Generative Adversarial Networks-based False Data Injection: A Concern for Data Integrity of Power Networks

Rehan Nawaz, Rabbaya Akhtar, Saad Ullah Khan, Siqi Bu, *Senior Member, IEEE,* and Muhammad Habib Mahmood

***Abstract*—The digitalisation of power grids has increased data integrity and security risks, enabling subtle manipulations to cause power theft, false tripping, and compromised controls. Effective false data attacks must exhibit complex nonlinear behaviour, preserve network dependencies, and account for real-time system conditions to evade advanced false data detection (FDD) methods. Constructing such attacks is highly challenging due to the restricted access to critical system information. In this paper, three deep learning-based False Data Injection (FDI) attacks for power networks are proposed, highlighting the vulnerabilities of existing false data detection defences. A holistic comparative analysis of attack constructed by three variants of Generative Adversarial Networks (GANs), including GAN, Wasserstein GAN (WGAN) and Conditional GAN (CGAN), is presented. The proposed false data attacks are assessed against five diverse FDD defences to assess all factors of possible FDI attack failure. IEEE Case-5, Case-14, Case-30 and Case-118 bus systems are simulated as target networks with realistic demand modelling. The simulation results display a progressive improvement of FDI success from GAN to CGAN.**

***Index Terms*—False Data Injection, False Data Detection, Power network cyber attacks, Cybersecurity**

I. Introduction

THE power systems are continuously evolving towards more reliable operation with enhanced power quality, automated controls and self-healing properties. However, these advanced control infrastructures set forward a high dependency on complex, wide-span, real-time communication infrastructure for coordination between automated equipment, rendering power networks more prone to threats such as communication lags, failures, false communication, data integrity attacks and False Data Injection (FDI) etc. These attacks not only can cause havoc for the targeted equipment but may also threaten the security of the whole society connected to it [1].

The Measurement Units (MUs) perform real-time infield measurements and communication of measured state parameters to the control units, which in turn receive, process and analyse the received data to initiate appropriate control action by actuators, if necessary. The control units (CUs) are responsible for maintaining network health through appropriate control actions based on the data from MUs. FDI Attacks target to deceive or mislead the CUs by manipulating the communicated data, resulting in the attacker's desired actions.

The advancements in False Data Detection (FDD) techniques make it challenging to design an effective FDI attack as randomly injected false data can be detected easily as an anomaly. Therefore, one false entry within real-time series data must satisfy all the dependencies and constraints of typically measured data to cover the inconsistency from actual data. Both FDI and FDD mechanisms operate concurrently; to evaluate emerging FDD methods effectively, it becomes imperative to develop more sophisticated FDI techniques to rigorously challenge the capabilities of FDD defence and highlight any potential weaknesses, ultimately strengthening the security measures for smart grids.

Recent research highlighted the increasing need for robust cyber security measures and enhanced countermeasures to secure the communication infrastructure of power systems [2]. State estimation-based methods, such as Bad Data Detection using AC state estimation (BDDAC) [3], [4] and Bad Data Detection using DC state estimation (BDDDC), were extensively explored. However, BDDDC was found to be ineffective when false data was crafted using the Jacobian matrix [6], [7], a technique commonly referred to as a Stealthy Attack (SA) or Online Data Integrity Attack (ODIA) [8].

Various methods were proposed in the literature to construct stealthy attacks, as these attacks required only a linear approximation of the Jacobian matrix. For instance, stealthy attacks could be generated using the measurement matrix alone [9]. A convex optimisation problem was formulated in [10] to create stealthy attacks using measurements from a limited number of

This work was supported in part by the Hong Kong Research Grants Council under Grant 15205424, and in part by The Hong Kong Polytechnic University through the Research Student Attachment Programme.

Rehan Nawaz is with the Department of Electrical and Computer Engineering, Air University, Islamabad, Pakistan, and also with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: rehan._.nawaz@outlook.com).

Rabbaya Akhtar is with the School of Energy Systems, Lappeenranta-Lahti University of Technology (LUT), Lappeenranta, Finland (e-mail: rabbaya.akhtar@lut.fi).

Saad Ullah Khan is with the Department of Electrical and Computer Engineering, Air University, Islamabad, Pakistan (e-mail: saadullahkhan@au.edu.pk).

Siqi Bu is with Department of Electrical and Electronic Engineering, Shenzhen Research Institute, Research Centre for Grid Modernisation, Research Institute for Smart Energy, Policy Research Centre for Innovation and Technology, International Centre of Urban Energy Nexus, and Centre for Advances in Reliability and Safety, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: siqi.bu@polyu.edu.hk).

Muhammad Habib Mahmood is with Teradata Corporation, Islamabad, Pakistan (e-mail: muhammadhabib.mahmood@teradata.com).

0000–0000/00$00.00 © 2021 IEEE

sensors. It was also demonstrated that such attacks could be constructed with partial information about the local area [11]. The Least Effort Attack (LEA) was introduced to minimise resource usage, enabling stealthy false data injection with minimal effort and using a minimal number of sensors [12]. Additionally, an innovative approach called the ICA Attack (ICAA) was developed, which constructed stealthy attacks without requiring prior network information by leveraging Independent Component Analysis (ICA) [13].

Despite their sophistication, stealthy attacks could be effectively identified by machine learning-based detection algorithms, i.e., False Data Detection using Support Vector Machine (FDDSVM) using Principal Component Analysis (PCA) for feature reduction and Gaussian Kernel Function within SVM framework demonstrated high accuracy. [14], [15]. However, FDIA using Linear Regression (FDILR) [6] and FDIA using Linear Regression with time stamp (FFDILRT) [7] bypassed FDDSVM and claimed that any machine learning-based FDD algorithm can be bypassed if the FDI attack is also machine learning based. However, these attacks failed to perform well against BDDAC.

AC state estimation, with its highly nonlinear nature and the lack of an analytical solution, is a more challenging defender. In [16], Cyber-Attacks using Graph Theory (CAGA) and line parameter data were proposed and tested against AC state estimation. FDI derived from admittance and susceptance of the target bus and the associated transmission lines, two attack strategies were introduced: Unobservable FDI Attacks on Bus (UFDIAB) and Unobservable FDI Attacks on Super-Bus (UFDIASB) [17]. In [18], a Topology Learning Aided FDI Attack (TLAFDIA) based on AC state estimation parameters was proposed. Perfect and Imperfect Attacks (PIA), an FDI approach with varying knowledge of states, were introduced in [19]. Network Parameter Coordinated False Data Injection (NP-FDI) was explored in [20], formulating optimisation problems to minimise the number of compromised measurements. Similarly, [21] presented False Data Injection Attacks Against Synchronization Systems in Microgrids (FDIASSM). Evolutionary techniques such as genetic algorithms and neural networks were utilised in [22] to construct least-effort attack vectors. An FDI technique was introduced in [23], estimating a nonlinear function $h(x)$ mapping the state vector to the measurement vector using weighted least squares on historical measurement data. In [7], FDI using Delta Thresholds (FDIDT) was proposed, where prior measurement vectors from a dataset were injected as false measurements. A GAN-based approach in [24] developed FDI Attacks by extracting the physical model (FDIAPM) using historical data. The Auto Encoder and WGAN-based Attack (AE-WGAN), proposed in [25], employed an autoencoder to learn the unknown state estimator model, used with a WGAN to derive the original data distribution. The Neural Network and GAN-based Attack (NN-GAN) in [26] applied a similar two-stage process. In [27], Cyber Attacks on Demand-Response (CADR) in renewable integrated Smart Grids were introduced. Optimal Data Injection Attacks (ODIA) leveraging matrix theory to transform the objective of maximising the trace of remote estimation error covariance into a solvable convex optimisation problem were formulated for cyber-physical systems, considering resource constraints [28]. Optimal FDIA (OFDIA) was framed utilising an optimal state feedback injection law [29]. Event-based Stealthy FDI Attack (ESFDIA) was proposed in [30], allowing precise false data injection to evade BDDAC, while a State-perturbation-based Adversarial Example and FDIA (S-AFDIA) was introduced in [31] as another strategy to bypass BDDAC. An Optimal Stealthy Attack modelled by Linear Quadratic Gaussian dynamics (OSA-LQG) was proposed in [32], enabling attackers to maximise quadratic cost by intercepting and altering transmitted measurements.

While these FDI techniques performed well against BDDAC, they were susceptible to detection by advanced methods incorporating time-series dependencies. For instance, the Nonlinear Function-based Variable Dummy Value Model (NF-VDVM) [33], [34] and False Data Detection using Temporal Behaviours (FDDTB) [35], [36] demonstrated the ability to detect these advanced attacks, highlighting the need for more robust FDI techniques.

Despite the significant contributions of the techniques presented for FDI, some grey areas are insufficiently addressed in the previous literature. The most common is that,

a. *Use of practically inaccessible information*
   Most FDIs construct attack vectors by models dependent on the inside system information of the scope, which is practically inaccessible to an outsider
b. *DC approximations*
   Most FDIs lack practicality as they are modelled or assessed through DC approximations-based models
c. *Consideration of Real-time system state*
   Most FDIs do not consider the power system's continuously varying state, e.g., highly diverse demand patterns
d. *Lack of comprehensive testing*
   Different FDD algorithms assess different aspects of FDIs; an undetectable FDI under state estimation FDD may not necessarily be successful under FDDTB or NF-VDVM, so comprehensive testing is the key.
e. *Impractically lenient safe thresholds*
   The safe threshold for AC state approximation-based testing is kept impractically lenient to display the falsely improved performance of the proposed attack
f. *Non separation of FDI and FDD sides*
   Most FDIs use the same dataset for training both FDI&D algorithms, which defies the fact that in the real world, attacking and defending sides are blind to each other
g. *Consideration of time series data patterns*
   The variation in parameters with time, i.e., the daily demand routine and seasonal variations in power consumption, are often neglected while modelling the attack vector

Considering the aforementioned research gaps, the main contributions of this paper are as follows:

- This study introduces three novel deep learning-based FDI methods: False Data Injection using Generative Adversarial Network (FDIGAN), False Data Injection using Wasserstein Generative Adversarial Network (FDI-WGAN), and False Data Injection using Conditional Gen-

erative Adversarial Network (FDICGAN). These methods are void of approximations and entirely focus on historical system measurement vectors, eliminating the need for inside information like system topology and line parameters, thus enhancing practicality.
- The proposed FDI methods incorporate demand variability and leverage a blind dataset for training and testing, effectively simulating real-world attack and defence scenarios.
- The effectiveness of the proposed FDI methods is rigorously evaluated against diverse FDD techniques, including BDDDC, FDDSVM, FDDTB, NF-VDVM, and BDDAC under stringent detection thresholds to evaluate different aspects such as inter and intra dependencies, safe data patterns, time series dependencies, nonlinear predefined hidden correlations, respectively. A comparative analysis of all three proposed FDI techniques is presented.
- By exposing specific vulnerabilities in existing detection mechanisms, this work not only highlights the limitations of current FDD methods but also offers critical insights for advancing the development of more robust and resilient FDD defences.

The rest of the paper is organised as follows: Section-II explains the proposed attack configuration and implemented model in detail. Section-III summarises the facts about test systems, test cases and the resulting outcomes, and a comparative analysis of all proposed attacks is defined in Section-IV. Section-V sums up the accomplishments, shortcomings and future work aspects of the proposed FDI attacks.

II. Proposed Technique

The Raw data from MUs is extracted, combined with pseudo measurements and transferred using data exchange techniques, resulting in a defined vector known as a measurement vector [37]. The constituents of a measurement vector M at time instant t are shown in Eq. 1.

$$M^t = \begin{bmatrix} V^t & \varphi^t & P_i^t & Q_i^t & P_{ij}^t & Q_{ij}^t & P_{ji}^t & Q_{ji}^t \end{bmatrix} \quad (1)$$

Here V^t, φ^t, P_i^t and Q_i^t represent vectors of measured bus voltage magnitudes, bus voltage phase angles, measured real power injections and measured reactive power injections to all buses at time instant t respectively, with the dimension of $1 \times m$ for m bus system. Moreover, P_{ij}^t, Q_{ij}^t, P_{ji}^t and Q_{ji}^t are the vectors of forward and reverse real and reactive powers respectively, of all transmission lines at time instant t with the dimension of $1 \times n$. n is the number of transmission lines. Hence, M^t has an overall dimension of $1 \times (4m + 4n)$. The measurement vectors aggregate over t instants to form a matrix $\mathbf{M}$ as shown in Eq. 2.

$$\mathbf{M} = \begin{bmatrix} M^1 & M^2 & \dots & M^t \end{bmatrix}^T \quad (2)$$

This measurement matrix M with dimension $t \times (4m + 4n)$ is used as a training dataset for FDIA to assimilate false data vectors of the power network. This paper utilises a relatively modern deep learning paradigm, Generative Adversarial Network, and its modified variants for constructing successful FDI attacks after thoroughly learning the nonlinear distributions of training data. FDI constructs false measurement vectors closely replicating real measurement vectors; any preprocessing of the diversified on-field measurements to form a complete measurement vector is outside the scope of this work.

A. FDI Attacks

The effective FDI attack creates deceptive measurement vectors, altering specific parameters while maintaining dependencies such as Kirchhoff's laws and time series patterns for stealthiness. Initially, machine learning techniques such as linear regression and multivariate polynomial regression proved impractical due to the complex nature of power system data [6], [7]. Consequently, deep generative models were pursued, starting with autoencoders transitioning to GANs due to their superior capabilities. Proposed FDI attacks are established using GAN and its two modified variants, WGAN and CGAN. The comparative analysis of resulting attack scenarios is presented by testing through various state-of-the-art FDD techniques. False Negative Rate (FNR) is used as an evaluation metric which portrays Bypassing Accuracy (BA) against FDD methods as defined in Eq. 3.

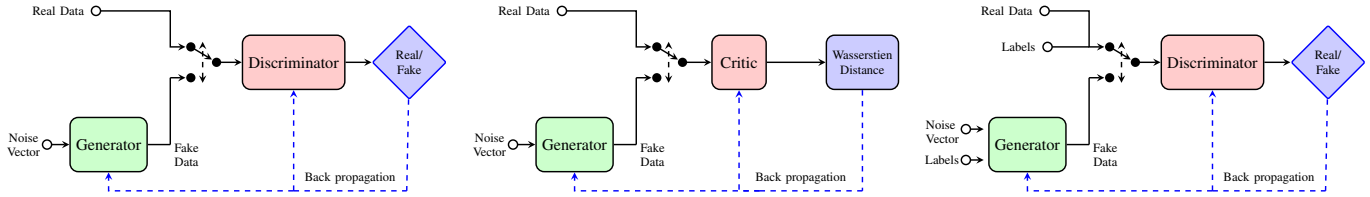
$$\text{BA} = \frac{\text{False Negative}}{\text{False Negative} + \text{True Positive}} \quad (3)$$

1) FDIGAN: GAN-based false data attack vectors for targeted power systems are assembled; a class of deep generative models consists of two Convolutional neural networks: a generator (G) and discriminator (D). Both generator and discriminator compete with each other through zero-sum games, where one's loss is another's gain. The generator and discriminator are trained simultaneously through mini-batch stochastic gradient descent. For the mini-batch of m, the number of training examples and noise samples. The generator works on learning the linear and highly nonlinear distributions in existing data to generate unseen data following identical distributions, while the discriminator trains to differentiate the fake data from the real data. GANs are efficient with high-dimensional data with highly complex, nonlinear feature relationships. Binary cross-entropy fosters competition in GANs training through the min-max game as shown in Eq. 4.

$$\begin{aligned} \min G \max D \; V(D,G) = \; & \mathbf{E}_{x \sim p_{data}(x)} \left[\log D(x)\right] \\ & + \mathbf{E}_{z \sim p_z(z)} \left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right] \end{aligned} \quad (4)$$

Here, $D(x)$ is the discriminator's output for a real sample x, $D(G(z))$ is the discriminator's output for the generated sample $G(z)$ from input noise vector z, $\mathbf{E}$ is expectation and p_{data} and p_z are the probability distribution of the original data and estimated probability distribution respectively. The Minmax GAN loss involves the simultaneous optimisation of both the D and G models. The discriminator maximises the Eq. 5 to penalise itself for misclassification of real and fake instances.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \log D(x^i) + \log(1 - D(G(z^i)) \quad (5)$$



(a) Functional block diagram of GAN (b) Functional block diagram of WGAN (c) Functional block diagram of CGAN

Fig. 1: Block diagram of Generative Adversarial Networks and its variants

Subsequently, the generator receives a reward if it successfully deceives the discriminator but faces penalties otherwise. Training the generator involves minimising the Eq. 6.

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^i)) \tag{6}$$

This recurrent process persists until the global optimality of $p_g = p_r$ is reached [38]. p_g and p_r are the generator's estimated and real data distribution, respectively. The operation of the GAN model is graphically displayed in Fig. 1a. The constituent measurement vectors from the measurement matrix are treated as real training samples, jointly following a highly nonlinear unknown distribution. The proposed FDI employs GAN to learn the unknown distributions and correlations of real data until the generated data is refined to an extent where the discriminator fails to distinguish it from real data. The resultant trained generator model is saved for constructing FDI vectors for attack.

GAN uses JS divergence to measure the similarity between the actual and estimated distribution, which is the weighted sum of forward and backward KL divergence [38]. However, the JS divergence exhibits zero or vanishing gradients with non-overlapping distributions. The Mode Collapse in GANs due to multi-modal distributions as suggested by nonlinear load flow equations, complex connectivity of the network and inter-dependency of parameters in training data, and the issue of vanishing gradient due to JS divergence limit the training capability. This limitation is addressed by opting for WGAN.

2) FDIWGAN: The mode collapse issue is addressed in WGAN [39]; the cost function is based on Wasserstein distance, i.e., the minimum energy required to horizontally move the estimated probability distribution p_g to the actual distribution p_r.

$$W(p_r, p_g) = \inf_{\gamma \in \prod(p_r, p_g)} \mathbf{E}_{(x,y)\in\gamma} \|x - y\| \tag{7}$$

Here $\mathbf{E}$ is the expectation, $\Pi(p_r, p_g)$ is the transport plan of probability mass over x and y to transform p_r to p_g. The Wasserstein distance is then the cost of the optimal transport plan. Linear programming is used to find the optimal solution. WGAN introduces a critic with the same role as the discriminator network. However, it predicts the Wasserstein distance, with objective function in Eq.8, instead of classifying the samples coming from the distribution p or q.

$$f_{critic}(w) = \max_{w \le W} \mathbf{E}_{x\sim p_r}[f_w(x)] - \mathbf{E}_{z\sim P(z)}[f_w(G(z))] \tag{8}$$

By introducing a stochastic gradient and applying the expectation over a mini-batch of m number of training examples and noise samples, the generator's training function is given in Eq. 9.

$$\nabla_{\theta} \frac{1}{m} \sum_{i=1}^{m} f(G(z^i)) \tag{9}$$

This process persists until the global optimality of $p_g = p_r$ is reached, and the resultant trained generator model is saved for constructing the FDI attack vectors. The operation of the WGAN model is graphically displayed in Fig. 1b.

The WGAN resolves mode collapse and vanishing gradient problem, but the generator is unaware of the time instant and real-time data patterns, i.e. load demand and power flows at that instant. Therefore, temporal behavioural FDD methods such as FDDTB can still detect these attacks, suggesting further investigation into supervised multi-class learning as training individual WGANs for each time instant proved computationally impractical, resulting in the identification of CGAN as a more practical alternative.

3) FDICGAN: CGAN, being a class of supervised learning, requires labelled data for training where both the generator and the discriminator are conditioned on some auxiliary information, such as class labels. As a result, the model can learn the class-wise multi-modal input-to-output mapping utilised here by treating each time instant as a distinct class and training the generator and discriminator separately for each time instant. The generator of the CGAN has an additional layer that enables data generation for specific classes. The discriminator also has one additional layer that results in the added responsibility of identifying the classes of the real data in addition to distinguishing between real and false data. If l_1^t is the label for measurement vector 1 received at time instant t, then the complete label vector L is shown in 10.

$$L = \begin{bmatrix} l_1^t & l_2^t & l_3^t & \dots & l_n^t \end{bmatrix}^T \tag{10}$$

The measurement matrix $\mathbf{M}$ and label vector L are jointly used as a dataset. The loss function of CGAN is shown in Eq. 11.

$$\begin{aligned} \min G \max D\, V(D,G) = {} & \mathbf{E}_{x\sim p_{data}(x)} [\log D(x|y)] + \\ & \mathbf{E}_{z\sim p_z(z)} [\log (1 - D(G(z|y)))] \end{aligned} \tag{11}$$

Here, the $G(z|y)$ and $D(x|y)$ are conditional probabilities demonstrating the generation and discrimination of data given label l. The generator's update equation is given in Eq. 12.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \log D(x^i|y^i) + \log(1 - D(G(z^i|y^i)) \tag{12}$$

Eq. 13 gives the discriminator's training equation.

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^i|y^i)) \tag{13}$$

The process is repeated till the discriminator fails to distinguish real and fake data, and the trained generator model is saved to generate false data vectors for attack. The operation of the CGAN model is graphically displayed in Fig. 1c.

The primary objective of FDIA is to construct measurement vectors so that they remain indistinguishable from actual data while injecting false measurements into the power system. These falsified measurement vectors can mislead the control mechanisms, potentially leading to inaccurate decision-making, operational inefficiencies, or even system instability. The specific objectives of FDIA include energy theft, false tripping, load shedding and intentional delays or blocking of data. The methodology involves manipulating target variables within the measurement vector to achieve a particular objective while simultaneously adjusting other parameters to maintain consistency with the power system's inherent dependencies. For example, energy theft can be achieved by falsifying load measurements, while voltage manipulation can result in improper protection system activation or control errors. The simulation results substantiate the capability of attack vectors constructed by FDICGAN to achieve any malicious intent of the attacker without being detected by commonly employed FDD defences. To illustrate the objective of the proposed FDIA, attack vectors are generated to target energy theft through misleading information by displaying less power consumption.

III. Simulation Scope

A simplified depiction of the presented work is shown in Fig. 2; the simulations are carried out using programming environments of MATPOWER 7.1, MATLAB 2020a, and Python 3.9.12 on a system with specifications of Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz (8 CPUs), 16GB RAM, Windows 11 Pro 64-bit, GPU-NVIDIA GeForce RTX 2080.

Datasets for IEEE test systems used are generated by the MATPOWER power flow program for 1 year, considering $\pm 10\%$ variation for a single instant with 0.1% white noise to make the data non-ideal. The generative models for proposed FDI algorithms are trained in Python, and the false measurement vectors constructed by trained models are injected into the target power system. The defence side FDD methods are implemented in MATLAB. The total number of samples for any parameter of all three systems over one year is 48×365. The data set for FDI contains $17,520$ measurement vectors. Another dataset is generated with a variation of around ± 12-15% for a single time instant for one year, which is used to train FDD algorithms. This approach ensures practicality by keeping the attacker and defender blind to each other. The FDI side trains GAN, WGAN, and CGAN, while the FDD side incorporates the training of SVM and NF-VDVM,

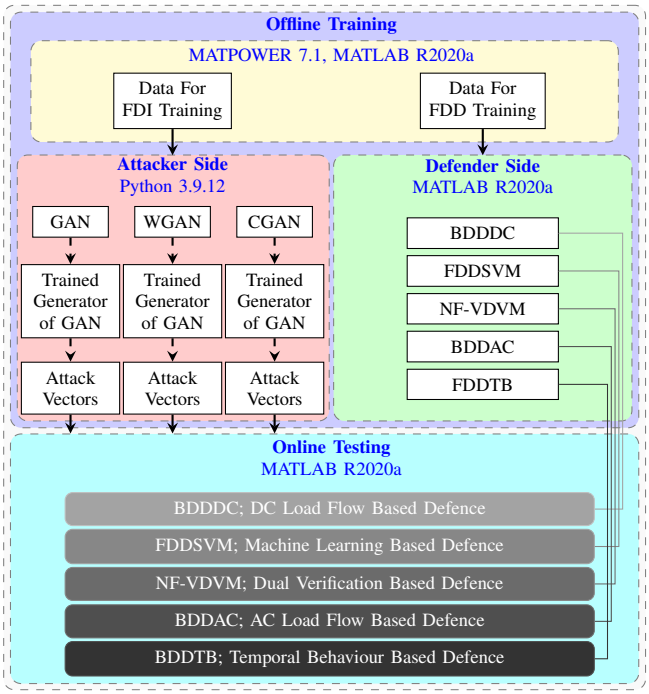


Fig. 2: Overview of proposed FDIs against trained FDDs

the evaluation of maximum safe thresholds for BDDDC and BDDAC and the assessment of $\Delta_{\min}$ and $\Delta_{\max}$ for FDDTB.

TABLE I: Computational cost of offline training (in hours) and online attack vector generation (in milliseconds) for the proposed FDI methods.

FDI Method	Offline Training Time *(h)*	Online Attack Gen. Time *(ms)*
FDIGAN	25∼30	60∼ 100
FDIWGAN	30∼35	90∼130
FDICGAN	120∼125	150∼200

The simulations consider the modified IEEE case-5 bus system (5 buses and 6 transmission lines), IEEE case-14 bus system (14 buses and 20 lines), IEEE case-30 bus system

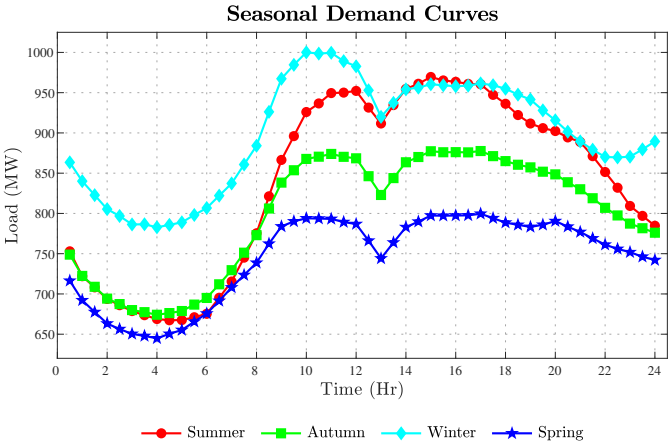
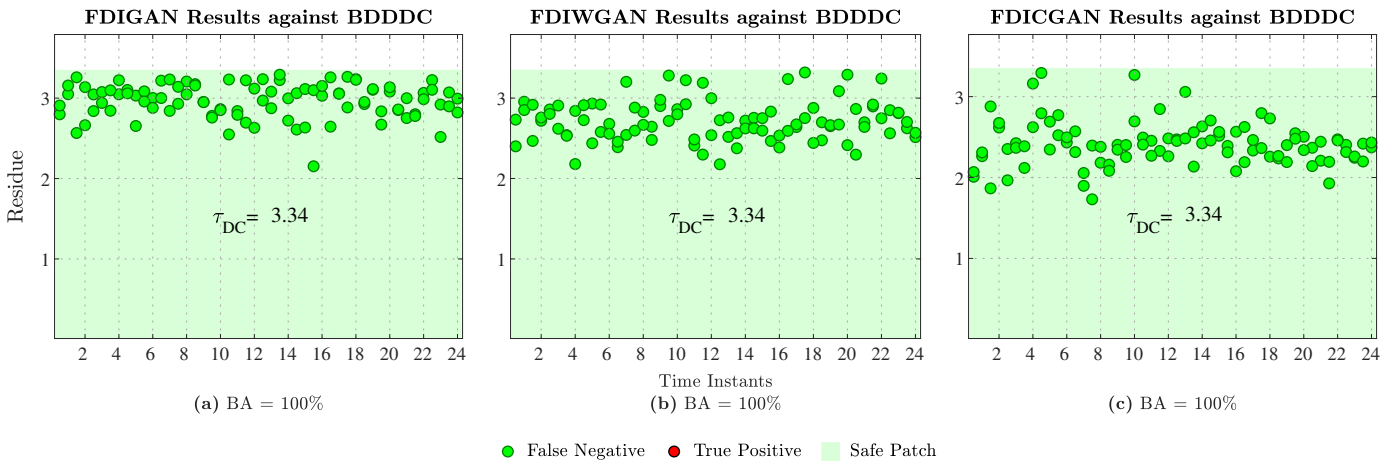
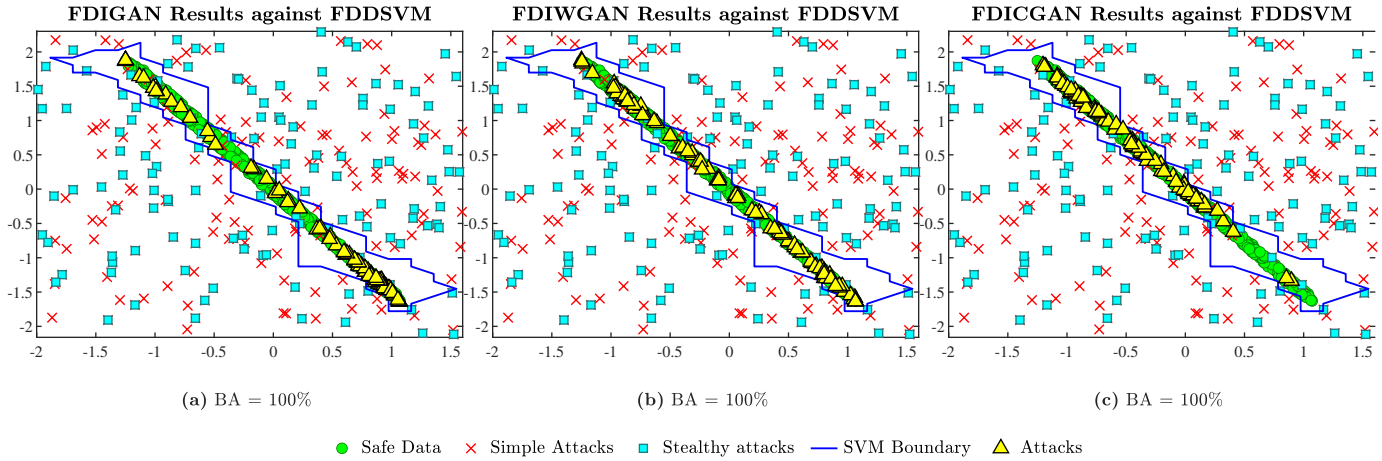


Fig. 3: Average daily demand curves considering seasonal load variation.



(a) BA = 100% (b) BA = 100% (c) BA = 100%

Fig. 4: Performance of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against BDDDC.



(a) BA = 100% (b) BA = 100% (c) BA = 100%

Fig. 5: Performance of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against FDDSVM.

(30 buses and 41 lines) and IEEE Case-118 bus system (118 buses and 186 lines) as target power networks producing a single measurement vector of 1×44, 1×136, 1×284 and 1×1216, respectively. To account for load dynamics, the realistic data of energy demand of South Korea is utilised from [40]. The daily demand curve is based on the mean demand with seasonal variations and 30 30-minute time step, mapped to IEEE systems load ratings as shown in Fig. 3.

IV. Results and Discussion

The proposed FDI models are trained offline for up to 10,000 epochs with a batch size of 64, and the associated computational costs are presented in Table I. A total of 96 attack vectors for two days, considering a time step of 30 minutes, are constructed by each trained generator model. To induce the FDIA, these attack vectors are sequentially injected into the target power system and then rigorously assessed by several advanced FDD algorithms.

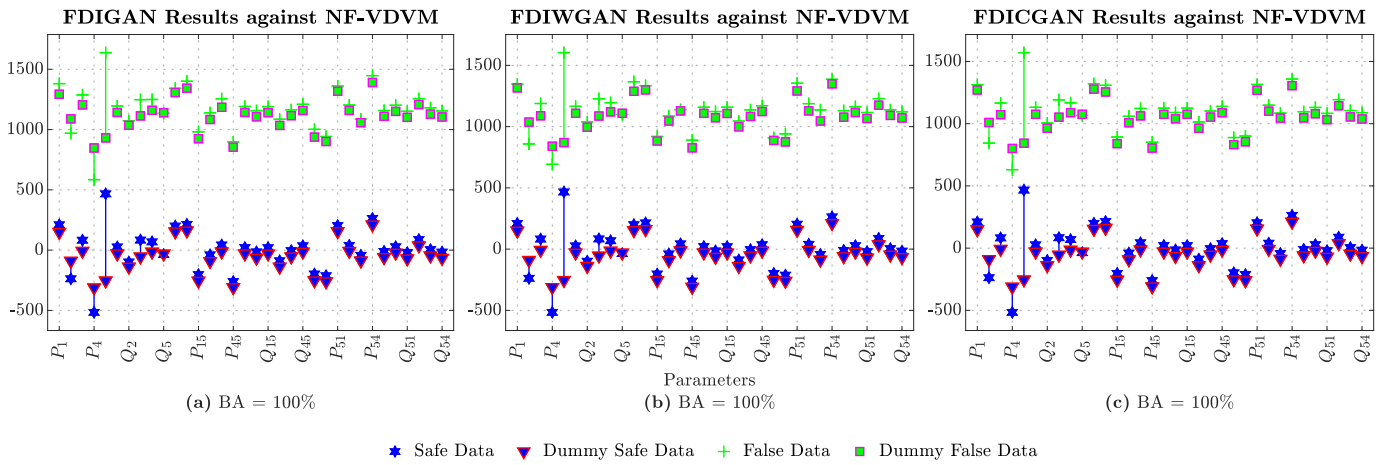
A. Performance of proposed FDIs against different FDDs

The following discussion presents a holistic analysis of performance stats for all three proposed FDIs against various FDD defences.

1) Performance against BDDDC: The safe threshold τ_{DC} from definite real data is evaluated to be 3.34. Fig. 4 displays the comparative performance of induced attacks against BDDDC defence. The residues of all 96 false measurement vectors from all three FDIs lie under the safe threshold. Hence, all three proposed attacks can successfully bypass BDDDC with perfect BA of up to 100%.

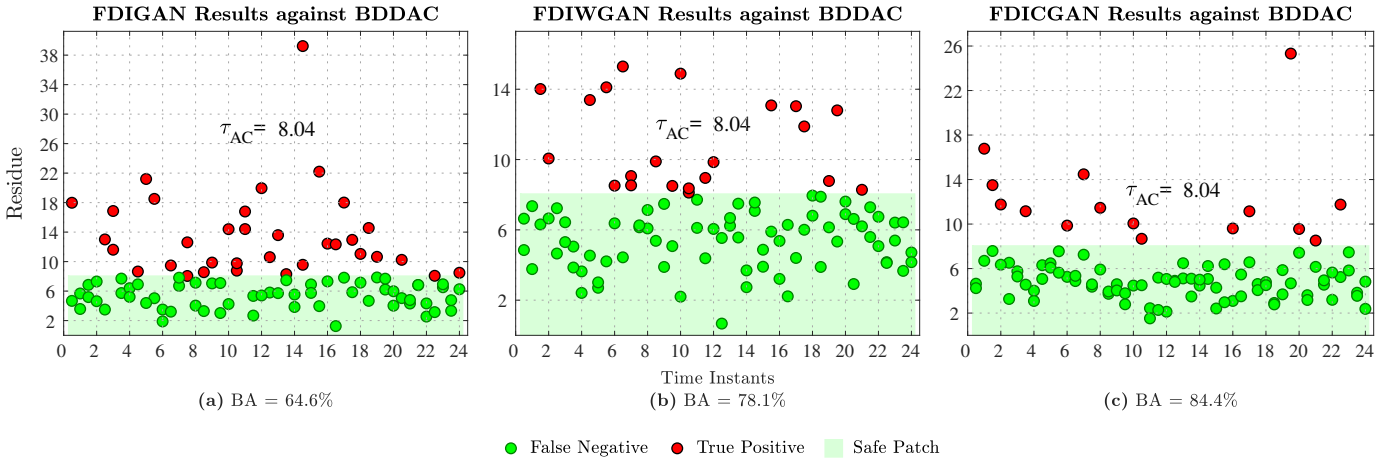
2) Performance against FDDSVM: FDDSVM is a machine learning-based defence. As evident from Fig. 5, all the attack vectors constructed through proposed attack models are well within safe bounds, suggesting that FDDSVM completely fails to detect the proposed FDIs as the attacks are also based on similar learning algorithms.

3) Performance against NF-VDVM: NF-VDVM is a dual verification-based defence used as a test bench with double-sized measurement vectors, so the FDI models are modified to generate measurement vectors with the updated dimensions. The results of a measurement vector from all three FDI techniques, without showing state variables, are displayed in Fig. 6; deep learning-based FDI techniques can effectively evade NF-VDVM with perfect BA of up to 100% by successful learning of the hidden correlation between the dual vectors of actual data.



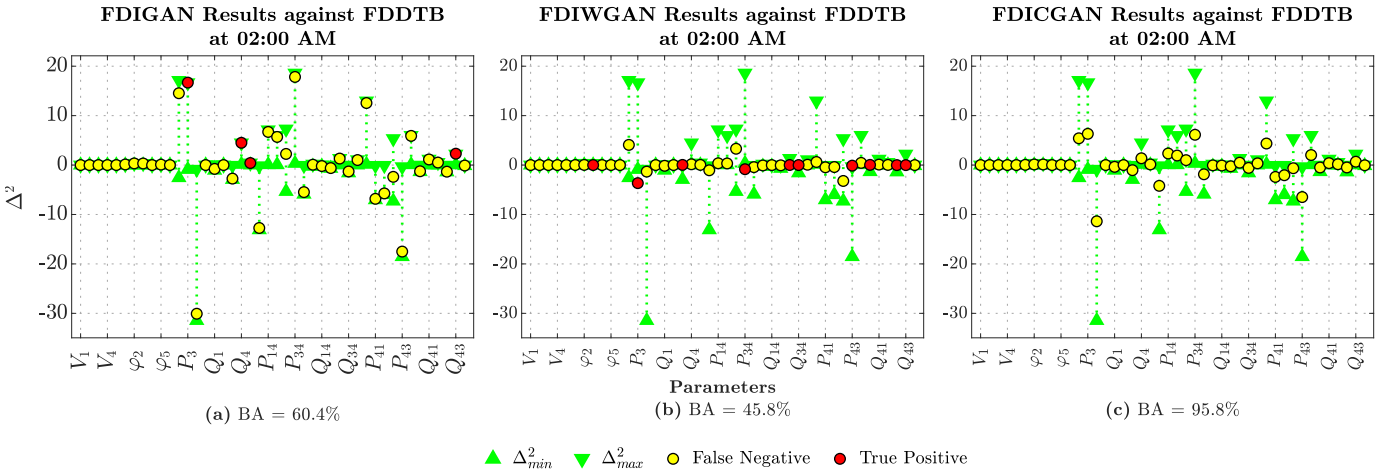
(a) BA = 100% (b) BA = 100% (c) BA = 100%

Fig. 6: Performance of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against NF-VDVM.



(a) BA = 64.6% (b) BA = 78.1% (c) BA = 84.4%

Fig. 7: Performance of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against BDDAC.



(a) BA = 60.4% (b) BA = 45.8% (c) BA = 95.8%

Fig. 8: Performance of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against FDDTB.

4) Performance against BDDAC: BDDAC deploys AC load flow analysis; Fig. 7 shows that the safe threshold τ_{AC} is evaluated as 8.04. Due to the strict threshold and the highly nonlinear and complex behaviour of AC state estimation, all three FDIAs are detectable in some instances. However, the BA improves noticeably from GAN to CGAN, i.e., 64.6%, 78.1% and 84.4%, respectively.

5) Performance against FDDTB: FDDTB evaluates the induced FDIs through the assessment of temporal behaviours and time series dependencies. Fig. 8 compares the performance results of FDIAs for a single time instant. In the case of FDIGAN and FDIEWGAN, some parameters failed to adhere

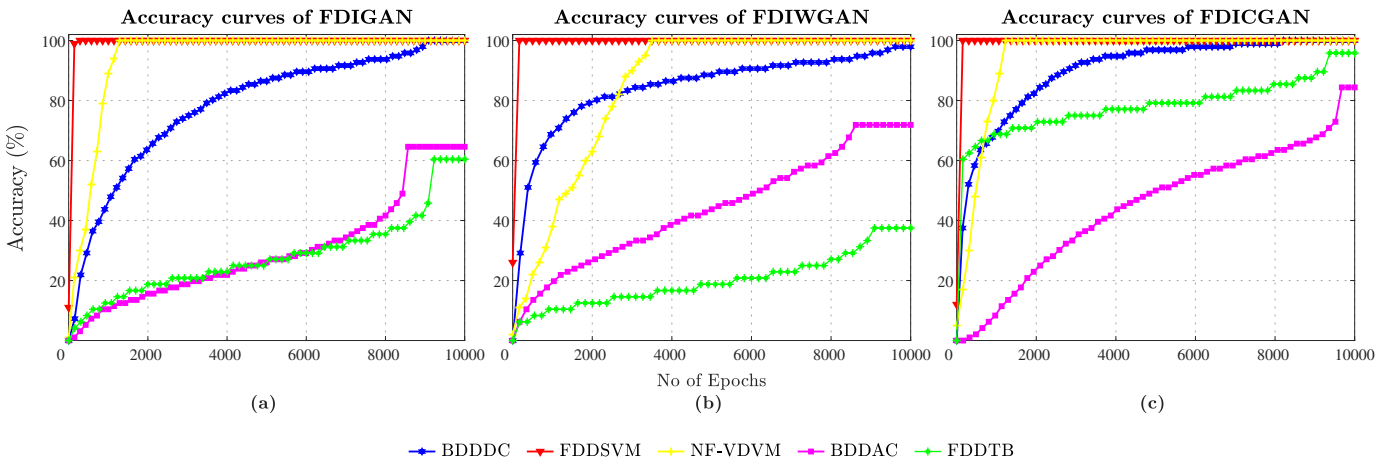


Fig. 9: Comparison of bypass accuracies of (a). FDIGAN, (b). FDIWGAN and (c). FDICGAN against FDD algorithms.

to the safe bounds due to lack of time instant information; overall BA of 60.4% and 45.8% is achieved, respectively. However, the FDICGAN, trained with time step labels, shows significantly promising improvement with a BA of 95.8% against FDDTB, which proves its effectiveness in attack vector construction of any specific time instant.

To summarise, Fig. 9 shows a comparison of bypassing accuracies of all three proposed FDI attacks against stated FDD defences with respect to the number of epochs and the quantitative results of the mean bypassing accuracies on all the stated test systems are shown in TABLE II. It can be observed that GAN, being an ML/DL approach, proved effective in the case of BDDDC, FDDSVM and NF-VDVM. However, the lack of time consideration and highly nonlinear complex multi-modal distributions proved to be a barrier in the case of FDDTB and BDDAC, respectively. Fig. 9 illustrates that the curves of BDDDC and BDDAC for FDIWGAN demonstrate better results as compared to GAN. However, the inadequacy against FDDTB due to the random injection of attack vectors persists. This issue is addressed by using the multi-class supervised generative model CGAN, which has exhibited remarkable success in BA against all FDD defences, as shown in Fig. 9 and TABLE II. Moreover, the attributes

TABLE II: Comparative analysis of the performance of proposed FDI techniques with previous FDI techniques in view of some state-of-the-art FDD Algorithms.

Methods	Ref.	A	B	C	D	E
SBFDI	[41]	✗	✗	✗	✗	✗
UMBFDI	[41]	✗	✗	✗	✗	✗
SUBFDI	[41]	100%	✗	✗	✗	✗
SA	[8]	100%	✗	✗	✗	✗
LEA	[12]	100%	✗	✗	✗	✗
ICAA	[13]	100%	✗	✗	✗	✗
PCAA	[42]	100%	✗	✗	✗	✗
LAIA	[11]	100%	✗	✗	✗	✗
FDILR	[6]	100%	100%	✗	✗	41.0%
FDILRT	[7]	100%	100%	✗	✗	66.7%
FDIDT	[7]	100%	100%	✗	100%	76.9%
AE-WGAN	[25]	100%	-	-	89.3%	✗
NN-GAN	[26]	100%	-	-	90%	✗
FDIGAN	Prop.	100%	100%	100%	64.6%	60.4%
FDIWGAN	Prop.	100%	100%	100%	78.1%	45.8%
FDICGAN	Prop.	100%	100%	100%	84.4%	95.8%

A: BDDDC **B**: FDDSVM **C**: NF-VDVM **D**: BDDAC **E**: FDDTB
✗: Failed to Bypass -: Not tested on respective FDD

achieved by the proposed FDIs are summarised in TABLE III, which supports the test results and provides a comparative

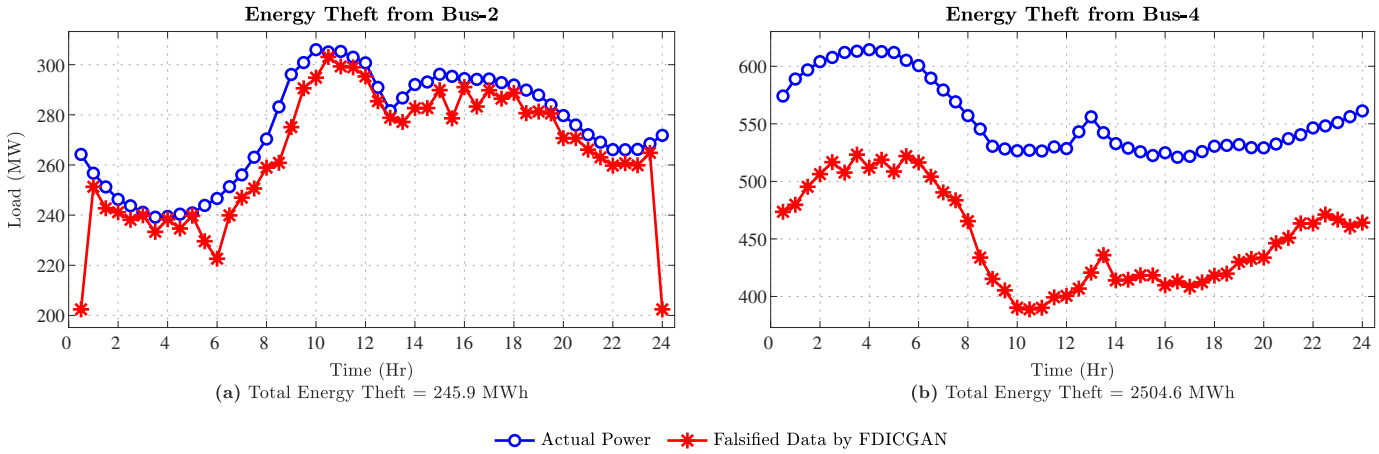


Fig. 10: Undetected energy theft in a day trough FDICGAN in IEEE case-5 target bus system.

TABLE III: Comparison of FDI Techniques regarding key attributes for effective practical application.
A: Attack is independent of any inside system information
B: Attack is void of any approximations
C: Real-Time demand variation is considered
D: Attack is tested through BDDAC
E: Strict safe thresholds for detection
F: Blind data for attack and defence
G: Consideration of time series patterns

FDI Techniques	A	B	C	D	E	F	G
ODIA [8], LAIA [11], LEA [12], ICAA [13], FDILR [6], FDILRT [7]	✓	✗	✗	✗	✗	✗	✗
CAGA [16], UFDIAB [17], UFDIASB [17], TLAFDIA [18], PIA [19], NP-FDI [20], FDIASSM [21]	✗	✓	✗	✓	✓	✗	✗
FDIAPM [24], FDIDT [7]	✓	✓	✗	✓	✗	✗	✗
AE-WGAN [25], NN-GAN [26], CADR [27], ODIA [28], OFDIA [29], ESFDIA [30], S-AFDIA [31], OSA-LQG [32]	✓	✓	✓	✓	✓	✗	✗
Proposed FDIGAN, FDIWGAN	✓	✓	✓	✓	✓	✓	✗
Proposed FDICGAN	✓	✓	✓	✓	✓	✓	✓

✗: Not Accounted ✓: Accounted

analysis with existing schemes, highlighting areas for further improvement in FDD defences.

Energy theft incidents considering Bus-2 and Bus-4 as target buses of the IEEE Case-5 bus system are depicted in Fig. 10. Throughout the day, a cumulative undetected energy theft of 2750.5 MWh is achieved through FDICGAN.

V. Conclusion

This article proposes three effective deep learning-based FDI techniques. The efficacy of the proposed methods is endorsed by simulation results that show noticeable performance against several diverse FDD algorithms, particularly FDICGAN. The proposed attack schemes enable the attacker to effectively manipulate the control units based on well-designed attack vectors without using any sensitive inside information of the target power system. The proposed attack mechanisms adapt well to demand variations and changing load flows. Critically implemented comparison of proposed techniques with previously presented schemes and with each other shows the limitations and progressive improvements, highlighting the FDICGAN as the most effective attack. Testing against diverse FDD defences proves the ability to generate false data to manipulate the system without being detected by existing FDD defences. Proposed undetectable FDIA can severely disrupt systems by compromising data integrity, misleading decision-making, introducing security risks, causing financial losses, and jeopardising safety in critical infrastructure, as demonstrated by an implemented example of energy theft. The emphasis is to highlight potential vulnerabilities of existing FDDs and that the cyber security measures to counter the consequences should be prioritised to safeguard the systems and data. The only limitation of the proposed attack is that it deems constant power generation, i.e., the impact of distributed generation in smart grids is not considered. However, this aspect is targeted in future work along with improving FDD methods for detecting the proposed attacks.

References

[1] Z. Wang and Y. Wang, "Pulse-coupled oscillators resilient to stealthy attacks," *IEEE Tran. on Signal Processing*, 2018.
[2] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Tran. on Smart Grid*, 2011.
[3] 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 Tran. on Smart Grid*, 2017.
[4] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network info." *IEEE Tran. on Smart Grid*, 2015.
[5] F. F. Wu and W.-H. Liu, "Detection of topology errors by state estimation (power systems)," *IEEE Tran. on Power Systems*, 1989.
[6] R. Nawaz, M. A. Shahid, I. M. Qureshi, and M. H. Mehmood, "Machine learning based false data injection in smart grid," in *1st International Conference on Power, Energy and Smart Grid (ICPESG)*, 2018.
[7] R. Nawaz, R. Akhtar, M. A. Shahid, I. M. Qureshi, and M. H. Mahmood, "Machine learning based false data injection in smart grid," *International Journal of Electrical Power & Energy Systems*, 2021.
[8] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, 2014.
[9] S. Tan, W.-Z. Song, M. Stewart, J. Yang, and L. Tong, "Online data integrity attacks against real-time electrical market in smart grid," *IEEE Tran. on smart grid*, 2016.
[10] M. G. Kallitsis, G. Michailidis, and S. Tout, "Correlative monitoring for detection of false data injection attacks in smart grids," in *International Conference on Smart Grid Communications*. IEEE, 2015.
[11] 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 Tran. on Smart Grid*, 2017.
[12] Q. Zhang, Z. Li, C. Seatzu, and A. Giua, "Stealthy attacks for partially-observed discrete event systems," in *IEEE 23rd International Conference on Emerging Technologies and Factory Automation*. IEEE, 2018.
[13] M. Esmalifalak, H. Nguyen, R. Zheng, L. Xie, L. Song, and Z. Han, "A stealthy attack against electricity market using independent component analysis," *IEEE Systems Journal*, 2015.
[14] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, 2014.
[15] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Tran. on Information and System Security (TISSEC)*, 2011.
[16] G. Hug and J. A. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Tran. on smart grid*, 2012.
[17] R. Deng and H. Liang, "False data injection attacks with limited susceptance information and new countermeasures in smart grid," *IEEE Tran. on Industrial Informatics*, 2018.
[18] M. Higgins, J. Zhang, N. Zhang, and F. Teng, "Topology learning aided false data injection attack without prior topology information," in *2021 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2021.
[19] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013.
[20] C. Liu, H. Liang, and T. Chen, "Network parameter coordinated false data injection attacks against power system ac state estimation," *IEEE Tran. on Smart Grid*, 2020.
[21] A. S. Mohamed, M. F. M. Arani, A. A. Jahromi, and D. Kundur, "False data injection attacks against synchronization systems in microgrids," *IEEE Tran. on Smart Grid*, 2021.
[22] C. Tu, X. He, X. Liu, Z. Shuai, and L. Yu, "Resilient and fast state estimation for energy internet: a data-based approach," *IEEE Tran. on Industrial Informatics*, 2019.
[23] M. Du, G. Pierrou, X. Wang, and M. Kassouf, "Targeted false data injection attacks against ac state estimation without network parameters," *IEEE Tran. on Smart Grid*, 2021.
[24] R. Jiao, G. Xun, X. Liu, and G. Yan, "A new ac false data injection attack method without network information," *IEEE Tran. on Smart Grid*, 2021.
[25] N. C-E and Y. Weng, "Attack power system state estimation by implicitly learning the underlying models," *IEEE Tran. on Smart Grid*, 2022.

[26] A. B, S. Mishra, and A. Verma, "Deep adversary based stealthy false data injection attacks against ac state estimation," in *2022 IEEE PES 14th Asia-Pacific Power and Energy Engineering Conference*, 2022.
[27] D. Tang, Y.-P. Fang, and E. Zio, "Vulnerability analysis of demand-response with renewable energy integration in smart grids to cyber attacks and online detection methods," *Reliability Engineering & System Safety*, vol. 235, p. 109212, 2023.
[28] G. Wu, J. Sun, and J. Chen, "Optimal data injection attacks in cyber-physical systems," *IEEE transactions on cybernetics*, vol. 48, no. 12, pp. 3302–3312, 2018.
[29] 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, 2018.
[30] H. Guo, J. Sun, Z.-H. Pang, and G.-P. Liu, "Event-based optimal stealthy false data-injection attacks against remote state estimation systems," *IEEE Transactions on Cybernetics*, 2023.
[31] Z. Cheng, H. Ren, J. Qin, and R. Lu, "Security analysis for dynamic state estimation of power systems with measurement delays," *IEEE Transactions on Cybernetics*, vol. 53, no. 4, pp. 2087–2096, 2021.
[32] X.-X. Ren and G.-H. Yang, "Kullback–leibler divergence-based optimal stealthy sensor attack against networked linear quadratic gaussian systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 11 539–11 548, 2021.
[33] M. A. Shahid, F. Ahmad, F. R. Albogamy, G. Hafeez, and Z. Ullah, "Detection and prevention of false data injection attacks in the measurement infrastructure of smart grids," *Sustainability*, 2022.
[34] M. A. Shahid, F. Ahmad, R. Nawaz, S. U. Khan, A. Wadood, and H. Albalawi, "A novel false measurement data detection mechanism for smart grids," *Energies*, 2023.
[35] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Tran. on Smart Grid*, 2014.
[36] S. Bhattacharjee and S. K. Das, "Detection and forensics against stealthy data falsification in smart metering infrastructure," *IEEE Tran. on Dependable and Secure Computing*, 2018.
[37] X. Luo, X. Wang, X. Pan, and X. Guan, "Detection and isolation of false data injection attack for smart grids via unknown input observers," *IET Generation, Transmission & Distribution*, 2019.
[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, 2020.
[39] M. Arjovsky, S. Chintala, and L. B., "Wasserstein generative adversarial networks," in *Inter. conference on machine learning*. PMLR, 2017.
[40] Korea Power Exchange, "Electric Power Statistic Information (EPSIS)," Website, 2022. [Online]. Available: https://epsis.kpx.or.kr/epsisnew/selectEkgeEpsMepRealChart.do?menuId=030300
[41] B. Li, G. Xiao, R. Lu, R. Deng, and H. Bao, "On feasibility and limitations of detecting false data injection attacks on power grid state estimation using d-facts devices," *IEEE Tran. on Industrial Info.*, 2019.
[42] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using pca approximation method in smart grid," *IEEE Tran. on Smart Grid*, 2015.

Rehan Nawaz received the Ph.D. degree in Electrical Engineering in 2024, the M.S. degree in 2017, and the B.S. degree in 2015, all from the Department of Electrical and Computer Engineering, Air University, Islamabad, Pakistan. He is currently a Postdoctoral Fellow with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. He previously worked as a professional freelancer and as a Research Associate at Air University. He also participated in the Research Student Attachment Programme at The Hong Kong Polytechnic University. His research interests include data-driven and machine learning applications in power systems, with a focus on smart grid cybersecurity, malicious data injection and detection, power system optimization, and the reliability and resilience of cyber-physical systems.

Rabbaya Akhtar is currently pursuing a Ph.D. degree and serving as a Junior Researcher in the School of Energy Systems at Lappeenranta-Lahti University of Technology (LUT), Finland. She received her Bachelor's degree in Electrical Engineering from COMSATS University Islamabad, Pakistan, in 2017, and her Master's degree in Electrical Engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2022. She has previously worked as a Lab Engineer at Air University Islamabad, Pakistan, and as an Assistant Manager (Technical) at the National Transmission and Despatch Company (NTDC), Pakistan. Her research interests include smart grids and machine learning applications in energy systems.

Saad Ullah Khan received his B.Sc. degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2012, and Ph.D. degree in electronic and electrical engineering from Sungkyunkwan University, South Korea, in 2019, respectively. He is currently working as an Assistant Professor with the Department of Electrical and Computer Engineering, Air University, Islamabad, Pakistan. His research interests include power system operation, power system resilience, active distribution systems, energy management, and electric vehicles.

Siqi Bu (S'11-M'12-SM'17) He received the Ph.D. degree from the Electric Power and Energy Research Cluster, The Queen's University of Belfast, Belfast, U.K., where he continued his postdoctoral research before entering industry. He then joined National Grid UK as an experienced U.K. National Transmission System Planner and Operator. He is currently a Professor and Associate Head with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, Associate Director of the Research Centre for Grid Modernisation, and a Chartered Engineer with the U.K. Royal Engineering Council, London. His research interests include power system stability, operation and economics considering renewable energy integration, smart grid applications, and transport electrification.

Dr. Bu is an Editor of IEEE Transactions on Power Systems, IEEE Transactions on Consumer Electronics, IEEE Power Engineering Letters, IEEE Open Access Journal of Power and Energy, CSEE Journal of Power and Energy Systems, Protection and Control of Modern Power Systems, Journal of Modern Power Systems and Clean Energy, and Advances in Applied Energy. He is a Fellow of the IET, Chairman of the IET Hong Kong Power and Energy Section, Co-Chairman of IET DPSP 2025 and APSCOM 2025, and Technical Chairman of IEEE PESIM 2026.

Muhammad Habib Mahmood received the B.E. degree in Mechatronics Engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2003. He obtained his Master's degree in Computer Vision and Robotics jointly from the University of Burgundy, France; the University of Girona, Spain; and Heriot-Watt University, UK, in 2010, graduating with distinction. He earned his Ph.D. in Computer Vision from the University of Girona, Spain, in 2018. He is currently serving as a Principal Data Scientist at Teradata Corp, where he leads the development and deployment of advanced artificial intelligence solutions. His reserach interests include applied AI in industry, intelligent systems, computer vision, and the strategic development of AI services.