Attack Power System State Estimation by Implicitly Learning the Underlying Models

Napoleon Costilla-Enriquez, *Graduate Student Member, IEEE*, and Yang Weng, *Senior Member, IEEE*

***Abstract*—False data injection attacks (FDIAs) are a real and latent threat in modern power systems networks due to the unprecedented integration of data acquisition systems. It is of utmost importance to understand attacking mechanisms to design countermeasures. To successfully deploy a FDIA, most past FDIA strategies need privileged power system information, which is carefully held by the power system operator. Newer approaches circumvent this issue by solely relying on intercepted measurement data, but they lack mathematical warranties of succeeding. This paper exposes power systems' vulnerability by showing that it is possible to deploy an attack without confidential information and, at the same time, to have a high probability of being successful. We present a scheme that learns (1) the implicit power system measurement distribution and (2) a surrogate of the unknown state estimator model. The proposed framework utilizes a Wasserstein generative adversarial network to learn the measurement distribution and an autoencoder to learn the unknown state estimator model. Additionally, we present a convergence proof that ensures that the proposed framework converges to the power system measurement distribution. The proposed method is demonstrated to be successful via extensive simulation on IEEE 9-, 14-, 57-, 118-, and 300-bus test cases.**

***Index Terms*—False data injection attack, state estimation, no system information, adversarial examples, Wasserstein generative adversarial networks (WGANs), autoencoder (AE).**

I. Introduction

DATA revolution takes place worldwide in different disciplines, including power systems. To provide a robust grid with new but diversified components, modern power grids are on the road to integrate unprecedented real-time and offline data for monitoring, control, and protection. However, this new data-driven outlook makes the power grid more vulnerable than ever to cyber-attacks with dire consequences. For instance, the power system operator may take wrong corrective actions that can cause a blackout; wrong actions can also cause inaccurate energy prices in a real-time electricity market [1], [2].

To better protect the system, it is essential to understand potential attack mechanisms. Among various attack categories [3], [4], [5], False Data Injection Attacks (FDIA) gained the attention of the power system community after the work in [6], which showed that unobservable attacks against DC State Estimators (SE) are possible. In this type of attack, the attacker modifies measurement data such that the estimated states are different from the real ones [7], [8]. These first works have the following assumptions, which may be impractical:

(i) The attacker has access to the entire network information (e.g., line parameters, grid topology, state estimator model, and estimated states) [9], [10]. It is impractical to think that an attacker can gather all this data without an insider in the Independent System Operator (ISO). Since this information is guarded by the power system operator it is difficult for an attacker to have this knowledge.

(ii) These first studies rely upon the DC power flow model when power system operators use the AC power flow model in real-world settings. The reason, AC-based FDIAs are harder to design and deploy due to the inherent complexity of the nonlinear power flow equations [11], [12].

Subsequent work relaxed the first assumption. Specifically, [2], [6], [13], [14], [15], [16], [17] propose various frameworks to design FDIAs with only partial network information, but they still rely on a DC-based model. To relax the DC model's assumption, a few studies have focused on FDIA with an AC-based model [9], [18], [19]. However, all the aforementioned approaches construct an attack vector relying upon the power system underlying information; we can call these techniques model-based FDIAs.

Later works showed that it is also possible to deploy FDIA without knowing privileged power system information such as power system parameters and topology or the state estimator model. The only needed information to deploy a FDIA are the power system measurements, and we classify these kinds of attacks as model-free FDIAs. In modern power system networks, the information is sent via remote terminal units that are designed avoid system intrusion [20], [21]. However, conventional approaches such as security software and firewalls could be insufficient to protect the system against breaches and cyber threats [12]. For example, in 2015, a cyber-attack was successfully deployed on Ukraine's electricity infrastructure. Around one year before the attack, the attackers gained access to multiple industrial networks by using the malware tool

Manuscript received 23 August 2021; revised 24 January 2022 and 16 June 2022; accepted 4 August 2022. Date of publication 10 August 2022; date of current version 22 December 2022. This work was supported in part by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) through the solar Energy Technologies Office under Award DE-EE0009355; in part by the Department of Energy under Grant DE-AR00001858-1631; in part by the National Science Foundation (NSF) under Grant ECCS-1810537 and Grant ECCS-2048288; and in part by the Fondo de Sustentabilidad Energetica CONACYT-SENER, Mexico, under Grant 708642. Paper no. TSG-01352-2021. *(Corresponding author: Yang Weng.)*

The authors are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: ncenriqu@asu.edu; yang.weng@asu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2022.3197770.

Digital Object Identifier 10.1109/TSG.2022.3197770

BlackEnergy 3 (BE), [22]. This malware enables unauthorized network access with valid (stolen) user credentials to move laterally across internal utilities' system. In this incident, the attackers gained access to targeted networks using weaponized Microsoft Office files by embedding BE in Visual Basic macro scripts. This latent risk has been recognized by the National Academies of Sciences, Engineering, and Medicine [23]. In the same work, they conclude that the United States' power system network is vulnerable to cyber-attacks. Thus, for an attacker, it would be feasible to collect sensor measurements by exploiting the protection schemes [12].

The authors in [24] showed that it is possible to deploy a stealthy FDIA by using principal component analysis (PCA). The extension of this work in [25] proposed a geometric approach to carry out a FDIA based only on power system measurements. The authors in [26] proposed a data-driven attack that learns the system operation subspace from measurements around a linearized nominal state. The work in [27] presented a zero-parameter information attack that only requires power system's topology information. The works in [28], [29] employed machine learning techniques to carry out a FDIA. Specifically, they trained a generative adversarial network (GAN) to generate tampered power system measurements that will be stealthy with high probability. While the works in [28], [29] and our work use generative adversarial networks (GANs) to carry out a FDIA, our approach has some important differences. Both works in [28], [29] use the DC linear power flow model. In contrast, our proposed approach uses the AC non-linear power flow model. Whereas the work in [29] requires normal and tampered measurements to train a conditional adversarial network (cGAN), our approach only requires normal measurements, which is a more reasonable assumption. This means that our attack is more appealing at the level of the information needed to train our model.

The difficulty with the model-free FDIAs is that it is hard to ensure that the model-free approach truly captures some properties of the power system model to bypass tests, such as the Chi-squared test to obtain the trust from energy management systems. To show the power system vulnerability under this setting, we introduce a data-driven approach that generates tampered measurements with the desired properties to deploy a FDIA, and at the same time, to have mathematical guarantees about the model accuracy. We achieve this goal by (1) implicitly learning the power system measurement distribution from data; and (2) learning a proxy model for the unknown state estimator.

Specifically, we aim to design a flexible model that captures the complex underlying interactions in the power system to learn the measurement distribution from data. Nonparametric methods are flexible since they build models from data making as few assumptions as possible, which usually means utilizing statistical models that are infinite-dimensional [30]. While these type of models are flexible by keeping the underlying assumptions as weak as possible, they are computationally demanding due to the required increment of number of parameters [31], [32]. For example, the work in [33] shows that their nonparametric model grows in complexity as additional data is used to train the model. As real power systems could have thousands of buses and data measurements from many years, the number of parameters needed in non-parametric models are computationally intractable [31]. Therefore, we choose parametric models, which can be designed with a fixed number of parameters that depend upon the specific problem. In recent years, these parametric models have had tremendous success in the ML community because they can learn complex high-dimensional distributions (for example, images in high resolution). In power systems, for example, the work in [34] physics-informed parametrized neural networks (PINN) to learn the underlying power grid's parameters. In the parametric models, we introduce a framework utilizing generative adversarial networks (GANs) to learn the power system measurement distribution to create spurious measurements to deploy a FDIA, as GAN's loss function is fully specified. As a comparison, variational autoencoders (VAEs)'s loss function is only the evidence lower bound (ELBO), which is hard to be embedded into other learning. Even more importantly, we can present mathematical proof to show that the GAN reliably learns the power system measurement distribution. In specific, we use the Wasserstein Generative Adversarial Network (WGAN), which is guaranteed to converge under mild assumptions to the actual observed distribution [35].

In addition, to mimic the data distribution, one knowledge we do have is the form of residual error test. Therefore, we propose to boost our attack capability by learning the state estimator model for the residual error test. However, learning the state estimator model directly is difficult because neither the power system nor the state estimator is known. To circumvent this issue, we use a surrogate model to mimic the state estimator. The residual error test and an autoencoder (AE) share the same mathematical structure. Thus, an AE can be trained as a proxy to mimic the state estimator. We leverage this similarity and employ an AE as a proxy for the residual test error. Specifically, in our proposed scheme, we include this proxy as a regularization term, which helps to improve the quality of the created tampered measurements. Finally, a second regularization term is added to maximize the impact of the attack. Whereas the model-based attacks need the complete network information (e.g., line parameters, grid topology, state estimator model, and estimated states), our proposed model-free approach needs a dataset of the measurements of the considered network to work. And such a data set does not need all the measurements to be included, which is another advantage.

The performance of the proposed model-free FDIA is verified by simulations on the standard IEEE 9-, 14-, 57-, 118-, and 300-bus test networks. Also, to contrast the differences and advantages between our approach and the existing ones in the literature, we carry out comparisons between our proposed FDIA and three other successful methods reported in [9], [18], [25]. These results show that our proposed model-free is successful. Specifically, our proposed model-free attack tampers measurements in a way that can fool the power system operator with high probability.

The rest of the paper is organized as follows: Section II introduces the problem formulation, Section III presents our proposed model-free FDIA model, Section IV presents the convergence proof, Section V shows numerical experiments, and Section VI concludes the paper.

II. Problem Formulation

To show the proposed model-free FDIA attack, we first review the model-based approaches based on AC state estimation.

A. State Estimation With AC Power Flow Model

State estimation (SE) infers the state variables (i.e., voltage angles and voltage magnitudes) $\mathbf{x} = (x_1, \ldots, x_n)$ from a set of measurements $\mathbf{z} = (z_1, \ldots, z_m)$ [36], where n is the number of buses or nodes in the grid, and m is the number of measurements. Mathematically, we can describe the problem as $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$, where $\mathbf{h}(\cdot)$ is the physical (non-linear) relationship between state variables and measurements, and $\mathbf{e}$ is a vector that represents white noise from the collected measurements (e.g., SCADA or PMU). In practice, measurements are collected and sent to the power system operator, which obtains the estimated states $\hat{\mathbf{x}}$ by solving [37], [38]:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathbf{z} - \mathbf{h}(\mathbf{x}))^T \mathbf{W}^{-1} (\mathbf{z} - \mathbf{h}(\mathbf{x})) = \mathrm{SE}(\mathbf{z}), \quad (1)$$

where, for compactness, we define the state estimator operator $\mathrm{SE}(\cdot)$. The input of this operation is a vector of measurements and its output are the estimated states. However, the vector of measurements $\mathbf{z}$ may contain bad or wrong data due to telecommunication failures, meter errors, or even FDIAs [10], [39]. To estimate the states with confidence, the SE possesses a Bad Data Detector (BDD) module to detect and filter suspicious data.

1) Bad Data Detector (BDD): The measurement errors are assumed to follow a Gaussian distribution $e_i \sim \mathcal{N}(0, \sigma_i)$ [39] (where σ_i is the standard deviation of the i-th measurement). Therefore, the squared measurement residual error $\mathbf{r} = \|\mathbf{z} - \hat{\mathbf{z}}\|_2^2$ follows a Chi-square distribution χ_k, where k represents the number of independent variables in the power system, and $\hat{\mathbf{z}} = \mathbf{h}(\hat{\mathbf{x}})$ is the vector of estimated measurements. Then, the presence of errors in the measurements can be detected with the Chi-square test (or residual error test) [39], [40]. This test works as follows:

(i) Select the detection confidence probability p (e.g., 0.95), and compute its associated threshold value $\tau = \chi^2_{k,p}$ with $p = \Pr(J(\hat{\mathbf{x}}) \leq \chi^2_{k,p})$.

(ii) Compute the normalized measurement error $J(\hat{\mathbf{x}}) = \sum_{i=1}^{m} (z_i - h_i(\hat{x}_i))^2 / \sigma_i^2$.

(iii) If the inequality in (2) holds, bad data will be suspected, or else the measurements are assumed to be free of bad data.

$$J(\hat{\mathbf{x}}) \geq \tau. \quad (2)$$

B. Model-Based FDIA in AC State Estimation

A FDIA modifies the estimated states $\hat{\mathbf{x}}$ or measurements $\hat{\mathbf{z}}$ by changing the original SCADA and PMU measurements $\mathbf{z}$ with a maliciously tampered measurement vector, that is, $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$, where $\mathbf{a}$ is an attack vector. The attacker designs this attack vector to compromise the system's reliability by creating a wrong state estimate. For a FDIA to be successful, it must circumvent the bad data detector (2) [41]. The assumptions in the literature for a model-based FDIA about the attacker's knowledge are the following [6], [9], [42]: (1) the attackers can intercept and alter the power system measurements that are used to obtain the estimated states in the grid; (2) the attacker has access to the power system model, which includes transmission line parameters and topology information; and (3) the attacker possess the SE model or can obtain the estimated states of the network. Under these strong assumptions, the attacker would be able to launch a perfect FDIA [10]. In this perfect FDIA, the attacker can define the attack vector as $\mathbf{a} = \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c}) - \mathbf{h}(\hat{\mathbf{x}})$, where $\mathbf{c}$ is the vector of changes in the estimated states. In this scenario, if the original measurements $\mathbf{z}$ can pass the residual-based bad data detector test in (2), the corrupted measurements $\mathbf{z}_a$ will also pass this test [9].

The work in [9] proposed an FDIA needing only partial power system information. In this context, there are two types of variables. (1) Measurements and state variables that are not altered, which are denoted with subscript 1, $\hat{\mathbf{x}}_1$ and $\mathbf{z}_1 = \mathbf{h}_1(\cdot)$. (2) Measurements and state variables that are maliciously altered, which are denoted with subscript 2, $\hat{\mathbf{x}}_2$ and $\mathbf{z}_2 = \mathbf{h}_2(\cdot)$. If an attacker constructs the attack vector as

$$\mathbf{a}_2 = \mathbf{h}_2\big(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 + \mathbf{c}\big) - \mathbf{h}_2\big(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2\big), \quad (3)$$

the tampered measurements will have the same residual error as the real ones. Note that to create the attack vector in (3), the attacker must know the estimated values of the state variables appearing in $\mathbf{h}_2$, which is still a strong assumption. There are other types of FDIA. For example, if $\mathbf{a} \neq \mathbf{h}(\hat{\mathbf{x}} + \mathbf{c}) - \mathbf{h}(\hat{\mathbf{x}})$ but (2) holds, then the attack is called a generalized FDIA [43].

III. Proposed Model-Free FDIA

Contrary to the model-based FDIAs, the model-free models only make one assumption [24], [25], [28], [29]: The attackers can intercept and alter the power system measurements that are used to obtain the estimated states in the grid. So, in this section, we show a theoretically sound method to deploy a FDIA by only using the power system measurements. If we want to deploy an attack without any underlying power system knowledge, we have to learn an implicit model through observations, that is, from power system measurements (SCADA and PMU). This implicit model should capture the inherent non-linearity relationships between different measurements based on residual error tests. Also, this model should be able to create new tampered measurements such that they are overlooked by the power system operator but change the estimated states and measurements. To summarize, we present a data-driven approach based on a WGAN with two regularization terms. First, the measurement distribution is learned with the WGAN, $\mathbf{z} + \mathbf{e}$. Second, to pass the residual error test, a proxy of the unknown SE model is embedded into the WGAN as a regularization term, $\mathbf{h}(\mathbf{z})$. Finally, a regularization term is added to maximize the attack impact.

A. Learning the Measurement Distribution

Reference [44] introduced the idea of generative adversarial networks, which revolutionized the machine learning (ML) field. GAN is a framework to teach a Deep Learning (DL)

model the implicit training data distribution so that we can sample from it and generate new data from that same distribution; in our case, the power system measurement distribution. Specifically, rather than sampling directly from an (assumed) parametric distribution, the target random variable is generated as a deterministic transformation of a simple, independent noise source, for instance, a Gaussian distribution. GANs are made of two distinct models, a generator and a discriminator. Formally, the minimax objective of the GAN is

$$\min_G \max_D \quad \mathbb{E}_{\mathbf{z}\sim\mathbb{P}_r}\mathbb{E}_{\boldsymbol{\lambda}\sim\mathbb{P}_\lambda}\big[\log D(\mathbf{z}) + \log(1 - D(G(\boldsymbol{\lambda})))\big], \quad (4)$$

where D is a discriminative network, G is a generative network, $\mathbb{P}_r$ is the real data distribution, and $\boldsymbol{\lambda}$ is the latent space, which it is sampled from an independent distribution $\mathbb{P}_\lambda$; that is, $\boldsymbol{\lambda} \sim \mathbb{P}_\lambda$ (usually a Gaussian distribution).

However, GANs have some issues, such as vanishing gradient and the lack of guarantee to convergence. The work in [45] presented the Wasserstein GAN (WGAN) that solves these issues. Also, WGANs possess stronger mathematical guarantees. For example, the authors in [35] proved that (under mild assumptions) the generator in the WGAN will converge to the true data distribution $\mathbb{P}_r$. Therefore, in this work, we will use this type of WGAN. These models are made of two distinct neural networks, a generator G and a discriminator D (or critic). The minimax objective of the WGAN is

$$\min_G \max_{D\in\mathcal{D}} \quad \mathbb{E}_{\mathbf{z}\sim\mathbb{P}_r}\mathbb{E}_{\boldsymbol{\lambda}\sim\mathbb{P}_\lambda}[D(\mathbf{z}) - D(G(\boldsymbol{\lambda}))], \quad (5)$$

where $\mathcal{D}$ is the set of 1-Lipschitz functions [45]; $\mathbb{P}_r$ is the real data distribution; $\boldsymbol{\lambda}$ is known as the latent space, and it is sampled from an independent distribution $\mathbb{P}_\lambda$. The generator G learns the real distribution $\mathbb{P}_r$, which, in our context, this real distribution is the set of historical observed measurements $\mathcal{Z} = \{\mathbf{z}_i \in \mathbb{R}^m\}_{i=1}^L$ (where L is the number of elements in the dataset), where $\mathbf{z}_i = \mathbf{h}(\mathbf{x}_i) + \mathbf{e}_i$. In other words, G implicitly learns to generate samples from the underlying model $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$.

B. Learning the State Estimator Model

To gain trust from the power system operator, the created tampered measurements, $\tilde{\mathbf{z}} = G(\boldsymbol{\lambda})$, must pass the residual error test in (2). This residual error for the tampered measurements is given as

$$\tilde{\mathbf{r}} = \left\|\tilde{\mathbf{z}} - \hat{\tilde{\mathbf{z}}}\right\|^2, \quad (6)$$

where $\hat{\tilde{\mathbf{z}}} = \mathbf{h}(\hat{\tilde{\mathbf{x}}})$ is the vector of estimated tampered or fake measurements, and $\hat{\tilde{\mathbf{x}}} = \mathrm{SE}(\tilde{\mathbf{z}})$ is the vector of estimated states from tampered measurements. As (2) suggests, the smaller the residual error $\tilde{\mathbf{r}}$, the bigger the probability of passing the test for a given tampered measurement, $\tilde{\mathbf{z}}$. In other words, a given vector of tampered measurements, $\tilde{\mathbf{z}}$, should produce a similar estimated vector, $\hat{\tilde{\mathbf{z}}} = \mathbf{h}(\hat{\tilde{\mathbf{x}}})$. However, in this model-free approach, we do not have access to the state estimator model $\mathbf{h}(\cdot)$. This non-linear function $\mathbf{h}(\cdot)$ can be thought of as a mapping from the measurement space to the estimated measurement space. For a vector of real measurements, the estimated measurements will be similar so that the residual error is low. This state estimator function $\mathbf{h}(\cdot)$ is unknown. Still, given its properties, it is possible to learn it from data and create a proxy to impose the same behavior in the tampered measurements.

The residual error expression in (6) resembles the loss function from an autoencoder (AE). Thus an AE model is a natural option to learn a proxy model of the unknown state estimator function $\mathbf{h}(\cdot)$. An autoencoder is a neural network that aims to produce or replicate its input to its output [46]. To do this, the autoencoder is trained to learn an encoding for a particular distribution and then with such encoding, learn to reconstruct the input distribution. To learn a meaningful encoding, the model's architecture prioritizes which traits from the input should be learned. By this process the autoencoder learns to ignore superfluous data, which could be noise. We will see how this autoencoder property improves the generation of fake measurements in Section V-E. Mathematically, the autoencoder is represented as a function, that is, $\mathrm{AE}(\cdot)$, and it is trained with the squared loss function:

$$\|\mathbf{z} - \mathrm{AE}(\mathbf{z})\|^2. \quad (7)$$

A trained AE with real measurements with (7) will learn the unknown function $\mathbf{h}(\cdot)$ that will minimize the residual error in (6). Once the autoencoder is trained (denoted as AE^*), the loss function in (7) can be embedded into (5) to incentivize the generation of tampered measurements that will produce similar estimated measurements, and thus lower the residual error. This can be done by adding the regularization term $\|\tilde{\mathbf{z}} - \mathrm{AE}^*(\tilde{\mathbf{z}})\|_2^2$ in (5):

$$\min_G \max_{D\in\mathcal{D}} \quad \mathbb{E}_{\mathbf{z}\sim\mathbb{P}_r}\mathbb{E}_{\boldsymbol{\lambda}\sim\mathbb{P}_\lambda}\Big[D(\mathbf{z}) - D\big(\tilde{\mathbf{z}}\big) + \|\tilde{\mathbf{z}} - \mathrm{AE}\big(\tilde{\mathbf{z}}\big)\|^2\Big], \quad (8)$$

where $\tilde{\mathbf{z}} = G(\boldsymbol{\lambda})$.

C. Maximize the FDIA Impact

The WGAN in (8) implicitly learns the underlying model that generates the observed data [44], [45]. To train a WGAN with (8), we need to sample $\mathbf{z}$ from the true data distribution $\mathbb{P}_r$. However, the generator in (8) conventionally takes a random signal as input and maps it to the true data distribution space; that is, $\boldsymbol{\lambda} \sim \mathbb{P}_\lambda$, where $\mathbb{P}_\lambda$ is usually a Gaussian distribution. This means that we do not have any control over the created fake measurements. To successfully attack a power system, we want these fake measurements, produced by our WGAN, to create different states from the actual ones. The attacker can only see and modify observed measurements. Thus, the attacker can attempt to markedly change the unobservable states by stealthy and sizeably manipulating the intercepted measurements to perform a successful FDIA. To accomplish this, we need to generate tampered measurements from the observed ones.

If we want to generate tampered measurements from the observed ones, rather than using a random distribution $\mathbb{P}_\lambda$ as latent space to feed our generator, we use the power system measurements as input to the generator, that is, $\mathbb{P}_\lambda = \mathbb{P}_r$. The result is that the generator's latent space is not fed with an

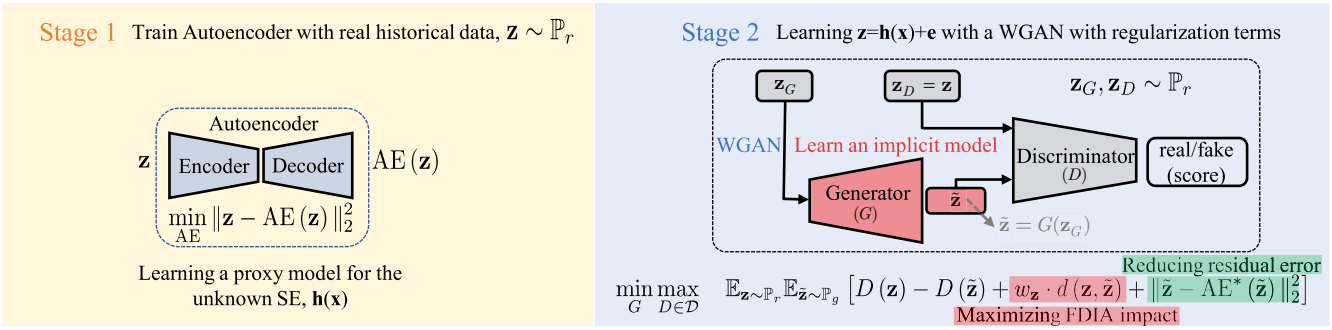


Fig. 1. Proposed model-free architecture with a WGAN and two regularization terms to deploy an FDIA.

arbitrary random distribution: it is fed with the power system measurement distribution. Specifically, we are conditioning the WGAN with respect to the actual measurement vector $\mathbf{z}$, as depicted in Fig. 1. This is desirable because in this way, rather than creating tampered measurements from an arbitrary distribution, they are constructed based on the observed ones. Furthermore, the created tampered measurements will differ from those received as input due to a regularization term that we include in our model, as we explain below.

To successfully deploy an FDIA, we want to incentivize the generator to construct measurements that will produce different measurements from those received as input. This will provoke the SE with high a likelihood to produce erroneous estimated states, the main objective in a FDIA. To accomplish this, we can incentivize the model to generate such fake measurements with the regularization term $w_{\mathbf{z}} \cdot d(\mathbf{z}, \tilde{\mathbf{z}})$ in (9) (the first regularization term in Fig. 1 in red), where $\tilde{\mathbf{z}} = G(\mathbf{z})$, $d(\cdot)$ is a distance function (e.g., mean squared or mean absolute distance), $d(\mathbf{z}, \tilde{\mathbf{z}})$ represents the distance between the original measurement and the generated one, and $w_{\mathbf{z}}$ is a hyper-parameter that represents the weight of this distance. This regularization term incentivizes the WGAN to produce a tampered measurement vector $\tilde{\mathbf{z}}$ that will generate completely wrong estimated measurements. Finally, we can explicitly induce sparsity in the attack vector. This sparsity property is desirable and essential because the attacker has to alter fewer measurements to successfully deploy a FDIA, [47]. We can add it into the model in (9) in the paper with the regularization term, $w_{\text{sparse}} \cdot \|\mathbf{z} - \tilde{\mathbf{z}}\|_1$, where w_{sparse} is the weight of the sparsity regularization term. This leads to the following loss function

$$\min_{G} \max_{D\in\mathcal{D}} \mathbb{E}_{\mathbf{z}\sim\mathbb{P}_r}\mathbb{E}_{\tilde{\mathbf{z}}\sim\mathbb{P}_g}\big[D(\mathbf{z}) - D\big(\tilde{\mathbf{z}}\big) + \|\tilde{\mathbf{z}} - \mathrm{AE}\big(\tilde{\mathbf{z}}\big)\|_2^2 + w_{\mathbf{z}} \cdot d\big(\mathbf{z}, \tilde{\mathbf{z}}\big) + w_{\text{sparse}} \cdot \big\|\mathbf{z} - \tilde{\mathbf{z}}\big\|_1\big]. \tag{9}$$

Training the WGAN with regularization terms adds complexity to the training process. If the regularization term becomes too large with respect to the original WGAN loss, the generator will struggle to learn the correct distribution. If the regularization term is too small, it will not have any effect on the training process. Thus, the regularization term will not fulfill its purpose. To solve this issue, a dynamic weight is introduced to control the size of $d(\mathbf{z}, \tilde{\mathbf{z}})$ throughout the training phase. This weight must maintain a balance between the generator loss term $D(\tilde{\mathbf{z}})$ and the regularization term $d(\mathbf{z}, \tilde{\mathbf{z}})$, so that the WGAN learns the desired distribution, and at the same time, the regularization term accomplishes its purpose. We can achieve this balance by setting the regularization term to be half of the generator loss term. We express this as $\frac{1}{2}|D(\tilde{\mathbf{z}})| = w_{\mathbf{z}} \cdot d(\mathbf{z}, \tilde{\mathbf{z}})$. Then, the result of such dynamic weight $w_{\mathbf{z}}$ is described in (10) where $t > 1$ is the iteration number in the training phase. This dynamic weight adapts during training, controlling the impact of the regularization term.

$$w_{\mathbf{z}}^{(t)} = \frac{1}{2} \cdot \left| \frac{D\big(\tilde{\mathbf{z}}^{(t-1)}\big)}{d\big(\mathbf{z}^{(t-1)}, \tilde{\mathbf{z}}^{(t-1)}\big)} \right|. \tag{10}$$

To summarize, our proposed architecture is shown in Fig. 1 with two stages. First, an autoencoder is trained with historical SCADA and PMU measurement data. Second, the WGAN is trained with the same data and the two regularization terms: (1) one incentivizes the WGAN to produce measurements that will pass the residual error test and (2) another to maximize the impact of the attack. More important features are described below, and the complete algorithm for our proposed FDIA is in Algorithm 1.

(i) The inputs for the generative network are actual power system measurements instead of random noise. This gives us control over the created measurements.
(ii) The generator is incentivized to generate measurements that will be different than the ones as input, causing an incorrect estimation of state variables and measurements.
(iii) The generated tampered measurements will have a small residual error, thus passing the residual error test with high probability.

Note that our proposed approach can be easily formulated to deploy an attack on a specific area in the power system, as proposed in [18]. Specifically, a FDIA can be launched in a specific area by tampering the measurements within the area under attack and not modifying the sensor measurements at boundary buses. In this way, the attacker only has to get the sensor measurements in the specific area under attack, which would reduce the amount of collected data. For conciseness and sake of clarity, we will analyze our proposed FDIA in the complete power grid.

Algorithm 1: Training Process of the Proposed Scheme to Create Tampered Measurements to Deploy a FDIA

Inputs : Dataset $\mathcal{M} = \{\mathbf{z}_i \in \mathbb{R}^m\}_{i=1}^{L}$, batch size b, number of iterations of the critic per generator iteration n_{critic}, generator and discriminator learning rates α, clipping parameter c.

Output: Generator (G) network.

Train an AE with the real measurements from the dataset $\mathcal{M}$ and the loss function $\mathcal{L} = \|\mathbf{z} - \mathrm{AE}(\mathbf{z})\|^2$.

for *number of training iterations* **do**

  for $k = 1, \ldots, n_{critic}$ **do**

    Sample a minibatch of b samples $\left\{\mathbf{z}_D^{(1)}, \ldots, \mathbf{z}_D^{(b)}\right\} = \left\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(b)}\right\} \sim \mathbb{P}_r$ from the measurement dataset $\mathcal{M}$.

    Sample a different minibatch of b samples and create a minibatch of fake measurements $\left\{G\left(\mathbf{z}_G^{(1)}\right), \ldots, G\left(\mathbf{z}_G^{(b)}\right)\right\} = \left\{\tilde{\mathbf{z}}^{(1)}, \ldots, \tilde{\mathbf{z}}^{(b)}\right\} \sim \mathbb{P}_g$.

    Train the critic (or discriminator): Gradient ascent on the critic: $\max_{D \in \mathcal{D}} \quad \mathbb{E}_{\mathbf{z} \sim \mathbb{P}_r} \mathbb{E}_{\tilde{\mathbf{z}} \sim \mathbb{P}_g} D(\mathbf{z}) - D\left(\tilde{\mathbf{z}}\right)$.

    Clip discriminator weights in the range $[-c, c]$.

  end

  Sample real and fake measurements: $\left\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(b)}\right\} \sim \mathbb{P}_r$ and $\left\{\tilde{\mathbf{z}}^{(1)}, \ldots, \tilde{\mathbf{m}}^{(b)}\right\} \sim \mathbb{P}_g$.

  Train the Generator: Gradient descent on generator: $\min_{G} \quad \mathbb{E}_{\mathbf{z} \sim \mathbb{P}_r} \mathbb{E}_{\tilde{\mathbf{z}} \sim \mathbb{P}_g}\Big[D(\mathbf{z}) - D\left(\tilde{\mathbf{z}}\right) + w_d \cdot d\left(\mathbf{z}, \tilde{\mathbf{z}}\right) + \|\tilde{\mathbf{z}} - \mathrm{AE}\left(\tilde{\mathbf{z}}\right)\|_2^2\Big]$.

end

Get generator G that creates tampered measurements.

IV. WGAN Guarantee for the Proposed Regularization Terms

The last section presented a framework to create fake power system measurements to deploy a FDIA. However, to successfully deploy a FDIA without relying upon the underlying power system model, we need to be confident that our learned model will produce measurements that look legit so that the residual error test does not detect them. To show that our proposed framework converges to the underlying measurement distribution, we present mathematical proof that certifies the WGAN convergence to the measurement distribution, thus creating fake measurements that look real. The only requirement for this proof to work is to have data to train the WGAN.

Generative adversarial networks can be understood as minimizing a moment matching loss defined by a set of discriminator functions [35], mathematically

$$\min_{\nu \in \mathcal{G}} \left\{ \begin{array}{l} d_{\mathcal{F}}\left(\hat{\mu}_m, \nu\right) := \\ \sup_{f \in \mathcal{F}} \mathbb{E}_{x \sim \hat{\mu}_m} \mathbb{E}_{\tilde{x} \sim \nu} f(x) - f(\tilde{x}) + w_{\mathbf{z}} \cdot d(x, \tilde{x}) \end{array} \right\}, \quad (11)$$

where $\hat{\mu}_m$ is the empirical measure of the observed data (in this case the power system measurements), and $\mathcal{F}$ and $\mathcal{G}$ are the sets of discriminators and generators, respectively. The practical WGANs take $\mathcal{F}$ as a parametric function class, that is, $\mathcal{F}_{nn} = \{f_\theta(x) : \theta \in \Theta\}$ where $f_\theta(x)$ is a neural network indexed by parameters θ that take values in $\Theta \subset \mathbb{R}^p$.

Notation and Definitions: X denotes a subset of $\mathbb{R}^d$. For each continuous function $f : X \to \mathbb{R}$, we define the maximum norm as $\|f\|_\infty = \sup_{x \in X} |f(x)|$, and the Lipschitz norm $\|f\|_{Lip} = \sup\{|f(x) - f(y)|/\|x - y\| : x, y \in X, x \neq y\}$, and the bounded Lipschitz (BL) norm $\|f\|_{BL} = \max\{\|f\|_{Lip}, \|f\|_\infty\}$. The set of continuous functions on X is denoted by $C(X)$, and the Banach space of bounded continuous functions is $C_b(X) = \{f \in C(X) : \|f\|_\infty < \infty\}$.

Weak Convergence: If $\mathcal{F}$ is discriminative, then $d_{\mathcal{F}}(\mu, \nu) = 0$ implies $\mu = \nu$. This means that the learned distribution is the same as the observed one. In reality, we cannot strictly get $d_{\mathcal{F}}(\mu, \nu) = 0$. Rather, we have $d_{\mathcal{F}}(\mu, \nu) \to 0$ for a sequence of ν_n and want to establish the weak convergence $\nu \rightharpoonup \mu$.

Theorem 1: Let (X, d_X) be any metric space. If $span\mathcal{F}$ is dense in $C_b(X)$, we have $\lim_{n \to \infty} d_{\mathcal{F}}(\mu, \nu_n) = 0$ implies that the learned distribution ν_n weakly converges to the real observed distribution μ.

In our context, the observed distribution μ corresponds to the set of observed power system measurements. Fig. 2 gives the intuition for the convergence proof. The learned distribution ν_n (in red) converges to the real one μ (in blue) as $n \to \infty$. In other words, the WGAN is learning to create samples that look as taken from the true observed distribution μ.

Proof: Given a function $g \in C_b(X)$, we say that g is approximated by $\mathcal{F}$ with error decay function $\epsilon(r)$ if for any $r \geq 0$, there exists $f_r \in span\mathcal{F}$ with $\|f_r\|_{\mathcal{F},1} \leq r$ such that $\|f - f_r\|_\infty \leq \epsilon(r)$. We note that $\epsilon(r)$ is a non-increasing function with respect to r. We know that the closure of $span\mathcal{F}$ is equal to the space of bounded continuous functions $C_b(X)$, that is, $cl(span\mathcal{F}) = C_b(X)$, then we have $\lim_{r \to \infty} \epsilon(r) = 0$. Now denote $r_n := d_F(\mu, \nu_n)^{-\frac{1}{2}}$, $f_n := f_{r_n}$ and $w_{\mathbf{z}} = 1/r_n$. We have $|\mathbb{E}_\mu g - \mathbb{E}_{\nu_n} g| + w_{\mathbf{z}} \cdot d(x, \tilde{x}) \leq |\mathbb{E}_\mu g - \mathbb{E}_\mu f_n| + |\mathbb{E}_\nu g - \mathbb{E}_\nu f_n| + |\mathbb{E}_\mu f_n - E_{\nu_n} f_n| + w_{\mathbf{z}} \cdot d(x, \tilde{x}) \leq 2\epsilon(r_n) + r_n d_{\mathcal{F}}(\mu, \nu_n) + w_{\mathbf{z}} \cdot d(x, \tilde{x}) = 2\epsilon(r_n) + 1/r_n + w_{\mathbf{z}} \cdot d(x, \tilde{x})$. If $\lim_{r \to \infty} d_F(\mu, \nu_n) = 0$, we have $\lim_{r \to \infty} r_n = \infty$. Given that $\lim_{r \to \infty} \epsilon(r) = 0$, we prove that $\lim_{n \to \infty} |\mathbb{E}_\mu g - \mathbb{E}_{\nu_n} g| + w_{\mathbf{z}} \cdot d(x, \tilde{x}) = 0$. Since this holds true for any $g \in C_b(X)$, we conclude that ν_n weakly converges to μ. If $\mathcal{F} \subseteq BL_C(X)$ for some $C > 0$, we have $d_{\mathcal{F}}(\mu, \nu) \leq C d_{BL}(\mu, \nu)$ for any μ, ν. Because the bounded Lipschitz distance metrizes the weak convergence, we obtain that $\nu_n \to \mu$ implies $d_{BL}(\mu, \nu_n) \to 0$, and $d_{\mathcal{F}}(\mu, \nu_n) \rightharpoonup 0$. ■

Theorem 1 guarantees us that the learned distribution ν by the WGAN will converge to the observed one μ. This idea is depicted in Fig. 2. The blue points represent the real measurements, and the red ones represent the fake measurements. At the beginning, the red points are random because the WGAN is not trained ($n = 1$). However, as training progresses, the WGAN produces samples (red points) that look more similar to the blue ones. Ideally, the fake samples will be indistinguishable from the real ones. In other words, our model will create fake measurements that look like real ones. This means that the WGAN captures the underlying power system's interactions that produce the observed measurements.

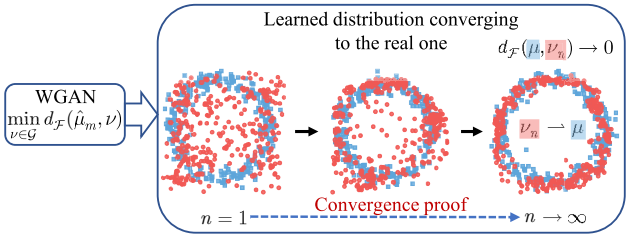


Fig. 2. Intuition for the WGAN convergence proof to the true observed distribution.

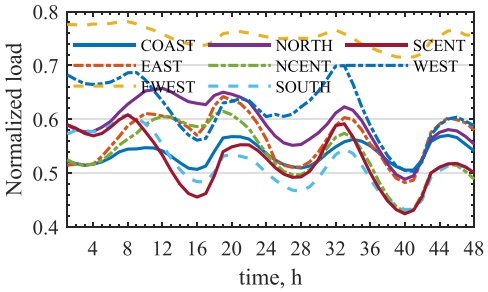


Fig. 3. ERCOT hourly normalized load data for 2021.

V. Experiments

This section will show how we deploy FDIAs on power grids with our proposed WGAN framework without knowing their mathematical or physical model. To show the contributions and generality of our approach, we carried out extensive experiments on different power networks.

First, we train a WGAN with historical SCADA and PMU measurements to demonstrate that the output of the WGAN converges to the true distribution of observed power system measurements, $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$. Note that the sampling rate of PMU measurements is faster than the sampling rate of SCADA measurements. We use PMU measurements alongside with SCADA measurements when the SCADA measurements are available. We will also show that the fake measurements will pass the residual error test, corroborating the aforementioned convergence theorem. Second, we show that the trained WGAN creates different measurements (and therefore states) from the actual ones. This will show that the regularization term works, and it is maximizing the FDIA impact. Next, we show that our proposed framework is more reliable than the model-based ones by showing that our WGAN produces more realistic measurements. This implies that our model is capturing the underlying power system model. Finally, an ablation study is carried out to show that embedding a surrogate state estimator model, $\mathbf{h}(\mathbf{x})$, improves the proposed framework to create tampered measurements that pass the residual error test. We carried out the aforementioned experiments in various test cases with similar results. Specifically, we use the small IEEE 9-bus test case to illustrate how our framework works. Then, we perform the same simulations in the IEEE 14-, 57-, 118-, and 300-bus networks to demonstrate that our proposed method scales well with larger power system networks.

A. Data Generation and Model Architecture

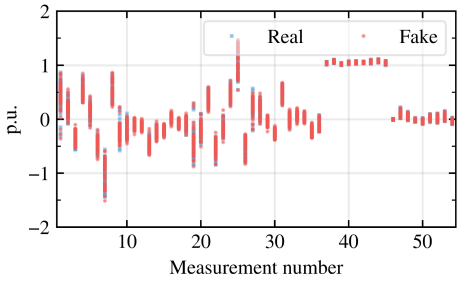
1) Data Generation: For both the 9- and 118-bus test cases, we consider all the active and reactive power flow measurements through transmission lines and transformers as SCADA measurements, and voltage magnitudes and angles as PMU measurements. The 9-bus network has 9 branches, which gives us 36 SCADA measurements and 18 PMU measurements. The measurements are arranged as follows: 1–9 correspond to the sent active power through branches, 10–18 correspond to the sent reactive power, 19–27 are the received active power measurements, 28–36 are the received reactive power on branches, 37–45 are the voltage magnitudes, and 46–54 are the voltage angles. The IEEE 118-bus network has 186 branches; thus, 980 measurements arranged as follows: 1–186 sent active power, 187–372 sent reactive power, 373–558 received active power, 559–744 received reactive power, 745–862 are the voltage magnitudes, and 863–980 are the voltage angles.

We obtain the power systems' measurements by solving L times the AC power flow under different load conditions using MATPOWER [48]. To simulate the 24-hour fluctuation, we use the real yearly load data from the Electric Reliability Council of Texas (ERCOT) for 2021 [49]. ERCOT reports 8 weather zones: *COAST*, *EAST*, *FWEST*, *NORTH*, *NCENT*, *SOUTH*, *SCENT*, and *WEST*. Fig. 3 depicts the load profiles of these zones for 2 days in 2021. For our simulations, we multiply each busload with the normalized loading parameter associated with a randomly selected area, γ, obtained from these realistic profiles. Similarly, we also adjust generation by scaling the generation profiles by multiplying them by the same loading parameter, γ, [50], [51]. To make it more realistic, we add white noise to all measurements according to the standard deviation associated with the measurement devices. That is, active power flow: 0.02 p.u., reactive power flow: 0.04 p.u., active power injection: 0.02 p.u., reactive power injection: 0.04 p.u., PMU voltage magnitude: 0.0001 p.u., and PMU voltage angle: 0.006 rad, according with [52]. Finally, if we do not find an AC power flow solution, we do not include it in the dataset. This data generation approach will give us rich data variety with the power system under different load conditions. The same procedure is used to generate data for the IEEE 14-, 57-, and 300-bus test cases.

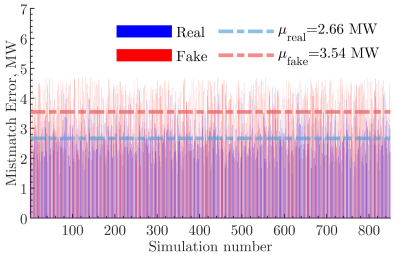
2) Model Architecture: The architecture of our proposed WGAN model is inspired by the architecture of the DCGAN [53] with the following modifications to adapt it to our power system data. Since the sensor measurement vectors are one-dimensional, we use fully connected layers instead of convolutional layers. The generator, G, consists of 5 layers with ReLU activation function for all layers except for the output, which uses tanh. The discriminator, D, is composed of 5 layers with LeakyReLU activations with the slope of the leak set to 0.2.

B. Learning the Implicit Power System Measurement Model

This section tests if the learned distribution by the WGAN converges to the true underlying power system measurement distribution, $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$. We train the WGAN according to Algorithm 1 with a dissimilarity weight $w_{\mathbf{z}} = 0.5$. We use the hyper-parameters from [45]: $n_{critic} = 5$, learning rates



(a) Real vs. fake tampered measurements for the 9-bus test case. Note that the fake measurements look like the real ones.



(b) Power flow mismatch error for the real and fake measurements.

Fig. 4. Learning an implicit power system model with the proposed WGAN architecture for the 9-bus test case using real load profiles from ERCOT [49].

$\alpha = 0.00005$ (for autoencoder, generator, and discriminator), clipping parameter $c = 0.01$, batch size $b = 64$, and Adam adaptive learning algorithm [54]. Also, we train the AE and the WGAN models for all test cases for 10 and 100 epochs, respectively. The normalized load from the Electric Reliability Council of Texas (ERCOT) for 2021 [49] contains hourly data for one year, which means that there are 8,760 load samples. From these 8,760 samples, we split the set into a training and a test dataset with 7,760 and 1,000 randomly chosen samples, respectively. This yearly data contains seasonal variation, so it captures the behavior of a real power system throughout the year. Note that both the AE and WGAN models are trained with this data, as indicated in Algorithm 1. Fig. 4(a) shows 100 measurement samples from the real dataset and 100 created fake measurements for the 9-bus test case. We can see in Fig. 4(a) generated fake measurements compared with real measurements from our dataset; the fake measurements (in red) follow the same pattern or distribution as the real ones (in blue); in fact, they overlap the real measurements, but they are not exactly the same. This means that the WGAN learned the true power system measurement distribution instead of memorizing the dataset. Note that Theorem 1 guarantees the model convergence with enough training data. In our numerical experiments, we trained our models by creating training and testing datasets of 7,760 and 1,000 samples, respectively. With these training datasets our models successfully learned the underlying power system measurement distribution. Also note that the our procedure to create the dataset produces rich distributions of sensor measurements, Fig. 4(a). For example, the measurement no. 1 has a range from 0p.u. to over 1p.u. (which corresponds to an active branch power flow measurement).

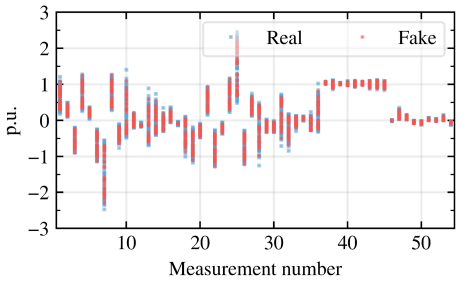
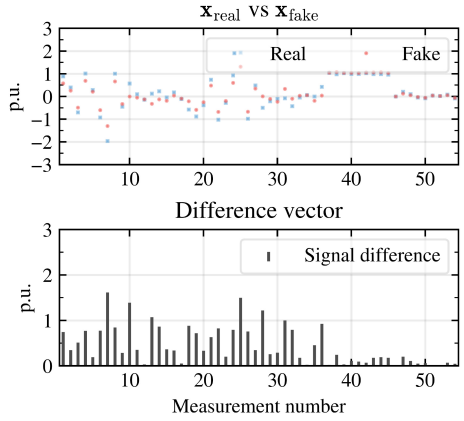


Fig. 5. Sensor measurement distribution with VRES for the 9-bus test system.

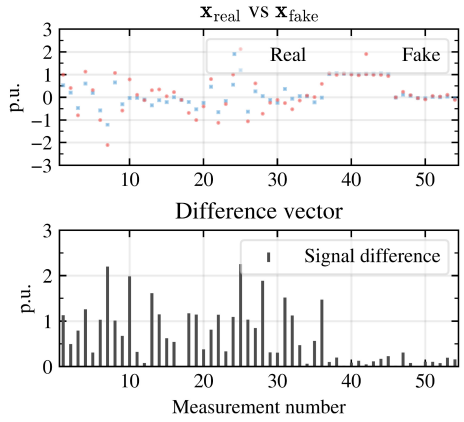
To assess if the trained WGAN learned the implicit power system measurement distribution, we carry out a power flow mismatch analysis, as follows. If we add power injection measurements in the set of measurements, the power flow balance at the i-th bus should be $\sum_{j\in\delta^+(i)} f^p_{(i,j)} + e_j = p_i^{\text{inj}} + e_i$, where $\delta^+(i)$ is the set of adjacent buses to bus i, $f^p_{(i,j)}$ is the power flow on branch (i,j), and e_j and e_i are the measurement errors associated to active power flow and injection, respectively. Under this setting, the power flow mismatch will not be zero due to measurement errors, that is, $|\sum_{j\in\delta^+(i)} f^p_{(i,j)} - p_i^{\text{inj}}| > 0$. We compute this power mismatch error $|\sum_{j\in\delta^+(i)} f^p_{(i,j)} - p_i^{\text{inj}}|$ for all the buses in the system for both real and fake tampered measurements. Fig. 4(b) shows the results, where each bar, blue for real and red for fake measurements, indicates the average power flow mismatch in the whole system for one simulation. In the same figure, we can see that the power flow mismatches of the real and tampered fake measurements are very close: 2.66 MW for the real measurements and 3.54 MW for the tampered fake measurements. This is remarkable because the WGAN does not know the power system topology, and it does not have information about which measurements should comply with the power flow balance. Yet, the WGAN produces fake tampered measurements that are within 1 MW, on average, with respect to the real measurements, as shown in Fig. 4(b).

Including variable renewable sources such as wind and solar generation that vary significally from one day to the next could produce a more diverse sensor measurement distribution. To test this idea, we use the 9-bus test case, and we take the normalized wind and solar aggregated generation data from the RTS-GMLC [55]. Then, we include the wind generation on bus 5 and the solar generation on bus 6 with different penetration values. For a penetration of 30%, we can see the sensor measurement distribution in Fig. 5. This distribution looks a little bit wider than the one without VRES in Fig. 4(a). Notice that both sensor measurement distributions look alike, which means that our original procedure to generate datasets creates rich sensor measurement distributions. Thus, the datasets for the remaining experiments will be created without adding VRES into the simulations.

1) Analyzing Attack's Vector Sparsity: We can test the attack vector's sparsity by taking the absolute difference between the real and tampered measurement vectors, that is, $|\mathbf{z} - \tilde{\mathbf{z}}|$. To test this idea, we take the real and tampered

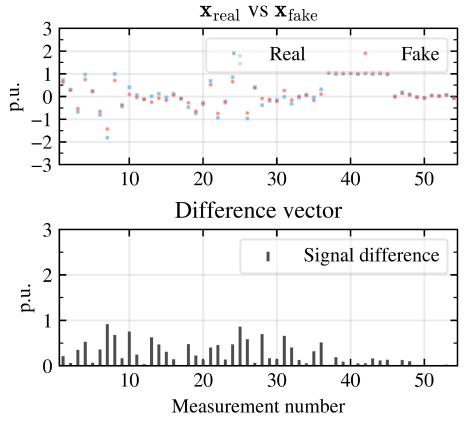


(a)

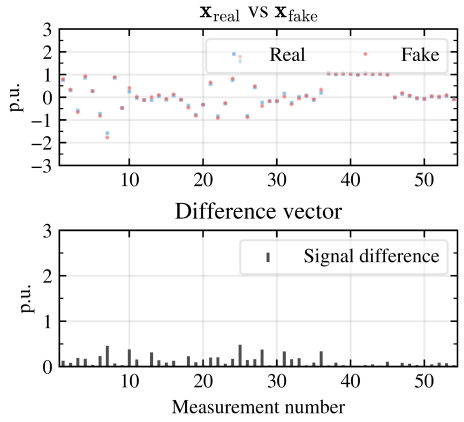


(b)

Fig. 6. Examples of absolute difference vectors.



(a)



(b)

Fig. 7. Examples of absolute difference vectors with the sparsity regularizer with $w_{\text{sparse}} = 0.5$.

measurements for the 9-bus test case, with $w_{\text{sparse}} = 0$, and we show two examples of specific sets of real and tampered measurements in Fig. 6(b). In the top part of the Figure, we can see the real and tampered measurements. In the inferior part of the Figure, we can see the absolute difference vectors, $|\mathbf{z} - \tilde{\mathbf{z}}|$. Note that even though $w_{\text{sparse}} = 0$, these vectors contain many zero values indicating the property of sparsity.

We train the WGAN following the same procedure in the paper for the 9-bus test system with the addition of the sparsity regularizer with a weight of 0.5, that is, $w_{\text{sparse}} = 0.5$. To test the sparsity of the results, we follow the same experiment design from the last example. Specifically, we take the real and tampered measurements for the 9-bus test case, and we show two examples of specific sets of real and tampered measurements in Fig. 7. In the top part of the Figure, we can see the real and tampered measurements. In the inferior part of the Figure, we can see the absolute difference vectors, $|\mathbf{z} - \tilde{\mathbf{z}}|$. As expected, when sparsity is explicitly taken into account, the attack vectors (absolute difference vectors in Fig. 7) present more sparsity than those in Fig. 6(b), where no sparsity is expressly considered in the model. However, the differences between real and tampered measurement vectors for the sparse FDIA are smaller than the FDIA that does not explicitly take into account the sparsity.

The model's results without including sparsity, $w_{\text{sparse}} = 0$, present sparsity and produce more changes in the tampered measurements. Thus, the remaining experiments will be done without explicitly including sparsity.

2) Analyzing Attack Vector: We can assess an attack vector's impact by taking the absolute difference between the real and tampered measurement vectors, that is, $|\mathbf{z} - \tilde{\mathbf{z}}|$. To test this idea, we take $1,000$ real and tampered measurements for the 9-bus test case, and we show two examples of specific sets of real and tampered measurements in Fig. 6(b). In the top part of the Figure, we can see the real and tampered measurements. In the inferior part of the Figure, we can see the absolute difference vectors, $|\mathbf{z} - \tilde{\mathbf{z}}|$. Note that in the $1,000$ samples, the mean magnitude of the attack vector is 15.05 units. Also, the attack vector, in specific sensor measurements, dramatically changes the real values. Under this context, the operator could take wrong corrective actions that will interfere with the correct and safe operation of the electric grid. This means that the attack will damage the system and lead to catastrophic events.

C. Deploying FDIAs Without Power System Knowledge

In the last section, we showed that a WGAN can learn the power system measurement distribution. This section shows

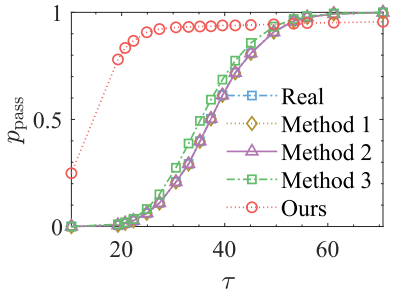


Fig. 8. Comparison of passing the residual error test with different methods for the 9-bus test case.

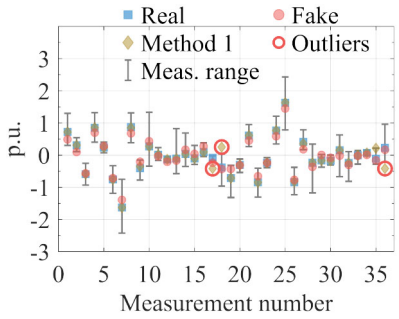


Fig. 9. Comparison of the tampered measurements by the model-based Method 1 [9] with our model-free approach for the 9-bus test case.

how we deploy a FDIA with our proposed framework, which is given by (9) and (10).

1) Deploying a FDIA With Fake Tampered Measurements: Our objective is to create fake tampered measurements $\tilde{\mathbf{z}}$ that generate estimated measurements and state variables as different as possible from the real ones. At the same time, for an attack to be successful, these measurements should pass the residual error test. Fig. 9 shows an instance of a real measurement vector and a created fake one for the 9-bus test network. The fake tampered measurements are within the historical range from the dataset and look similar to the real ones. However, they produce significant changes in voltage magnitudes v and voltage angles δ with respect to the real states, as shown in Fig. 10. Furthermore, the fake measurements pass the test in (2), which means that the control center will not notice the FDIA.

2) Comparison Against Other FDIA Methods: To assess the advantages and differences between our proposed model-free FDIA framework, we compare it against the model-based FDIA presented in [9] and described by (3)—we will refer to this FDIA as Method 1. This model-based attack has the same residual error as the original measurements as proven in [9]. However, the Method 1 produces measurements that are out of the historical range from the historical measurements.

To prove the last point, we perform the following experiment. We use the fake vector in Fig. 10, where we can see that the voltage magnitude in bus 5 goes from 1 to 1.05 p.u. We use Method 1 to tamper the state $v_5 = 1.05$ p.u. using (3). Fig. 9 shows the real measurements (in blue), the created tampered measurements by our proposed framework (in red), the

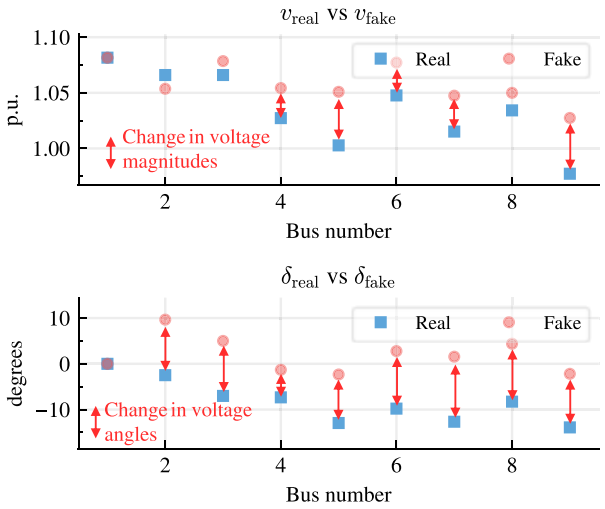


Fig. 10. Example of a real and a fake measurement vector for the 9-bus test case.

created tampered measurements by Method 1, and the historical measurement range from our data generation (gray bar). In the same Figure, we see that the created measurements by the WGAN are within or very close to the historical range. In contrast, some tampered measurements by Method 1 are far away from the real historical measurements. In specific, we see in Fig. 9 that measurements 18 and 36 show a large distance from the historical range. The key observation is: Even though Method 1 produces measurements with the same residual error as the real ones, these measurements will still look suspicious. The power system operator would realize that the tampered measurements 18 and 36 are outliers with respect to the historical ones, as shown in Fig. 9. In contrast, in the same Figure, we can see that our fake tampered measurements are within the range of historical measurements and also pass the residual error test (for a confidence of $p = 0.95$). Thus, making them less suspicious for the power system operator. This means that our attack design is more advantageous at the stealth level.

We also carried out a sensitivity analysis for different confidence values p. In this sensitivity analysis, we compare our method against three techniques in the literature: Method 1 introduced in [9], Method 2 from [25], and Method 3 proposed in [18]. This sensitivity analysis is carried out with the residual error test. Thus, the results only depend on the residual error produced by the FDIA approaches. In other words, the range of historical measurements does not affect the success rate. Methods 1 and 2 produce the same residual error as the real measurements; this means that if the real measurement passes the residual error test, the tampered measurements by these methods will pass as well. Method 3 is an attack on a specific area, and we chose to delimit this area by the buses 5 and 6. An important characteristic of this technique is that the residual error of the tampered measurements can be lower than the real residual. The authors in [18] attribute it to the fact that the tampered measurements will be more consistent (i.e., free of noise errors); thus, reducing the overall residual error.

TABLE I
COMPARISON OF DIFFERENT FDIAS

	Ours	M1 [9]	M2 [25]	M3 [18]
Is power system model needed?	✗	✓	✗	✓
Same residual as originals?	✗	✓	✗	✓
Measurements needed to deploy attack	All or Area under attack	All	All	Area under attack
Tampered measurements within historical range?	✓	✗	✗	✗

To compare these methods, we made 1,000 simulations with the same procedure described in Section V-A, and we tamper the real noisy measurements with our proposed approach and Methods 1, 2, and 3. For a given confidence value p, we compute its corresponding threshold $\tau = \chi^2_{k,p}$, and obtain the probability of each measurement to pass the residual error test for the specified threshold, that is, $\Pr(J(\mathbf{z}) \geq \tau)$. We repeat this process for each simulation and each aforementioned method, and we obtain the success rate of passing the residual error test. This is the probability of the simulations to pass the error test, and we call it p_{pass}. We repeat this experiment for several values $p \in (0, 1)$, and the result is shown in Fig. 8. We can see that as the threshold τ increases, the probability to pass the residual error test p_{pass} increases as well. Given that Methods 1 and 2 (in brown and purple, respectively) tampered the measurements such that the residual error is the same as the real one (in blue), they (almost) follow perfectly the real curve. Method 3 (in green) is close to the real curve but just slightly above due to the behavior of this technique, as we previously explained. Note that Methods 1 and 2 produce the same p_{pass} as the real noisy measurements in Fig. 8. This is because both methods are guaranteed to have the same residual error as the real noisy measurements by design, as indicated in (3) (see proof in [9]).

It is important to note that we trained our model with noisy measurements, and the method did not have access to the underlying power system model. The key finding is that despite using only noisy measurements, our approach produces tampered measurements with lower residual errors, outperforming all other methods. We ascribe this due to the regularization term that contains the AE in (9), $\|\tilde{\mathbf{z}} - \text{AE}(\tilde{\mathbf{z}})\|_2^2$. As discussed in Section III-B, an autoencoder has a denoising effect on the on the noisy measurements. This will be proved with an ablation study in Section V-E. A summary of the qualitative traits of each of the aforementioned methods is shown in Table I, where it is shown that our proposed algorithm is the only one that tampers measurements so that they are within the historical range.

3) Comparison Against Other Model-Free FDIA Method: To make a fair comparison, we train our proposed model with the same methodology indicated in the paper with the difference that we use the DC power flow model as the work in [29] does. This framework requires normal and tampered measurements to train a conditional adversarial network

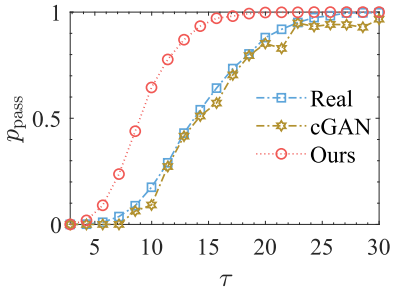


Fig. 11. Comparison of passing the residual error test with the cGAN, [29], for the 14-bus test case.

TABLE II
COMPARISON OF PASSING THE RESIDUAL ERROR TEST WITH THE CGAN, [29]

Test Case	Success Rate (%)	
	Ours	cGAN
9-bus	95.5	92.7
14-bus	95.7	95.78
57-bus	89.3	93.6
118-bus	97	91.4
300-bus	91	93.1

(cGAN). However, the work in [29] does not clearly indicate how the dataset of tampered measurements is obtained. For simplicity, we use the well-known FDIA proposed in [9] to create the dataset of tampered measurements. We evaluate both approaches on the 14-bus test. For a given confidence value p, we compute its corresponding threshold $\tau = \chi^2_{k,p}$, and obtain the probability of each measurement to pass the residual error test for the specified threshold, that is, $\Pr(J(\mathbf{z}) \geq \tau)$. We repeat this process for each simulation and each aforementioned method, and we obtain the success rate of passing the residual error test. This is the probability of the simulations to pass the error test, and we call it p_{pass}. We repeat this experiment for several values $p \in (0, 1)$, and the result is shown in Fig. 11. We carry out the same experiments for the IEEE 9-, 57-, 118-, and 300-bus test cases for a confidence value $p = 0.95$. The results are shown in Table II.

4) Validate Scalability of the Proposed Approach: Finally, we show that our approach scales to bigger power system networks. To demonstrate it, we test our model-free FDIA on the IEEE 118-bus network. The created fake tampered measurements pass the residual error test, and Fig. 12 shows that the created fake measurements provoke significant changes in the voltage angles, leading to a successful FDIA.

Also, a sensitivity analysis, like the one in the previous section, is carried out for the IEEE 9-, 14, 57-, 118-, and 300-bus test cases, and the results are shown in Fig. 13. In the same Figure, we can see that our FDIA method outperforms the ones proposed in the literature.

Finally, we validate the scalability of our proposed approach. As previously mentioned, the AE and the WGAN models for all the test cases are trained for 10 and 100, respectively. The number of training samples and the number of iterations for all test cases are fixed since we used real yearly load data from the Electric Reliability Council of Texas (ERCOT) for 2021 [49]. Also, the number of layers is fixed

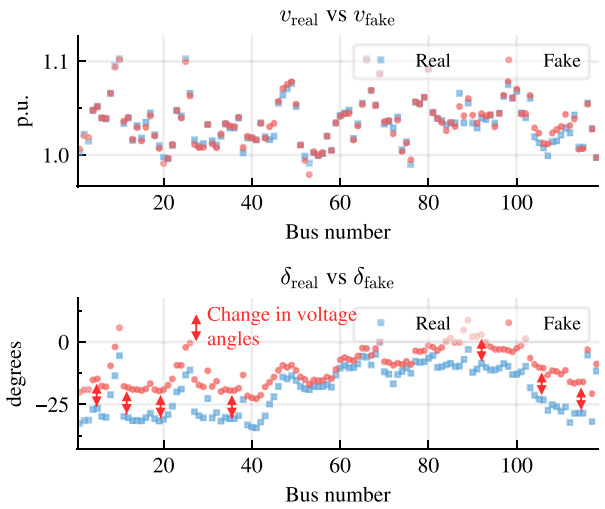


Fig. 12. Example of a real vs a fake measurement for the 118-bus test case. Note that the fake measurements produce different states.

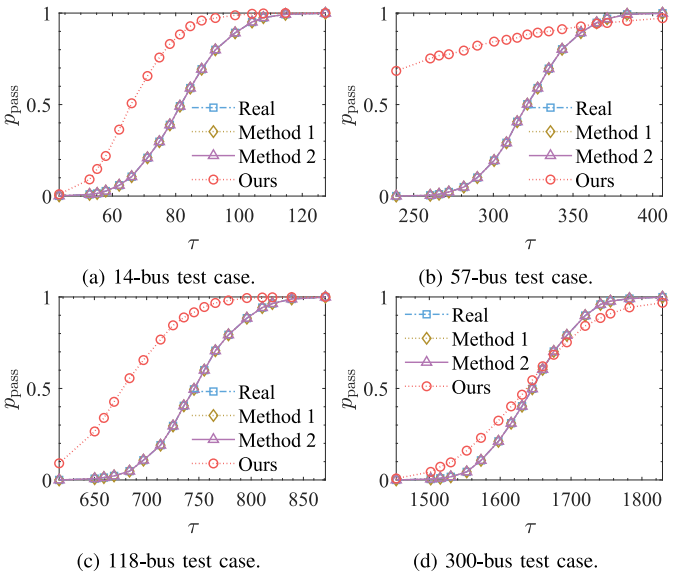


Fig. 13. CDF comparison for many test cases.

to be 5 for both the generator and discriminator for all the experiments. The only component that varies is the dimensionality, which depends upon the power system size. Thus, our proposed approach presents good scalability with respect to the power system size. We can test this by measuring the training times for the AE and WGAN models. Fig. 14 shows such training times. We can see that training the surrogate state estimator (i.e., AE) for 10 epochs takes less than 40 sec for all test cases. Training the WGAN model for 100 epochs takes less than 530 sec for all test cases. We can see that the training times for the models' convergence for 1 year of data are low. Thus, our proposed attack could be easily deployed in real-world settings.

D. Comparison of Different Defenses

The Chi-squared test could be, in some cases, inaccurate due to the approximations of errors by residuals [39]. So, in this

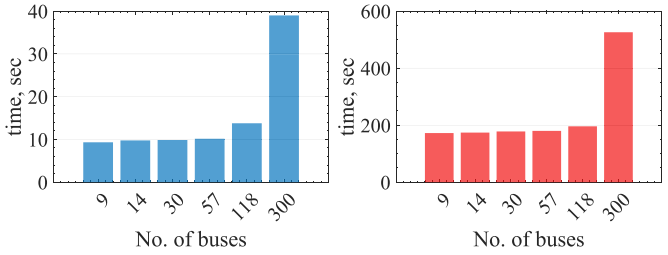


Fig. 14. Training times for AE and WGAN for different test cases.

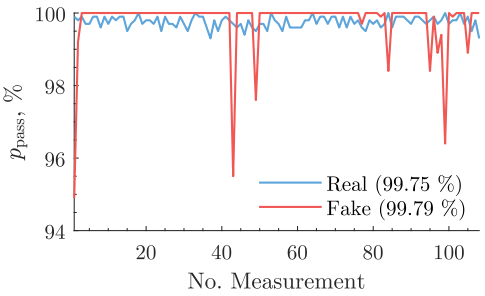


Fig. 15. Largest normalized residual statistical test for the 14-bus test system.

section, we show how our proposed algorithm performs against more sounding defenses. In the literature, there exist numerous defenses with different traits. For example, defenses that do not use temporal correlations and ones that make use of them. In the realm of defenses that exploit temporal patterns to detect FDIAs, there are works such as the moving-target defense (MTD) [56], [57] or the work in [41]. However, our proposed FDIA scheme does not take into account inter-temporal correlation, so it would be unfair to test our attack against such defenses. Thus, in this section, we choose defenses that utilize data measurements at a specific time interval to detect spurious data. Specifically, we test our proposed attack against the largest normalized residual statistical test (LNRT) [39], [58] and a recent deep learning-based detector that consists of an adversarial autoencoder [59].

1) Largest Normalized Residual Statistical Test (LNRT): The LNRT is more robust than the classical Chi-squared test for bad data detection and identification [39], [58]. The normalized value of the residual for the measurement i can be computed as $r_i^{\text{norm}} = \frac{|r_i|}{\sqrt{\Omega_{ii}}}$, where $\sqrt{\Omega_{ii}}$ is the diagonal entry in the residual covariance matrix. This normalized residual entry has a standard normal distribution, that is, $r_i^{\text{norm}} \sim \mathcal{N}(0, 1)$. Then, the largest element in the set $\{r_i^{\text{norm}}\}_{i=1}^{M}$ is compared against a chosen threshold to decide if bad data is presented. If this threshold is set to 3, then the confidence level is 99.7%. We carry out this test for the 14-bus test system for each real and fake measurement, and the results are shown in Fig. 15, where the average is 99.75% for real measurements and 99.79% for tampered measurements with our proposed method. We carry out the same experiments for the IEEE 9-, 57-, 118-, and 300-bus test cases for a confidence value $p = 0.997$. The results are shown in Table III.

2) Deep Learning-Based Detector: There are recent learning-based detectors to detect FDIAs. The work in [59],

TABLE III
COMPARISON OF DIFFERENT DEFENSE MECHANISMS AGAINST A FDIA. $^{*}p = 0.997$, $^{\dagger}p = 0.95$

Test Case	Success Rate (%)		
	LNRT*	Chi-squared†	AEE [59]
9-bus	98.3	95.5	93
14-bus	99.79	95.7	92.6
57-bus	92	89.3	86.5
118-bus	99.4	97	92.3
300-bus	93.5	91	84.4

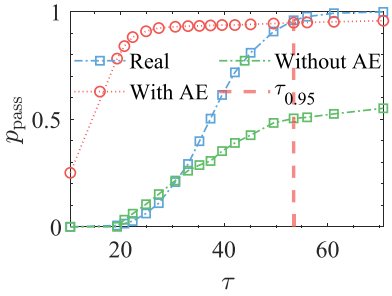


Fig. 16. Probability of passing the residual error test for 9-bus test case with and without AE.

for example, proposed a scheme that consists in an adversarial autoencoder (AEE). The AAE network is trained in three stages: the reconstruction phase, the adversarial phase, and the supervised phase. For a model-based FDIA, this AEE has a detection accuracy of 96.25% and 97.85% for the 13- and 123-bus distribution networks. We test this defense against our proposed model-free FDIA for the IEEE 9-, 14-, 57-, 118-, and 300-bus test cases, and the results are shown in Table III. In this table, we can see that our proposed approach has a lower success rate for the AAE defense than for the Chi-squared and LNRT. Nonetheless, our method still exhibits a high success rate (above 80%) for all the tested cases.

E. Ablation Study

This section presents an ablation study to show the impact of the SE's surrogate model in the proposed framework. The experiment design is similar to the one presented in previous sections. We made 1,000 simulations with the same procedure described in Section V-A. For a given confidence value p, we compute its corresponding threshold $\tau = \chi^2_{k,p}$, and obtain the probability of each measurement to pass the residual error test for the specified threshold, that is, $\Pr(J(\mathbf{z}) \geq \tau)$. We repeat this process for each simulation for the real and proposed framework with and without AE for the 9-bus test case. Next, we obtain the success rate of passing the residual error test, p_{pass}. We repeat this experiment for several values $p \in (0, 1)$, and the result is shown in Fig. 16. In the same Figure, we can see that the model without the AE has a lower probability of passing the residual error test throughout all the thresholds. We can also see that the model without the AE (green line) always have around the same or lower probability of passing the residual error test than the real measurements. As discussed in Sections III-B and V-C2, whereas the model with the AE has a denoising effect the model without the AE can only learn from the noisy measurement data.

TABLE IV
IMPACT OF INCLUDING AN AE

Test Case	Success Rate (%)	
	With AE	Without AE
9-bus	95.5	63.7
14-bus	95.7	81.1
57-bus	89.3	61
118-bus	97	54.33
300-bus	91	70.6

We carry out the same experiments for the IEEE 14-, 57-, 118-, and 300-bus test cases for a confidence value $p = 0.95$. The results are shown in Table IV, which shows that the model with the AE has a higher success rate than the one without it.

VI. CONCLUSION

We presented an architecture to create tampered measurement vectors to carry out a FDIA without knowing the power system underlying information. The architecture is framed into an optimization framework that considers the WGAN loss function and two regularization terms to control the attack measurement vectors. We validated our proposed framework with several power systems, in which we created fake measurements to create a bad data injection attack without knowing the underlying power system model. These fake measurements passed the residual error test to detect bad data and gave completely wrong estimated state variables and measurements, which would compromise the electric grid's reliability. This work proves that for an attacker, it is not required to have access to all power system information. Thus, more research is needed to keep power systems safe from these attacks.

REFERENCES

[1] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2010, pp. 226–231.
[2] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 659–666, Dec. 2011.
[3] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, and H. Zhu, "Securing smart grid: Cyber attacks, countermeasures, and challenges," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 38–45, Aug. 2012.
[4] Y. Zhou, Y. Liu, and S. Hu, "Smart home cyberattack detection framework for sponsor incentive attacks," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1916–1927, Mar. 2019.
[5] N. Costilla-Enriquez and Y. Weng, "Exposing cyber-physical system weaknesses by implicitly learning their underlying models," in *Proc. Asian Conf. Mach. Learn.*, 2021, pp. 1333–1348.
[6] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Security*, vol. 14, no. 1, pp. 1–33, 2011.
[7] M. Mohammadpourfard, A. Sami, and Y. Weng, "Identification of false data injection attacks with considering the impact of wind generation and topology reconfigurations," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1349–1364, Jul. 2018.
[8] M. Mohammadpourfard, Y. Weng, and M. Tajdinian, "Benchmark of machine learning algorithms on capturing future distribution network anomalies," *IET Gener. Transm. Distrib.*, vol. 13, no. 8, pp. 1441–1455, 2019.
[9] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sep. 2012.

[10] Z. Wang, H. He, Z. Wan, and Y. Sun, "Detection of false data injection attacks in AC state estimation using phasor measurements," *IEEE Trans. Smart Grid*, early access, Feb. 10, 2020, doi: 10.1109/TSG.2020.2972781.

[11] N. Costilla-Enriquez, Y. Weng, and B. Zhang, "Combining Newton-Raphson and stochastic gradient descent for power flow analysis," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 514–517, Jan. 2021.

[12] M. Jin, J. Lavaei, and K. H. Johansson, "Power grid AC-based state estimation: Vulnerability analysis against cyber attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 5, pp. 1784–1799, May 2019.

[13] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 382–390, Jun. 2011.

[14] J. Valenzuela, J. Wang, and N. Bissinger, "Real-time intrusion detection in power system operations," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[15] Y. Mo *et al.*, "Cyber–physical security of a smart grid infrastructure," *Proc. IEEE*, vol. 100, no. 1, pp. 195–209, Jan. 2012.

[16] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Proc. Conf. Inf. Sci. Syst.*, 2010, pp. 1–6.

[17] Y. Zhang, L. Wang, Y. Xiang, and C.-W. Ten, "Power system reliability evaluation with SCADA cybersecurity considerations," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1707–1721, Jul. 2015.

[18] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2239–2248, Sep. 2017.

[19] L. Jia, R. J. Thomas, and L. Tong, "On the nonlinearity effects on malicious data attack on power system," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–8.

[20] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 24–45, Feb. 2015.

[21] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Comput. Netw.*, vol. 57, no. 5, pp. 1344–1371, 2013.

[22] J. Styczynski and N. Beach-Westmoreland. "When the Lights Went Out: A Comprehensive Review of the 2015 Attacks on Ukrainian Critical Infrastructure." 2019. [Online]. Available: https://www.boozallen.com/s/insight/thought-leadership/lessons-from-ukranians-energy-grid-cyber-attack.html

[23] National Academies of Sciences, Engineering, and Medicine, *Enhancing the Resilience of the Nation's Electricity System*. Washington, DC, USA: Nat. Acad. Press, 2017.

[24] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[25] W.-L. Chin, C.-H. Lee, and T. Jiang, "Blind false data attacks against AC state estimation based on geometric approach in smart grid communications," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6298–6306, Nov. 2018.

[26] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1102–1114, Mar. 2015.

[27] Z. Zhang, R. Deng, D. K. Yau, and P. Chen, "Zero-parameter-information data integrity attacks and countermeasures in IoT-based smart grid," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6608–6623, Apr. 2021.

[28] S. Ahmadian, H. Malki, and Z. Han, "Cyber attacks on smart energy grids using generative adverserial networks," in *Proc. Global Conf. Signal Inf. Process.*, 2018, pp. 942–946.

[29] M. Mohammadpourfard, F. Ghanaatpishe, M. Mohammadi, S. Lakshminarayana, and M. Pechenizkiy, "Generation of false data injection attacks using conditional generative adversarial networks," in *Proc. Innovat. Smart Grid Technol. Europe*, 2020, pp. 41–45.

[30] L. Wasserman, *All of Nonparametric Statistics*. New York, NY, USA: Springer, 2006.

[31] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*. New York, NY, USA: Springer, 2006.

[32] M. Hollander and J. Sethuraman, "Nonparametric statistics: Advanced computational methods," in *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam, The Netherlands: Elsevier, 2001.

[33] G. Ji, M. C. Hughes, and E. B. Sudderth, "From patches to images: A nonparametric generative model," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1675–1683.

[34] S. Lakshminarayana, S. Sthapit, and C. Maple, "Application of physics-informed machine learning techniques for power grid parameter estimation," *Sustainability*, vol. 14, no. 4, p. 2051, 2022.

[35] P. Zhang, Q. Liu, D. Zhou, T. Xu, and X. He, "On the discrimination-generalization tradeoff in GANs," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[36] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control*. Hoboken, NJ, USA: Wiley, 2013.

[37] A. Tarali and A. Abur, "Bad data detection in two-stage state estimation using phasor measurements," in *Proc. IEEE PES Innovat. Smart Grid Technol. Europe*, 2012, pp. 1–8.

[38] Y. Weng, R. Negi, and M. D. Ilić, "Probabilistic joint state estimation for operational planning," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 601–612, Jan. 2019.

[39] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC Press, 2004.

[40] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust data-driven state estimation for smart grid," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1956–1967, Jul. 2017.

[41] 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 Trans. Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[42] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2016–2025, Jul. 2016.

[43] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, Jul. 2017.

[44] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. Red Hook, NY, USA: Curran Assoc., 2014, pp. 2672–2680.

[45] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 214–223.

[46] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (Adaptive Computation and Machine Learning Series). London, U.K.: MIT Press, 2016.

[47] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2158–2169, Mar. 2019.

[48] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[49] "Hourly Load Data Archives." Electric Reliability Council of Texas (ERCOT). [Online]. Available: http://https://www.ercot.com/gridinfo/load/load_hist/ (Accessed: Jun. 14, 2022).

[50] V. Ajjarapu and C. Christy, "The continuation power flow: A tool for steady state voltage stability analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 416–423, Feb. 1992.

[51] F. Milano, "Continuous Newton's method for power flow analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 50–57, Feb. 2009.

[52] M. S. Shahriar, I. O. Habiballah, and H. Hussein, "Optimization of phasor measurement unit (PMU) placement in supervisory control and data acquisition (SCADA)-based power system for better state-estimation performance," *Energies*, vol. 11, no. 3, p. 570, 2018.

[53] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–15.

[55] E. Preston and C. Barrows, "Evaluation of year 2020 IEEE RTS generation reliability indices," in *Proc. IEEE Int. Conf. Probabilistic Methods Appl. Power Syst.*, 2018, pp. 1–5.

[56] Z. Zhang, R. Deng, D. K. Yau, P. Cheng, and J. Chen, "Analysis of moving target defense against false data injection attacks on power grid," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2320–2335, 2019.

[57] S. Lakshminarayana and D. K. Yau, "Cost-benefit analysis of moving-target defense in power grids," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1152–1163, Mar. 2021.

[58] J. Zhao and L. Mili, "Vulnerability of the largest normalized residual statistical test to leverage points," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4643–4646, Jul. 2018.

[59] 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.