldsymbol{z}'$ and $\boldsymbol{H}$.

In the subsequent sections, we elaborate the models and algorithms for detection, localization, and measurement recovery, respectively.

IV. AE-BCV-based FDIA Detector

The detection of FDIA aims to determine whether the measurement vector is falsified by FDIA. Therefore, we formulate the FDIA detection as a binary classification problem. However, the detection of FDIA is faced with two major challenges. First, the large size of the measurement vector causes unobvious variation on a subset of measurements, reducing the sensitivity of the classifier for the change of measurement distribution. Second, the ever-changing operating points of the power grid (especially the topology of the power grid) pose great challenges in training the classifier, as the training datasets cannot cover all possible unseen operational scenarios in the training stage. The unseen probability distributions of measurements in the test dataset will be viewed by the classifier as a novel category that is not included in the training, which may lead to poor performance in the identification of FDIA.

In view of these challenges, we propose to apply AE [32] as the feature extractor combined with a decision maker based on BCV [33]. AE is used to learn a compressed mapping of the original higher-dimensional multi-modal distribution of the state/measurement vector into a lower-dimensional space. We notice that the existing methods of dimensionality reduction, e.g., principal component analysis, aim to remove unimportant dimensions in the vector. However, each variable in the measurement vector is equally important, as FDIA may falsify on any variable. AE reduces the dimensions of measurement via the encoding process. The key features of the distribution of the state/measurement vector are retained during this process. The BCV-based decision maker transforms the classification problem into a hypothetical decision-making problem to overcome the second challenge. The BCV-based decision maker makes robust decisions by calculating the probabilities of a pair of mutually exclusive hypotheses. The impact of generalization error (due to unseen operating points of the power grid) will be greatly reduced in detecting FDIA, as compared with conventional classification-based methods.

A. AE-based Feature Extractor Model

AE is a feed-forward neural network consisting of an encoder q_d and a decoder p_d. As in Fig. 2, the input of AE $\boldsymbol{y}_d \in \mathbb{R}^{m+n-1}$ consists of $\boldsymbol{z}$ and $\hat{\boldsymbol{x}}$. The encoder aims to encode the joint PDF of $(\boldsymbol{z}, \hat{\boldsymbol{x}})$ with a compressed representation, i.e., the code $\boldsymbol{c}_d \in \mathbb{R}^d$. The decoder subsequently reconstructs the input as $\tilde{\boldsymbol{y}}_d \in \mathbb{R}^{m+n-1}$ with the compressed code. By training AE with the dataset, the reconstructed input will asymptotically approach the original distributions.

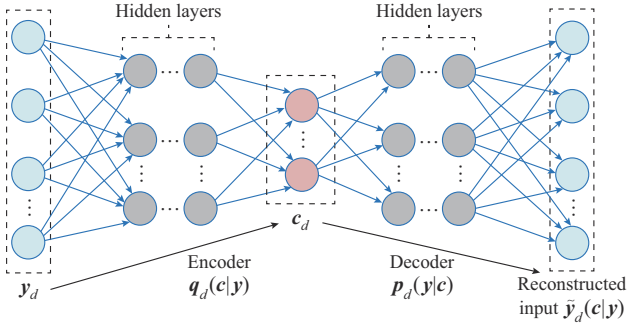

Fig. 2. AE-based feature extractor model.

Both the encoder and the decoder are essentially nonlinear mapping, which are briefly written as:

$$\boldsymbol{c}_d = \sigma(\boldsymbol{\omega}\boldsymbol{y}_d + \boldsymbol{b}) \tag{10}$$

$$\tilde{\boldsymbol{y}}_d = \sigma(\tilde{\boldsymbol{\omega}}\boldsymbol{c}_d + \tilde{\boldsymbol{b}}) \tag{11}$$

where $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ are the weights of mapping layers; $\boldsymbol{b}$ and $\tilde{\boldsymbol{b}}$ are the biases; and σ is the activation functions such as ReLU, Sigmoid, etc. Parameters in AE model, i.e., weights and biases, will be determined in the training process, which minimizes the mean square error between the input and reconstructed input as:

$$\min_{\boldsymbol{\omega},\tilde{\boldsymbol{\omega}},\boldsymbol{b},\tilde{\boldsymbol{b}}} \left\| \boldsymbol{y}_d - \sigma\{\tilde{\boldsymbol{\omega}}[\sigma(\boldsymbol{\omega}\boldsymbol{y}_d + \boldsymbol{b})] + \tilde{\boldsymbol{b}}\} \right\|_2 \tag{12}$$

After the AE is well-trained, we will only use the encoder as the feature extractor. The input will first map to the compressed code by the encoder. Then, this code will be sent to the BCV-based decision maker, as shown in Fig. 3.

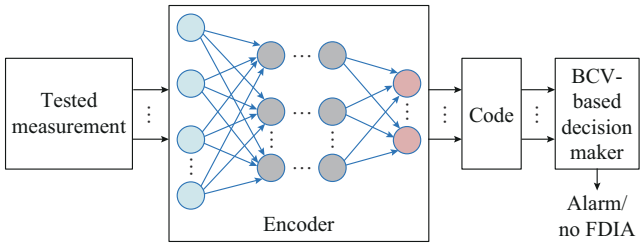

Fig. 3. BCV-based decision maker model.

B. BCV-based Decision Maker to Detect FDIA

The BCV-based decision maker proposed in [33] aims to perform face recognition under uncertain conditions. Inspired by BCV, we aim to determine whether the measurement vector has been falsified with uncertain topological change of the power grid by solving a hypothetical decision-making problem formulated as:

$$r(\boldsymbol{c}_1, \boldsymbol{c}_2) = \lg \frac{P(\Delta|H_1)}{P(\Delta|H_2)} \geq \tau_{\text{BCV}} \tag{13}$$

where $\boldsymbol{c}_1$ is the given code of a reliable measurement vectors from a training dataset containing N codes; $\boldsymbol{c}_2$ is the code of the detection input we want to test; H_1 is the hypothesis where $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are equally reliable; and H_2 is the hypothesis where $\boldsymbol{c}_2$ is falsified by FDIA. The difference in the two codes is given by $\Delta = \boldsymbol{c}_1 - \boldsymbol{c}_2$, and τ_{BCV} is a given threshold. Based on the maximum a posterior (MAP) rule, we make the decision by testing the log-likelihood ratio r, where r measures the similarity between $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ [33]. If r exceeds the given threshold, H_1 is true, indicating that the tested measurement vector is reliable. Otherwise, H_2 is true, indicating that FDIA has falsified the tested measurement vector.

With N training samples, the posterior probabilities of Δ based on H_1 and H_2 can be calculated by:

$$P(\Delta|H) = \prod_{u=1}^{N} P(\Delta^{(u)} = \boldsymbol{c}_1^{(u)} - \boldsymbol{c}_2^{(u)}|H_v) \quad v = 1, 2 \tag{14}$$

V. AE-GAN-based Localization and Recovery Algorithms

In this section, we aim to locate the falsified variables in the measurement vector and recover those variables. The localization and recovery process contain two steps: ① an AE-GAN model is proposed to generate a candidate set of distributions of normal measurements; ② a pattern match algorithm is proposed to locate the attacked point and identify the most likely candidate measurement in the candidate set.

A major challenge to fulfill this goal comes from the multi-modal distributions of power grid measurements, i.e., while the power grid is operating normally, there exists multiple reasonable distributions for the measurement vector. It is infeasible to fit all these reasonable distributions and determine which one is the most likely.

To overcome this challenge, AE-based GAN is used to capture the multi-modal distributions and generate a diverse candidate set of measurement distributions under normal operational conditions. By theoretical analysis and case studies, AE-GAN can generate an infinite number of multi-modal distributions that are highly similar with measurement vectors under normal operational conditions. Meanwhile, the AE-GAN model has overcome model collapse [34], even if the generation target has multi-modal distributions, which is common in other generative models such as GAN.

A. AE-GAN Model

The general GAN consists of two networks, i.e., the generator and the discriminator. The discriminator aims to classify the generated input, whereas the generator aims to cheat the discriminator by generating various distributions for normal/falsified measurements. These two networks will be trained by playing dynamic optimization games against each other [35]. The generator will be able to generate data with highly similar PDF as the input data.

In AE-GAN model, the decoder in AE will be set as the generator in GAN. The AE model approximates the distribution of generated measurements to the distribution of normal measurements. To generate as many candidate measurements as possible, the trained decoder uses Gaussian distribution as the input code, as the Gaussian distribution is the most extensively used input distribution of generative models. Thus, Gaussian noise is set as the reference in the GAN training process to approximate the input code of the decoder to Gaussian distribution.

Compared with conventional GAN models, the design of AE-GAN in Fig. 4 addresses two fundamental challenges in generating multi-modal measurement distributions.

1) Model collapse. The training of GAN has brittle convergence properties due to the model collapse in this problem. Model collapse is one kind of GAN training failure incurred by the multi-modal distributions of input data [34]. To avoid model collapse, we replay the input as Gaussian distribution and use the encoder in AE as the generator. The encoder maps the multi-modal distributions of power system measurements to the single-modal distribution. For the GAN part in AE-GAN, both the input and the generated data have single-modal distributions, and the model collapse is avoided.

2) Generate infinite multi-modal candidate data. Since the training of GAN approximates the generated data to the real input, the decoder is mapping the Gaussian distribution to multi-modal distribution of power system measurement data. Thus, a well-trained decoder can generate infinite multi-modal candidate data with infinite Gaussian random samples.

In Fig. 4, the encoder $\boldsymbol{q}_{lr}$ plays as the generator, and an additional discriminator is introduced to the network.

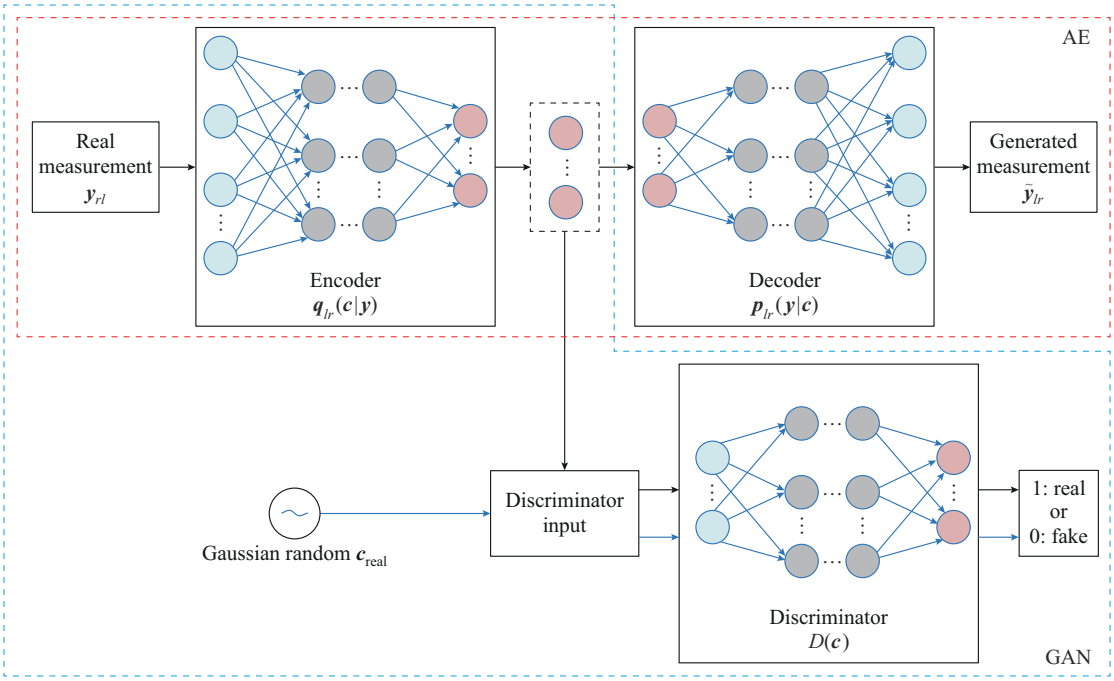

Fig. 4. AE-GAN model.

In each training step, the normal measurement $\boldsymbol{y}_{lr}$ is sent to AE to output generated measurement $\tilde{\boldsymbol{y}}_{lr}$, then the entire AE network is trained one time to minimize (12). Since we must enforce the code distribution obeying Gaussian distribution, this code will be regarded as the fake input of discriminator, denoted as $\boldsymbol{c}_{\text{fake}}$. Then, $\boldsymbol{c}_{\text{fake}}$ and the Gaussian random sample $\boldsymbol{c}_{\text{real}}$ are sent to discriminator $D(\boldsymbol{c})$, which will output the probability, and $\boldsymbol{c}$ is a reliable measurement vector. Based on the output result, the entire network is trained by [35].

$$\min_{\boldsymbol{q}_{lr}} \max_{D} V(\boldsymbol{q}_{lr},D)=\mathbb{E}_{\boldsymbol{c}_{\text{real}}\sim p_{\text{Gaussian}}}[\lg D(\boldsymbol{c}_{\text{real}})]+ \mathbb{E}_{\boldsymbol{c}_{\text{fake}}\sim \boldsymbol{q}_{lr}(\boldsymbol{y}_{lr})}[\lg(1-D(\boldsymbol{c}_{\text{fake}}))] \quad (15)$$

where $\boldsymbol{p}_{\text{Gaussian}}$ is a Gaussian distribution. By training with (15), the generated distribution will approach Gaussian distribution, and the Nash-equilibrium is achieved.

B. Localization and Measurement Recovery of FDIA

Among the massive measurements in normal operational conditions generated by AE-GAN, we can identify the most likely one to approximate the original measurement vector. Then, the falsified measurements are replaced by corresponding variables in the selected measurement vector.

To pinpoint the falsified measurement, we propose an iterative algorithm for FDIA localization in Algorithm 1, which repeats the screening with two actions: ① check suspected variables by comparing L_{origin} and $L_{removed}$; ② after removing suspected variable, compare the distributions of the falsified and generated measurements to maximize the accuracy of FDIA localization.

Then, a pattern match algorithm for measurement recovery is proposed in Algorithm 2 with two steps: ① replace falsified variables by the reconstructed variables; ② repeat the correction of generated measurements to improve recovery accuracy with the trained AE in the AE-GAN network. Note that at this point, the code encoded by recovered measurements is close to the code decoded by normal measurements. The repetitive correction is essential to searching the optimal distribution among the corresponding codes in the candidate set. Finally, we recover state variables by solving (3) with the recovered measurements.

Algorithm 1: iterative algorithm for FDIA localization

Create k generated measurement, $\boldsymbol{GM} \in \mathbb{R}^m$. Use random Gaussian sample and trained decoder in AE-GAN.
for each $\boldsymbol{GM}$ in k steps:
  Compute original distribution difference L_{origin} between $\boldsymbol{GM}$ and attacked measurement $\boldsymbol{AM}$ as: $L_{origin}=\|\boldsymbol{GM}-\boldsymbol{AM}\|_2$.
  for each i in m steps:
    Remove $\boldsymbol{GM}(i)$ in $\boldsymbol{GM}$ as $\boldsymbol{GM}'$; and remove $\boldsymbol{AM}(i)$ in $\boldsymbol{AM}$ as $\boldsymbol{AM}'$.
    Compute removed distribution difference $L_{removed}$ between $\boldsymbol{GM}$ and $\boldsymbol{AM}$ when removing the i element: $L_{removed}=\|\boldsymbol{GM}'-\boldsymbol{AM}'\|_2$.
    if $L_{origin}-L_{removed}\geq\tau_1$
      Define attacked location $\boldsymbol{AL}$ and let $\boldsymbol{AL}(i)=1$.
    end
  end
  Return the sequence number of nonzero variables in $\boldsymbol{AL}$, remove the variables based on the index in $\boldsymbol{GM}$ and $\boldsymbol{AM}$ as $\boldsymbol{GM}'$ and $\boldsymbol{AM}'$.
  Compute the distance $L_{distance}$ between the rest of $\boldsymbol{GM}$ and $\boldsymbol{AM}$: $L_{distance}=\|\boldsymbol{GM}'-\boldsymbol{AM}'\|_1$
  if $k=1$
    $\tau_2=L_{distance}$; define searched attacked location $\boldsymbol{SAL}$ and let $\boldsymbol{SAL}=\boldsymbol{AL}$; define reconstructed measurement in localization $\boldsymbol{RML}$ and let $\boldsymbol{RML}=\boldsymbol{GM}(k)$.
  else if $\tau_2>L_{distance}$
    $\tau_2=L_{distance}$; $\boldsymbol{SAL}=\boldsymbol{AL}$; $\boldsymbol{RML}=\boldsymbol{GM}(k)$.
  end
end

Algorithm 2: pattern match algorithm for measurements recovery

Define recovered measurement $\boldsymbol{RM}$ and let $\boldsymbol{RM}=\boldsymbol{AM}$.
$\boldsymbol{RM}=\boldsymbol{RML}$ for nonzeros in $\boldsymbol{SAL}$
Define final reconstructed measurement in recovery $\boldsymbol{FRMR}$ and let $\boldsymbol{FRMR}=\boldsymbol{RM}$.
for each i in k steps:
  Use $\boldsymbol{AE}$ to generate new recovered measurement $\boldsymbol{NRM}$ as: $\boldsymbol{NRM}=\boldsymbol{p}_{lr}(\boldsymbol{q}_{lr}(\boldsymbol{RM}))$ and let $\boldsymbol{RM}=\boldsymbol{NRM}$ for nonzeros in $\boldsymbol{SAL}$.
  Return the index of nonzero variables in $\boldsymbol{SAL}$ and remove the variables corresponding to the serial number in $\boldsymbol{NRM}$ and $\boldsymbol{AM}$ as $\boldsymbol{NRM}'$ and $\boldsymbol{AM}'$.
  Compute the distance between the rest of $\boldsymbol{NRM}$ and $\boldsymbol{AM}$: $L_{distance}=\|\boldsymbol{NRM}'-\boldsymbol{AM}'\|_1$.
  if $\tau_2>L_{distance}$: $\tau_2=L_{distance}$; $\boldsymbol{FRMR}=\boldsymbol{RM}$.
  end
end

VI. Case Studies

In this section, we evaluate the proposed attack model and the defense methodology against FDIA.

A. Experimental Setup

Our evaluations are conducted on a computer with 16 GB RAM, an Intel i7-8750H CPU, and an Nvidia RTX 2070 GPU. The dataset is generated by the proposed convex optimization models (8) and (9) with the power system simulation tool MATPOWER and the optimization modeling tool CVX. The proposed methodology is implemented with PyTorch.

B. General Setup of Dataset

1) Normal Cases

We test our methodology using IEEE 57-bus system, IEEE 118-bus system, and IEEE 300-bus system from MATPOWER, and a synthesized 415-bus system with 627 branches based on a part of the China Southern Grid (CSG). Power flow calculations are performed on these systems to obtain base cases.

We diversify normal operational cases as follows: ① perform Monte-Carlo simulations to conduct (configurable for training, validation, and test) line switch to the base cases; ② vary the bus power injection by 50% to 150% of the base case values; ③ perturb the bus power injection by white noise with the variance of 1% of the base case values.

2) FDIA Cases

Cases with falsified measurements by FDIA are generated by overlaying the injection data $\boldsymbol{a}$ and introduced error $\boldsymbol{c}$, both determined by (8) and (9) on normal operational samples.

3) Dataset for Training AE and AE-GAN

We combine the normal cases and FDIA cases into a dataset for training and evaluating AE and AE-GAN models. To ensure that some topology changes are unseen in the test stage, we construct training and validation datasets with 5% line switching, while the test dataset is constructed with up to 8% line switching. The sizes of datasets for training, validation, and test are set to be 80%, 10%, and 10%, respectively. We do not conduct additional normalization for the dataset because the measurements and states are recorded as per-unit value.

Each item in the dataset includes features and a label as in Table I, depending on the training purpose.

TABLE I
Features and Label of Dataset

Dataset	Feature	Label
AE feature extractor	$\boldsymbol{z}$, $\boldsymbol{x}$	Not applicable
AE-BCV detector	$\boldsymbol{z}$, $\boldsymbol{x}$	FDIA or not
AE-GAN	$\boldsymbol{z}$	Positions of falsified measurement

C. Case Study I: Impact of Proposed Attack Model

We study the performance of our FDIA model based on two aspects: the ability to bypass the BDD; and system-wide impact with limited resources and access to meters.

We categorize attacks into three levels: strong attack (SA), moderate attack (MA), and weak attack (WA). The intensity of attacks is measured by the magnitude of introduced error $\boldsymbol{c}$, determined by the parameter k in models (8) and (9). In untargeted FDIA, the total estimated error is set as k times larger than the sum of absolute values of real estimation. In targeted FDIA, the attacked points are randomly selected and the error on selected points is set as k times larger than base values. The ranges of k are shown in Table II.

TABLE II
Category of FDIA Intensity

Attack level	Range of k
SA	$10\le k\le 40$
MA	$5\le k<10$
WA	$2\le k<5$

1) The performance of bypassing the BDD: BDD works if and only if the residual threshold detection model (5) works successfully. Comparing the deference of residuals between normal and attacked scenarios, we can evaluate whether the attacked measurements can bypass BDD.

We conduct simulations on IEEE 118-bus and 300-bus systems. Each level of attack has been simulated 1100 times. We average the normal operational residual (NOR) and the difference caused by SA, MA, and WA of targeted, and untargeted attack models in Table III. The NOR and difference caused by attacks are calculated by:

$$NOR=\max\left(\left|\boldsymbol{z}-\boldsymbol{H}\hat{\boldsymbol{x}}\right|\right) \tag{16}$$

$$d_{\mathrm{NOR}}=NOR-\max\left(\left|(\boldsymbol{z}+\boldsymbol{a})-\boldsymbol{H}(\hat{\boldsymbol{x}}+\boldsymbol{c})\right|\right) \tag{17}$$

We can see that the differences of FDIA attack at each level are far less than the NOR, indicating the proposed FDIA models are able to bypass BDD.

2) System-wide impact assessment. We compare the performance of our model and the conventional FDIA model [6] on IEEE 118-bus and 300-bus systems. The results are shown in Table IV, indicating that the proposed attack models can affect more state variables with less compromised meters, as compared with the conventional model.

TABLE III
NORMAL RESIDUAL AND DIFFERENCE CAUSED BY UNTARGETED FDIA

Test system	*NOR*	d_{NOR}		
		WA	MA	SA
IEEE 118-bus	2.570	5.3×10^{-8} (TA) 6.9×10^{-8} (UA)	3.7×10^{-8} (TA) 9.6×10^{-7} (UA)	4.3×10^{-8} (TA) 2.8×10^{-8} (UA)
IEEE 300-bus	0.128	5.1×10^{-8} (TA) 3.4×10^{-7} (UA)	8.9×10^{-8} (TA) 8.4×10^{-8} (UA)	1.1×10^{-8} (TA) 6.3×10^{-8} (UA)

Note: TA represents targeted; and UA represents untargeted.

TABLE IV
COMPROMISED METERS AND AFFECTED STATES OF TWO MODELS

Test system	FDIA model	No. of compromised meters	No. of affected states
IEEE 118-bus	Conventional model	60-140	10
	Untargeted FDIA	33-34	117
	Targeted FDIA	18-49	10-48
IEEE 300-bus	Conventional model	50-140	10
	Untargeted FDIA	9-12	299
	Targeted FDIA	10-50	10-35

D. Case Study II: Accuracy of FDIA Detection

We evaluate the accuracy of FDIA detector as follows. ① We test our detectors under various attacks, including targeted and untargeted attacks at three intensity levels. ② For the sake of robustness, we train only one detector for each power system for all types of attacks. Meanwhile, several types of attacks unseen in the training process are included in the test process to verify the robustness of the proposed model. ③ We only use normal measurements to train AE. Using only normal measurements in the training can effectively compress the sample space of AE, which facilitates and accelerates the training. To train the BCV, normal measurements, targeted, and untargeted attacks are considered. Half of the training dataset is generated on the base topology, and the other half is based on topological changes with up to 5% line outages. We also limit the intensity level of the attack for training, where only SA is used in the training dataset. However, we test our detectors using WA, MA, and SA. In other words, the AE-BCV detector can detect less tangible FDIA events in realistic attack scenarios.

The detailed structure and the parameters of the AE-BCV detector are listed in Appendix A. The confusion matrices of the proposed detectors are listed in Table V, with over 95% accuracy on test systems. For a comprehensive comparison, we train two other detectors based on ANN and SVM and test them with the same datasets. From the comparison shown in Table VI, we observe that the proposed model outperforms the other two models.

To evaluate the robustness of the proposed detectors, we focus on the accuracy of the test using unseen test samples. The performance of detectors against unseen MA and WA is listed in Tables VII and VIII. We observe that the AE-BCV detector still outperforms other two detectors with unseen milder attacks. Meanwhile, we also observe that the advantage of AE-BCV expands as the intensity of the attack decreases. These observations indicate that the AE-BCV detector is more robust to the unseen test samples.

TABLE V
CONFUSION MATRIX FOR GENERAL FDIA DETECTION

Bus	Correct rate (%)	False positive (%)	False negative (%)
57-bus	96.7	0.6	2.7
118-bus	98.5	0.1	1.4
415-bus	97.5	0.2	2.3

TABLE VI
AVERAGE PERFORMANCE OF DETECTORS UNDER VARIOUS FDIA

Detector	Accuracy (%)		
	57-bus	118-bus	415-bus
AE-BCV	96.7	98.5	97.5
SVM	86.0	93.3	88.5
ANN	86.7	92.8	87.2

TABLE VII
PERFORMANCE OF DETECTORS AGAINST UNSEEN MA

Detector	Accuracy (%)		
	57-bus	118-bus	415-bus
AE-BCV	98.4	99.2	97.6
SVM	94.0	94.0	92.4
ANN	87.6	93.2	90.0

TABLE VIII
PERFORMANCE OF DETECTORS AGAINST UNSEEN WA

Detector	Accuracy (%)		
	57-bus	118-bus	415-bus
AE-BCV	90.8	95.2	93.2
SVM	50.8	80.0	62.8
ANN	60.4	78.8	60.4

We also make a rough comparison of the detection accuracy between the proposed AE-BCV detector and the detection models in [13], [14] with the IEEE 118-bus system, as the details of the model design of these models are not consistent. We notice that the state-of-the-art deep-learning-based methods achieve over 90% correct rates on the IEEE 118-bus system (92% in [13], and up to 98% in [14]). For unseen topological changes, the proposed detector can still achieve over 90% correct rate on the IEEE 118-bus system with an approximately equivalent performance [13], [14], as shown in Tables VII and VIII.

E. Case Study III: FDIA Localization and Data Recovery

We first visualize the training process of the AE-GAN to show that the Nash-equilibrium is achieved. Next, we analyze the performance of the localization algorithm. Finally, detailed cases of measurement recovery are demonstrated.

The training dataset for AE-GAN only contains normal measurements $\boldsymbol{z}$. And the testing set for localization and re-

covery only consists of measurements under untargeted attacks. We assume that each sample in the testing set has been identified as "under attack" by the detector. The topology change has been considered in the testing set, up to 8% line outages may occur based on the base topology. We have prepared a test set for all kinds of attacks.

1) The visualization of the AE-GAN training: it is crucial to train the discriminator and the generator at the same pace and maintain the confrontation between these two models. Otherwise, the equilibrium will not be achieved.

Figure 5 shows the training loss in IEEE 118-bus system, visualizing the training process of two models. The loss denotes the error between the outputs and expected results of the model in (12). To show the curve more intuitively, we use moving average of loss (MAL) for a sliding window of 10 steps to reflect the overall trend of the loss curve. For both curves, the quicker they decrease, the faster they are trained. After the first 200 steps, the curve of the discriminator decreases, whereas the curve of the generator increases, showing that they are in different training paces. However, the trends of the two curves reverse after 200 steps. Finally, the training pace of the two models gradually balances and reaches Nash-equilibrium.

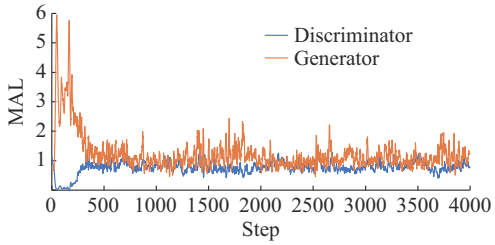

Fig. 5. Moving average of training loss in IEEE 118-bus system.

2) The performance of FDIA localization. The recovery process aims to minimize the search error measured by the distance to L_{distance}. Figure 6 shows Manhattan distance versus iteration of IEEE 118-bus system. We observe that Manhattan distance curves decrease, indicating the recovery algorithm can effectively discover more accurate measurements and make improvements on the localization results. The recovered measurement of IEEE 118-bus system is shown in Fig. 7. The FDIA localization results of IEEE 118-bus system and CSG 415-bus system are shown in Table IX, where the correct rate, positive false rate, and negative false rate in the table are normalized by the number of attack/unattacked points, rather than the number of total measurements.

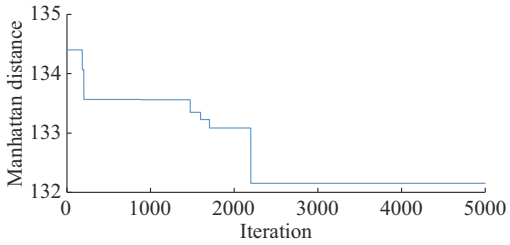

Fig. 6. Manhattan distance versus iteration of IEEE 118-bus system.

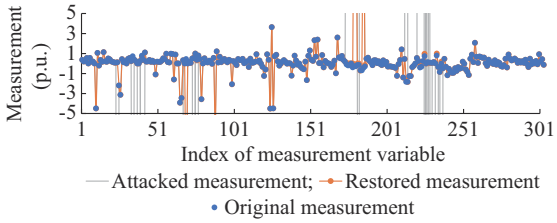

Fig. 7. Recovered measurements of IEEE 118-bus system.

TABLE IX
FDIA LOCALIZATION RESULT WITH VARIED INTENSITY OF ATTACK

Type	Intensity of attack	Correct rate (%)	Positive false rate (%)	Negative false rate (%)
IEEE 118-bus	SA	83.92	6.91	16.08
	MA	84.96	11.36	15.04
	WA	80.37	12.90	19.63
CSG 415-bus	SA	85.69	7.31	14.31
	MA	85.16	9.07	14.84
	WA	81.28	11.63	18.72

3) The performance of measurement recovery. As shown in Table X, the mean error of recovered measurements caused by FDIA drops significantly. For $\boldsymbol{z}(j_0)=0$, we will set $\boldsymbol{z}(j_0)=mean(|\boldsymbol{z}|)$. Otherwise, $|\boldsymbol{a}(j_0)/\boldsymbol{z}(j_0)|$ will become infinite. Therefore, the adverse impact of FDIA is mitigated.

TABLE X
ERROR OF MEASUREMENTS BEFORE OR AFTER RECOVERY WITH VARIED ATTACK INTENSITY

Type	Intensity of attack	Error (mean $\lvert\boldsymbol{a}/\boldsymbol{z}\rvert$) caused by FDIA	Error (mean $\lvert\boldsymbol{a}/\boldsymbol{z}\rvert$) after recovery
IEEE 118-bus	SA	16.50	0.85
	MA	5.70	0.61
	WA	3.15	0.46
CSG 415-bus	SA	21.69	0.92
	MA	5.99	0.53
	WA	2.68	0.32

F. Computation Time

The computation time is listed in Table XI.

TABLE XI
COMPUTATION TIME

Type	Detection		AE-GAN (s)	Localization and recovery time (s)
	Training time (s)	Classification time (s)		
IEEE 118-bus	441.56	0.011	324.85	14.69
CSG 415-bus	594.44	0.013	406.39	25.47

It is shown that, despite that the offline training of the classifier and the AE-GAN model are time-consuming, the computation time for online classification, localization, and recovery is short. Therefore, the proposed methodology is efficient for online defense against FDIA.

It is seen that the computation time increases slightly as the scale of the power system expands. On one hand, the AE-BCV detector is designed to capture the deep characteristics of joint PDF of $(\hat{x}, z)$. Therefore, FDIA can be detected if the falsified measurements cause the density of joint probability derivate from normal distributions. As a result, the training dataset for the AE-BCV detector is not necessarily large. On the other hand, note that the size of the neural network in the AE-BCV detector and AE-GAN (as shown in Appendix A) is medium, the training time of these neural networks depend primarily on the size of training dataset. Therefore, the training time will increase mildly as the size of the power system grows.

VII. Conclusion

To address the challenges of FDIA, we design a deep-learning-based methodology for detecting and locating FDIA using DC power flow and recovering the falsified measurement/state variables. Importantly, we aim to improve the generalization of the proposed methodology with uncertainty under operational conditions.

To this end, we first propose two attack models to demonstrate the system-wide impact of massive FDIA even with limited access to meter and sensors. Second, we design a robust AE-BCV detector to learn the deep feature of the joint probability function of state variables and measurements, and then classify FDIA by the MAP rule. The proposed detector outperforms the existing methods with over 95% detection accuracy for FDIA, even if the system structure, i.e., the operational topology of the power grid in the application context, is unseen to the detector in the training stage. Third, we design an AE-GAN to generate a diverse dataset containing measurement samples under the normal operational conditions. Subsequently, we design a pattern match algorithm to recover falsified measurements from the dataset based on Manhattan distance. From comprehensive case studies, the proposed methodology achieves an 80% localization accuracy and recovers the state variables close to the true values.

The advantages of the proposed methodology are as follows. First, the proposed AE-BCV detector can be directly applied with AC power flow model, as we train the joint probability distribution of states and measurements. Second, the proposed AE-GAN can generate a sufficiently large candidate set for falsified measurement localization without model collapse.

Our future research can be extended in the following aspects. First, the attack model can be modified to consider AC power flow model. Second, the algorithmic efficiency for localization and recovery can be improved by constructing a refined candidate set. Third, the proposed methodology can be applied in the context of dynamic SE with asymmetric information possessed by cyber attackers and defenders.

Appendix A

A. Detection Test Based on a Large Test System

This case is to verify the scalability of the model. The 1354-bus system in MATPOWER is used in the simulation. The topological changes are not considered, whereas other setting is the same as that in case studies. The positive false rate is 0.1% and the negative false rate is 4.7%.

B. Detailed Structure and Parameter Setup

The detailed structure and the parameters of the AE-BCV detector and AE-GAN are listed in Table AI and Table AII. The values of τ_1 for localization are 0.01 for IEEE 118-bus system and 0.5 for CSG 415-bus system.

TABLE AI
Neural Network Parameters of AE-BCV Detector

Type	System	Total hidden layers	Number of neurons in one layer		
			Input	Hidden	Output
Encoder	57-bus	14	192	200	80
	118-bus	14	420	200	80
	415-bus	14	1455	200	120
	1354-bus	14	4697	200	120
Decoder	57-bus	14	80	200	192
	118-bus	14	80	200	420
	415-bus	14	120	200	1455
	1354-bus	14	120	200	4697

TABLE AII
Setup of Parameter for AE-GAN

Type	Total hidden layers	Number of neurons in each hidden layer	Learning rate	Number of neurons of AE code
Encoder	2	1000	0.0001	
Decoder	2	1000	0.0001	120
Discriminator	2	500	0.0005	

References

[1] L. Che, X. Liu, Z. Li *et al*., "False data injection attacks induced sequential outages in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1513-1523, Mar. 2019.

[2] G. Liang, S. R. Weller, J. Zhao *et al*., "The 2015 Ukraine blackout: implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317-3318, Jul. 2017.

[3] H. Wang, J. Ruan, B. Zhou *et al*., "Dynamic data injection attack detection of cyber physical power systems with uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5505-5518, Oct. 2019.

[4] A. Sargolzaei, K. Yazdani, A. Abbaspour *et al*., "Detection and mitigation of false data injection attacks in networked control systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281-4292, Jun. 2020.

[5] F. Li and Y. Tang, "False data injection attack for cyber-physical systems with resource constraint," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 729-738, Feb. 2020.

[6] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 1-33, May 2011.

[7] O. Kosut, L. Jia, R. J. Thomas *et al*., "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645-658, Dec. 2011.

[8] Y. Song, X. Liu, Z. Li *et al*., "Intelligent data attacks against power systems using incomplete network information: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 4, pp. 630-641, Jul. 2018.

[9] G. E. Constante-Flores, A. J. Conejo, J. Wang, "Sensitivity-based vulnerability assessment of state estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 886-896, Jul. 2016.

[10] N. Živković and A. Sarić, "Detection of false data injection attacks us-

ing unscented Kalman filter," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 847-859, Sept. 2018.
[11] M. Ansari, V. Vakili, B. Bahrak *et al*., "Graph theoretical defense mechanisms against false data injection attacks in smart grids," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 860-871, Sept. 2018.
[12] B. Chen, H. Li, and B. Zhou, "Real-time identification of false data injection attacks: a novel dynamic-static parallel state estimation based mechanism," *IEEE Access*, vol. 7, pp. 95812-95824, Jul. 2019.
[13] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505-2516, Sept. 2017.
[14] J. Yu, Y. Hou, and V. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3271-3280, Jul. 2018.
[15] J. Cao, D. Wang, Z. Qu *et al*., "A novel false data injection attack detection model of the cyber-physical power system," *IEEE Access*, vol. 8, pp. 95109-95125, May 2020.
[16] Y. An and D. Liu, "Multivariate Gaussian-based false data detection against cyber-attacks," *IEEE Access*, vol. 7, pp. 119804-119812, Aug. 2019.
[17] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2218-2234, May 2020.
[18] X. Wang, X. Luo, M. Zhang *et al*., "Detection and isolation of false data injection attacks in smart grid via unknown input interval observer," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3214-3229, Apr. 2020.
[19] X. Luo, Y. Li, X. Wang *et al*., "Interval observer-based detection and localization against false data injection attack in smart grids," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 657-671, Jan. 2021.
[20] S. Wang, S. Bi, and Y. Zhang, "Locational detection of false data injection attack in smart grid: a multi-label classification approach," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8218-8227, Sept. 2020.
[21] M. Mohammadpourfard, A. Sami, and Y. Weng, "Identification of false data injection attacks with considering the impact of wind generation and topology reconfigurations," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1349-1364, Jul. 2018.
[22] H. Wang, J. Ruan, G. Wang *et al*., "Deep learning-based interval state estimation of AC smart grids against sparse cyber attacks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4766-4778, Nov. 2018.
[23] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2239-2248, Sept. 2017.
[24] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1665-1676, Jul. 2014.
[25] X. Huang, Z. Qin, and H. Liu, "A survey on power grid cyber security: from component-wise vulnerability assessment to system-wide impact analysis," *IEEE Access*, vol. 6, pp. 69023-69035, Jun. 2018.
[26] Y. Li, Y. Wang, and S. Hu, "Online generative adversary network based measurement recovery in false data injection attacks: a cyber-physical approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2031-2043, Mar. 2020.
[27] R. Deng, G. Xiao, R. Lu *et al*., "False data injection on state estimation in power systems–attacks, impacts, and defense: a survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411-423, Apr. 2017.
[28] A. Tajer, S. Sihag, and K. Alnajjar, "Non-linear state recovery in power system under bad data and cyber attacks," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 1071-1080, Sept. 2019.
[29] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262-282, Feb. 2000.
[30] Q. Yang, J. Yang, W. Yu *et al*., "On false data-injection attacks against power system state estimation: modeling and countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 717-729, Mar. 2014.
[31] Electric Power Research Institute, USA. (2016, Feb.). Electric power system resiliency: challenges and opportunities. [Online]. Available: https://www.naseo.org/Data/Sites/1/resiliency-white-paper.pdf
[32] Z. Chen, C. K. Yeo, B. S. Lee *et al*., "Autoencoder-based network anomaly detection," in *Proceedings of 2018 Wireless Telecommunications Symposium (WTS)*, Phoenix, USA, Apr. 2018, pp. 1-5.
[33] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, no. 11, pp. 1771-1782, Nov. 2000.
[34] Z. Lin, A. Khetan, G. Fanti *et al*., "Pacgan: the power of two samples in generative adversarial networks," in *Proceedings of Advances in Neural Information Processing Systems*, Montreal, Canada, Dec. 2018, pp. 1498-1507.
[35] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza *et al*., "Generative adversarial nets," in *Proceedings of International Conference on Neural Information Processing Systems*, Montreal, Canada, Dec. 2014, pp. 2672-2680.

Xiaoge Huang received the B. E. degree in electrical engineering from Guangxi University, Nanning, China, in 2018. From August 2018 to August 2019, he was a Research Assistant with School of Electrical Engineering, Guangxi University. His major research interests include power system cyber-security, renewable energy integration, optimal power flow, and deep learning.

Zhijun Qin received the B.E. and M.Sc. degrees from Huazhong University of Science and Technology, Wuhan, China, in 2000 and 2003, respectively, and the Ph.D. degree from the University of Hong Kong, Hong Kong, China, in 2015, all in electrical engineering. He worked as a Postdoctoral Research Fellow with the University of Hong Kong from 2015 to 2016. He is currently working as an Associate Professor with School of Electrical Engineering, Guangxi University, Nanning, China. His research interests include power system resilience, renewable energy integration, optimal power flow, and power system cyber-security.

Ming Xie received the B.E. degree in applied electronic technology from Guangxi Normal University, Nanning, China, in 2001, the M.Sc. degree in computer software and theory from Wuhan University, Wuhan, China, and the Ph.D. degree in computer system architecture from Wuhan University. Currently, he is working as a Senior Engineer with the Information Center, Guangxi Electric Power Grid Co., Ltd., Nanning, China. His research interest includes power system cyber-security.

Hui Liu received the M.S. degree in 2004 and the Ph.D. degree in 2007 from the College of Electrical Engineering at Guangxi University, Nanning, China, both in electrical engineering. He worked in Tsinghua University, Beijing, China, as a postdoctoral fellow from 2011 to 2013 and in Jiangsu University as a Faculty Member from 2007 to 2016. He visited the Energy Systems Division at Argonne National Laboratory, Argonne, USA, from 2014 to 2015. He joined the School of Electrical Engineering at Guangxi University in 2016, where he is a Professor and Deputy Dean. He is an Editor of the IEEE Transactions on Smart Grid and the IEEE Power Engineering Letters. He is also an Associate Editor of the IET Smart Grid and the IET Generation, Transmission & Distribution. His research interests include power system optimization, power system stability and control, electric vehicles, integrated energy systems, demand response, etc.

Liang Meng received the B.E. degree in electronic information engineering from Lijiang College, Guangxi Normal University, Lijiang, China, in 2010, and the M. Sc. degree in pattern recognition and intelligent system from Guangxi Normal University, Nanning, China, in 2014. Currently, he is an Engineer with the Information Center, Guangxi Electric Power Grid Co., Ltd., Guilin, China. His research interest includes power system cyber-security.