Locational Detection of False Data Injection Attacks in Smart Grids: A Graph Convolutional Attention Network Approach

Wei Xia, Deming He, and Lisha Yu

***Abstract*—As a typical application supported by the Internet of Thing (IoT), smart grids, which are critical complex cyber–physical infrastructures, are facing increasing cybersecurity threats. One major substantial cybersecurity threat to smart grids is false data injection attacks (FDIAs), which could bypass bad data detectors so as to disrupt the operation of power grids. In this work, we consider the locational detection problem for FDIAs, namely, detecting the presence of FDIAs and locating the compromised buses, by utilizing smart grid data (either power injection measurements or system state estimates). Regarding smart grid data residing on the inherent underlying graph structures of power grids, we model smart grid data as non-Euclidean graph signals. We correspondingly develop the FDIA detector based on the graph convolutional attention network (GCAT) for the locational detection of FDIAs, by considering the underlying graph topology of a power grid. The proposed FDIA detector leverages the graph attention mechanism, which could elastically assign the graph shift operators in the GCAT, to enhance the locational detection performance. Integration of the proposed FDIA detector in a noninvasive way could endow existing power systems with the capability of locational detection of FDIAs. Illustrative simulation results demonstrate the superior locational detection performance of the proposed GCAT-based FDIA detector, compared to the other state-of-the-art FDIA detectors.**

***Index Terms*—False data injection attack (FDIA), graph convolutional attention network (GCAT), power system security, smart grid.**

I. Introduction

SMART grids which integrate power grids with advanced information and communication technology (ICT) to enhance the grid efficiency and reliability, have been considered as one of the most important applications of Internet of Thing (IoT) technologies and can be perceived as one giant smart grid IoT network [1], [2], [3]. Operators of smart grids generally monitor real-time system states (such as the voltage phasor on each bus) of power grids so as to make appropriate decisions for the security and stability of power systems [4], [5]. System states of smart grids could generally be estimated [5] with the physical measurements, such as active/reactive power flows on branches and active/reactive power injections on buses from remote terminal units (RTUs), where the summation of power flow values on branches connected to a bus are referred to as the power injection value on the corresponding bus. The physical measurements from RTUs would be delivered to supervisory control and data acquisition (SCADA) systems [4] in order to estimate system states. Besides the RTUs, physical measurements can also be acquired by another sort of smart metering devices, namely, phasor measurement units (PMUs), which would produce estimates of voltage phasor at each bus [6]. Specifically, the system states can be estimated based on either the alternating current (AC) or direct current (DC) power flow model, both of which could describe the operational characteristics of each power grid component so as to facilitate the analysis of power grids [5]. The state estimation based on the AC power flow model is generally computationally expensive and might not converge to an optimal solution in certain scenarios [7]. Whereas the state estimation based on the DC power flow model is generally endowed with merits, such as higher reliability, simpler control, and more efficient interfacing with the renewable energy sources and energy storage units [8]. Moreover, the DC power flow model could also be utilized to approximate the AC power flow model [5], [7], [8].

The estimates of the system states of smart grids might be corrupted owing to the presence of bad data, such as large unexpected meter or communication errors. Fortunately, bad data can be detected by a bad data detector (BDD) [7], [9], [10]. However, besides bad data, another major and typical type of cybersecurity threat, namely, false data injection attacks (FDIAs) could bypass BDDs, and disturb the power system state estimation by injecting false data into the physical measurements [10], [11], [12], [13], [14]. Unware of the presence of FDIAs, power grid operators would be deceived by the corrupted system state estimates, and take actions rendering the operations of a power grid in an potentially unsafe mode, or even resulting in catastrophic consequences, such as overloading or destruction of grid devices.

Manuscript received 30 June 2023; revised 18 August 2023; accepted 28 September 2023. Date of publication 10 October 2023; date of current version 7 March 2024. This work was supported by the National Natural Science Foundation of China under Grant 61871104. *(Corresponding author: Wei Xia.)*

Wei Xia is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the School of Computer Science and Technology (School of Cyberspace Security), Xinjiang University, Ürümqi 830046, China (e-mail: wx@uestc.edu.cn).

Deming He is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: demingh@std.uestc.edu.cn).

Lisha Yu was with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China. She is now with the Protocol Product Center, China Satellite Network Exploration Company Ltd., Chongqing 401120, China (e-mail: lishayu_1998@163.com).

Digital Object Identifier 10.1109/JIOT.2023.3323565

Recently, investigation of the techniques of detecting the presence of FDIAs on smart grids has been attracting increasingly interest [12], [13], [15], [16], [17], [18], [19]. FDIA detection can be generally categorized into data-driven [12], [15], [16], [19] and model-based approaches [13], [17], [18]. The state-of-the-art deep-learning-based mechanism could be utilized to design (model-free) data-driven FDIA detectors [15], [16], relying only on historical data sets of systems. Specifically, based on the AC power flow model, the multilayer perceptron (MLP)-based FDIA detector has been demonstrated to be of high-detection accuracy, with five features selected by the principal component analysis of the measurement residuals [16]. On the other hand, based on the DC power flow model, the conditional deep belief network-based FDIA detector has been demonstrated to successfully detect over 90% of FDIAs [15], with the historical measurements of power flows utilized. By using the Kalman filter and recurrent neural network (KFRNN), Wang et al. [19] presented an effective two-level learner-based FDIA detection scheme based on the DC power flow model.

Both the aforementioned FDIA detectors [15], [16], [19] could detect the presence of FDIAs, but are incapable of simultaneously locating the attacked buses (meters) of a power system. Nevertheless, locating the attacked buses is generally essential, such that preventive actions, such as isolating the buses under attack or redispatching the power system, could be taken. Based on either the AC or DC power flow model, locational detection of FDIAs can be formulated as a multilabel classification problem, where each instance is associated with a set of labels [20] dictating which buses (meters) are under attack. Regarding the conventional multilabel learning algorithms, the multilabel classification problem can be generally addressed in two approaches [20]. On the one hand, we could convert the multilabel classification problem to another tractable classification problem leveraging the methods, such as the binary relevance (BR) [21], classifier chain (CC) [22], or label powerset (LP) [23], and apply the conventional single-label classification algorithms, such as the Gaussian Naive Bayes (GNB) [24], support vector machine (SVM) [25], [26], etc. On the other hand, we could solve the multilabel classification problem leveraging the modified single-label classification algorithms, such as the multilabel K-nearest neighbor (KNN) [27] and the decision tree (DT) [28]. Nevertheless, the aforementioned conventional multilabel learning algorithms might not perform quite satisfactorily the locational detection of FDIAs in smart grids, as experimentally verified hereinafter.

Recently, the FDIA detector [8] leveraging the convolutional neural network (CNN) has been demonstrated to outperform the aforementioned conditional deep belief network-based FDIA detector [15], based on the DC power flow model. The CNN-based FDIA detector [8] is a *noninvasive* plug-in, as the entire architecture incorporating the locational detector therein is established without altering or invading the existing power systems.

Different from the data-driven detection approaches, the model-based FDIA detectors [17], [18], [29], [30], [31], [32] generally depend on the knowledge of power systems (e.g., the near chordal sparsity of power grids [32]), yet without the need for historical data sets [33]. Recently, the irregular topology of a power system has been modeled as a *graph* in the model-based approaches [17], [18], with each grid bus and branch intuitively abstracted as a vertex (node) and an edge, respectively. The neighborhood of each node is the set of neighboring nodes, which includes the node itself and the nodes connected to it via an edge. Thereby, either the system states or the physical measurements from power grids could be modeled as graph signals in the non-Euclidean domain [34], and could be well manipulated with the graph signal processing (GSP) methodology, including graph filtering, graph Fourier transform, etc., extended from the classical digital signal processing tools.

Integration of the emerging GSP methodology in power systems, such as the GSP framework for power grids based on PMU data, the spectral graph analysis of power flows, FDIA detection, and so on, has been successfully validated in recent works [17], [18], [35], [36]. Specifically, with the graph signal filtered by the *spectral* graph filter, based on either the AC [18] or DC power flow model [17], the model-based FDIA detectors exploiting the inherent underlying graph structure of the power grid have been experimentally demonstrated to be effective in detecting FDIAs [17], [18].

It is promising to enhance the detection performance by exploiting the underlying graph structure of power grids, which, however, is generally overlooked by the aforementioned data-driven methods [8], [15], [16], [37]. On the other hand, although the model-based FDIA detectors [17], [18] cope with the underlying graph structure of power grids leveraging the spectral graph filters, they are generally incapable of locating the attacked buses (meters). Furthermore, the spectral graph filter-based FDIA detectors therein might not be readily transplanted into a different power system in another scenario, as the graph filters therein are generally customized along with potentially requisite manual tuning of the threshold for a certain system in a certain scenario.

By incorporating the underlying graph structure, the graph neural networks (GNNs) extended from the CNN methodology have recently been demonstrated to efficiently model the non-Euclidean data via nodal feature propagation and aggregation [34], [38]. Akin to the graph filters that could be categorized into spectral graph filters and spatial filters [39], the GNNs can be classified into spectral and spatial approaches as well [34].

Among the spectral approaches, the graph convolutional network (GCN) [40] has been recently applied to develop the locational detector for FDIAs in smart grids [41]. The convolutions in the Laplacian domain therein are approximated with Chebyshev polynomials of the preset Laplacian matrix that is determined by the magnitude of the admittance matrix of the corresponding grid. With the power injection measurement on each bus as the input feature, the GCN-based FDIA detector could achieve enhanced locational detection performance [41]. Appropriate Laplacian matrix in GNNs for locating FDIAs in smart grids is generally unknown. Hence, presetting the appropriate weights of the GSO is generally challengeable.

Alternatively, with spatial approaches operating on the neighboring nodes [34], convolutions could be defined directly on the graph instead of the Laplacian domain [40]. Among the spatial approaches, the celebrated graph convolutional attention network (GCAT) [38], [42] incorporates the graph attention mechanism into the nodal propagation to adapt the graph shift operator utilized in the spatial graph filter. Unlike the GCN using the fixed Laplacian matrix which could also be adopted as the graph shift operator [41], the GCAT leverages the graph attention mechanism such that the graph shift operator in GCAT could be assigned elastically with different weights.

In practice, a device attack (e.g., a compromised intelligent electric device, such as a circuit breaker breaks a circuit maliciously) [43] might take place in smart grids, or exteme natural factors, such as strong winds or heavy snow might result in transmission line failures or power outage [44]. Besides, new buses might join an existing power system. Under either of the above circumstances, the power grid topology would change. Consequently, the topology of a power grid in the testing phase may be not necessarily exactly the same as that utilized in the training phase for data-driven FDIA detectors. However, the aforementioned data-driven FDIA detectors, such as the MLP-based FDIA detector [16], CNN-based FDIA detector [8], and GNN-based FDIA detector [41], could not be applied directly to such scenarios. In contrast, owing to the graph attention mechanism, the dimensions of the trainable parameters are irrelevant to the number of nodes. Thus, the GCAT could be directly applied to graphs (power grid topologies) different from those in the training phase [42]. As a matter of fact, to the best of the authors' knowledge, there is a lack of investigations on the FDIA locational detectors based on not only the spatial GNN but also the GCAT approaches, despite their potential practical importance.

In this article, we consider filling the above gap. We consider herein the locational detection of FDIAs in smart grids by utilizing either power injection measurements or system state estimates. We first formulate the locational detection architecture incorporated with the noninvasive FDIA detector. Exploiting the underlying graph topology structure of the power grids, we develop the FDIA detector based on the spatial GNNs, which generates high-dimensional features via feature propagation and aggregation. To adapt the graph shift operator utilized in the GNN, we further reinforce the GCAT-based FDIA detector leveraging the graph attention mechanism. Illustrative simulation results validate the superior locational detection performance of the proposed GCAT-based FDIA detector. Moreover, the proposed GCAT-based FDIA detector could be applied to scenarios where the power grid topology in the testing phase is different from that in the training phase, as experimentally validated.

The main contributions of this work are summarized as follows.

1) We propose the architecture for FDIA detection, based on the spatial GNN architecture, such that high-dimensional features could be extracted from the non-Euclidean graph data residing on the underlying graph structure of a power system.
2) We develop the GCAT-based FDIA detector leveraging the graph attention mechanism, which could adapt the graph shift operator in the GNN. We further experimentally validate the performance of the proposed detector when the power grid topology in the testing phase is different from that in the training phase.
3) Regarding either power injection measurements or system state estimates, we propose the locational detection architecture for FDIAs. The proposed architecture could readily incorporate FDIA detectors with the existing power systems, including BDD in a noninvasive way.

Notation: The set of real numbers is denoted by $\mathbb{R}$. Scalars, vectors and matrices are denoted by lower case letters (a), lower case boldface letters ($\boldsymbol{a}$) and upper case boldface letters ($\boldsymbol{A}$), respectively. The ith entry of the vector $\boldsymbol{a}$ is denoted by $[\boldsymbol{a}]_i$, the (i,j)th entry of the matrix $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,j}$, and the ith row of $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,:}$. The transpose and the l_2-norm of the matrix or vector are denoted by $(\cdot)^{\mathrm{T}}$ and $\|\cdot\|$, respectively. The Kronecker product is denoted by the symbol $\otimes$. $(\cdot)^{-1}$ denotes the inverse of the matrix. The $M \times M$ identity matrix is denoted by $\boldsymbol{I}_M$.

II. Problem Formulation

A. Graph Representation of Smart Grids

We consider herein an N-bus power system consisting of M meters. Without loss of generality, we assume that $M \leq N$. The power system is modeled as a connected, undirected, weighted graph $\mathcal{G} \triangleq (\mathcal{V}, \mathcal{E})$ [17], [18], [41]. The vertex (node) set $\mathcal{V} = \{1, \ldots, N\}$ consists of the buses with connected loads or generators of the power system, and the cardinality of the vertex set is $|\mathcal{V}| = N$. The set of branches and transformers connecting bus $i \in \mathcal{V}$ with bus $j \in \mathcal{V}$ is denoted as $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, composed of $|\mathcal{E}| = E$ edges. The neighboring bus j of bus i indicates that bus j is connected to bus i. The neighborhood of bus i, denoted by $\mathcal{N}_i$, is defined as the set of neighboring buses to bus i, including bus i itself.

A graph signal supported on the vertex set of the graph can be represented as $\boldsymbol{u} \triangleq [u_1, u_2, \ldots, u_N]^{\mathrm{T}} \in \mathbb{R}^N$, where each entry u_i is associated with bus i, $i = 1, 2, \ldots, N$. Thus, either system state estimates or power injection measurements of the smart grids could be modeled as a non-Euclidean graph signal. Moreover, the graph shift operator $\boldsymbol{\Phi} \in \mathbb{R}^{N \times N}$ of the graph $\mathcal{G}$ replaces u_i (either the power injection measurement or the state estimate) at each node (bus) i with the linear transformation of the values in the neighborhood $\mathcal{N}_i$. Each entry $[\boldsymbol{\Phi}]_{i,j}$ of $\boldsymbol{\Phi}$ is nonzero if bus i and j are connected, or $i = j$; and $[\boldsymbol{\Phi}]_{i,j} = 0$, otherwise [45], [46]. Specifically, the one-hop shift of $\boldsymbol{u}$ is given by $\boldsymbol{\Phi u}$. The graph shift operator $\boldsymbol{\Phi}$ is generally sparse [39] accounting for the local graph structure of power grids, and thus could capture the sparsity and locality of the relationship among different entries of the graph signal $\boldsymbol{u}$ (either the power injection measurements or the state estimates) [38].

Graph signals could be processed by graph filters that take a graph signal as the input and produce another graph signal as the output, with the linear graph signal operator $\boldsymbol{G} : \mathbb{R}^N \rightarrow \mathbb{R}^N$, which is defined as a polynomial of the graph shift operator $\boldsymbol{\Phi}$ [38], [39]. The (convolutional) graph filter with the order

K can be given by

$$\boldsymbol{G} \triangleq \sum_{k=0}^{K} \boldsymbol{\Phi}^{(k:0)} \tag{1}$$

with $\boldsymbol{\Phi}^{(k:0)} \triangleq \prod_{k'=0}^{k} \boldsymbol{\Phi}_{k'} = \boldsymbol{\Phi}_k \boldsymbol{\Phi}_{k-1}, \ldots \boldsymbol{\Phi}_0$, where $\boldsymbol{\Phi}_0 \in \mathbb{R}^{N \times N}$ is a diagonal matrix, and each matrix $\boldsymbol{\Phi}_{k'}$ is the graph shift operator, $k' = 1, \ldots, K$. The output of the graph filter $\boldsymbol{G}$ can thus be regarded as a weighted linear combination of consecutively shifted graph signals, namely

$$\boldsymbol{o} = \boldsymbol{G}\boldsymbol{u} = \sum_{k=0}^{K} \boldsymbol{\Phi}^{(k:0)} \boldsymbol{u}. \tag{2}$$

With the graph filter $\boldsymbol{G}$ operating on the neighboring nodes, the value (either the power injection measurement or the state estimate) at each node is exchanged with its neighboring nodes via graph convolution (2). Specifically, the output of a Kth order graph filter is a sum of the aggregation of either the power injection measurement or the state estimate at each bus i coming from its k-hop neighborhood, $k = 0, \ldots, K$. Thus, the graph convolution (2) is K-localized, whose computational complexity is $\mathcal{O}(K(E+N))$. Obviously, data at each node can be propagated across the graph in a diffusion manner [38].

B. Power System State Estimation

Accurate system state estimation is crucial to the cybersecurity and stability of power systems. For simplicity, we consider herein the DC power flow model [8], [15], [30]. In the DC power flow model, the physical measurement $\boldsymbol{z}_o = [z_1, z_2, \ldots, z_M]^{\mathrm{T}} \in \mathbb{R}^M$ consists of the active power flows on the branches and active power injections on the buses from meters,[1] and can be expressed as [4], [5], [7], [8]

$$\boldsymbol{z}_o = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} \tag{3}$$

where the Jacobian matrix $\boldsymbol{H} \in \mathbb{R}^{M \times N}$ is associated with the topology and line impedances of the power system [17], each entry of the system state $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}} \in \mathbb{R}^N$ corresponds to the voltage phase angle of each bus, and $\boldsymbol{e} \in \mathbb{R}^M$ is the zero-mean additive white Gaussian noise (AWGN), with the covariance matrix $\boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I}_M$.

The system state of a power grid can be estimated via the conventional weighted least square (WLS) [5]

$$\hat{\boldsymbol{x}}_o = \left(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_o. \tag{4}$$

The system state estimation may be corrupted by bad data owing to various reasons, such as large unexpected meter errors or communication errors [10]. Nevertheless, bad data could be detected with conventional BDDs [10] through comparing the l_2-norm of the measurement residual

$$r = \|\boldsymbol{z}_o - \boldsymbol{H}\hat{\boldsymbol{x}}_o\|^2 \gtrless \tau_r \tag{5}$$

with the given threshold τ_r. The hypothesis that bad data exists in the physical measurements would be accepted if $r > \tau_r$. We could leverage the BDDs to ameliorate the reliability of state estimation in the presence of bad data [10].

C. False Data Injection Attacks

In addition to bad data, attackers could also launch FDIAs by injecting false data into the physical measurements to undermine the system state estimation. Thus, the physical measurement can be rewritten in a unified form

$$\boldsymbol{z} = \begin{cases} \boldsymbol{z}_o + \boldsymbol{a}, & \text{in the presence of FDIAs} \\ \boldsymbol{z}_o, & \text{otherwise} \end{cases} \tag{6}$$

where $\boldsymbol{a}$ denotes the attack vector.

Constructing the attack vector as $\boldsymbol{a} \triangleq \boldsymbol{H}\boldsymbol{c} \in \mathbb{R}^M$ rather than an arbitrary vector would probably mislead power grid operators into accepting a compromised system state estimate as anticipated by the attackers [10], [31], where $\boldsymbol{c} \in \mathbb{R}^N$ denotes the error vector induced by FDIAs. Accordingly, the system state estimate can also be generally rewritten in the following unified form:

$$\hat{\boldsymbol{x}} = \begin{cases} \hat{\boldsymbol{x}}_o + \boldsymbol{c}, & \text{in the presence of FDIAs} \\ \hat{\boldsymbol{x}}_o, & \text{otherwise.} \end{cases} \tag{7}$$

The error vector $\boldsymbol{c}$ is generally sparse, as not every bus would be attacked, due to the limitations on either the attackers' resources required to attack all the buses of a power grid or the physical constraint(s) in some specific meters [29], [31]. We use herein $\mathcal{V}_a \subset \mathcal{V}$ to denote the vertex (bus) set of the attacked buses. If bus i is attacked, the corresponding ith element $[\boldsymbol{c}]_i$ is nonzero.

Obviously, in light of (5), the measurement residual r would be unchanged even in the presence of FDIAs given $\boldsymbol{a} = \boldsymbol{H}\boldsymbol{c}$, since $r = \|\boldsymbol{z} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 = \|\boldsymbol{z}_o + \boldsymbol{a} - \boldsymbol{H}(\hat{\boldsymbol{x}}_o + \boldsymbol{c})\|^2 = \|\boldsymbol{z}_o - \boldsymbol{H}\hat{\boldsymbol{x}}_o\|^2$. It is also noteworthy that given $\boldsymbol{a} = \boldsymbol{H}\boldsymbol{c}$, $(\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{I})$ $\boldsymbol{a} = \boldsymbol{0}$, which is in accordance with the condition in [10, eq. (6)]. Moreover, for $|\mathcal{V}_a| > 0$, in light of [10, Th. 3.3], under the assumption that $M \leq N$, attackers could generate an attack vector $\boldsymbol{a}$ that could bypass the BDDs [10], [31].

In this work, we consider developing an approach to not only detecting false data injection attacks but also locating the attacked buses, utilizing either the physical measurements or the system state estimates. Regarding the non-Euclidean graph signals from smart grids, we consider developing FDIA detection techniques based on GNNs. We further develop an FDIA detector based on the GCAT incorporating the graph attention mechanism, so as to determine the unknown parameters of the graph shift operator in the GNN via learning.

III. Locational Detection Approach of FDIAs

In view of the underlying topologies of power grids, we now consider developing a noninvasive FDIA detector, which could be readily incorporated into existing power systems, by leveraging the GNN to capture the spatial proximity among the features on neighboring buses. Further, we consider exploiting the graph attention mechanism to adapt the graph shift operator by utilizing the nodal features.

A. Locational Detection Architecture

We consider herein a noninvasive locational detector for FDIAs, i.e., integration of the detector into an existing system would necessitate no alternation to the existing power system,

[1] In the DC power flow model, system states are usually voltage phase angles [10], as voltage magnitudes and reactive power flows are of little concern.

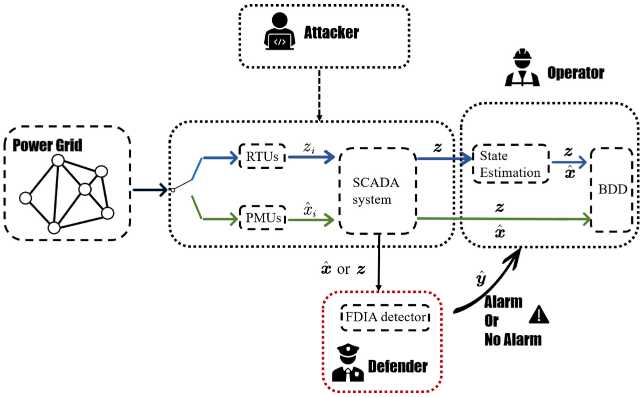


Fig. 1. Illustration of the locational detection architecture.

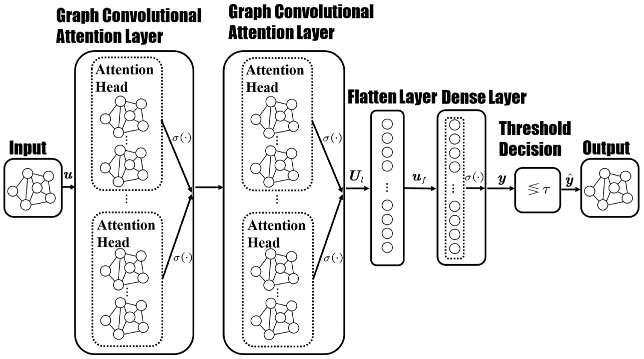


Fig. 2. Architecture of the proposed GCAT-based detector.

as illustrated in Fig. 1. Besides, integration of a BDD could ameliorate the accuracy of the system state estimates in the presence of bad data.

Each physical measurement $\boldsymbol{z}$, such as the active power flows on branches and active power injections on buses collected by RTUs [4], is delivered to SCADA systems for further processing and analysis [4]. In the presence of FDIAs, operators would obtain the physical measurements contaminated with false data [10], [11]. Given the physical measurement vector $\boldsymbol{z}$ from the SCADA system, the system state estimate $\hat{\boldsymbol{x}}$ can be obtained [5] within the state estimation module, as illustrated in Fig. 1. Moreover, those estimates of voltage phase angle can be obtained either from the state estimation module, or directly from the PMUs [6], [35].

In this work, we consider developing a noninvasive FDIA detector that could not only detect the presence of FDIAs but also locate the attacked buses. Each bus of the power system would be regarded as a node in the graph $\mathcal{G}$, and labeled as $i = 1, 2, \ldots, N$. With either the physical measurements (power injections) $\boldsymbol{z}$ or the system state (voltage phase angle) estimates $\hat{\boldsymbol{x}}$ of all buses utilized, each entry $\hat{y}_i$ of the output $\hat{\boldsymbol{y}} \in \mathbb{R}^N$ of the detector based on the GCAT equals either 1 or 0, correspondingly indicating that bus $i \in \mathcal{V}$ is attacked or not. Once the FDIAs are detected, alarms would be triggered to notify power grid operators of which buses have been under attack, such that operators could take preventive actions, such as isolating the buses under attack or redispatching the power system.

B. Architecture of the GCAT-Based Detector

The architecture of the proposed FDIA detector, which takes into account the non-Euclidean graph signals of smart grids, is illustrated in Fig. 2. Incorporation of the graph attention mechanism into the graph convolutional layer of the spatial GNN facilitates the adaptive learning of the unknown parameters of the graph shift operator(s). Hence, the resulting detector is dubbed GCAT-based FDIA detector hereinafter.

The FDIA detector consists of an input layer, two hidden graph convolutional attention layers indexed by l $(l = 1, 2)$, one flatten layer, one dense (fully connected) layer, and an output layer. The input graph signal $\boldsymbol{u} = [u_1, u_2, \ldots, u_N]^{\mathrm{T}}$ of the smart grids corresponds to the input layer of the GCAT-based FDIA detector. Whereas each entry of $\boldsymbol{u}$ is given by either the physical measurement or the system state estimate at each bus (i.e., $\boldsymbol{u} = \boldsymbol{z}$ if $M = N$, or $\boldsymbol{u} = \hat{\boldsymbol{x}}$). High-dimensional features could be extracted from the input graph signal $\boldsymbol{u}$ by the two hidden graph convolutional attention layers such that the attacked buses could be located. Moreover, each entry y_i of the output $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^{\mathrm{T}} \in \mathbb{R}^N$ of the dense layer is the probability of the corresponding bus i being attacked. The binary decision with respect to $\boldsymbol{y}$ yields $\hat{\boldsymbol{y}} \in \mathbb{R}^N$, each entry $\hat{y}_i$ of which indicates that bus i is attacked or not.

Specifically, each hidden graph conventional attention layer contains H independent attention mechanisms, as illustrated in Fig. 2. The output features $\boldsymbol{U}_{l-1} \in \mathbb{R}^{N \times F_{l-1}}$ of layer $l - 1$ are the input to the hidden graph convolutional attention layer l, with F_{l-1} being the number of features at layer $l - 1$. Note that each column of $\boldsymbol{U}_{l-1}$ is a graph signal. Whereas each ith row $[\boldsymbol{U}_{l-1}]_{i,:}$ of $\boldsymbol{U}_{l-1}$ is the nodal features of node $i \in \mathcal{V}$. The output features, $\boldsymbol{U}_l \in \mathbb{R}^{N \times F_l}$, are obtained via either concatenating or averaging the output of H independent graph convolutions (8) of each hidden graph convolutional attention layer l, based on the multihead (H-head) attention mechanism, as detailed hereinafter. Especially, the input of the hidden graph convolutional attention layer $l = 1$ is the input graph signal $\boldsymbol{u}$ with $F_0 = 1$.

The implementations of the proposed GCAT-based FDIA detector are detailed below.

C. Graph Convolutional Attention Layer

Graph convolutions in GNNs for non-Euclidean data of smart grids could be implemented via graph filtering [38]. In light of (1), the output of the graph filter $\boldsymbol{G}$ is a weighted linear combination of the shifted neighbors' nodal features, such that the spatial proximity between the features on neighboring buses could be extracted from the graph signal $\boldsymbol{u}$.

As illustrated in Fig. 3, the graph filter $\boldsymbol{G}$ incorporated at each graph convolutional attention layer l would filter each column of the input feature $\boldsymbol{U}_{l-1} \in \mathbb{R}^{N \times F_{l-1}}$, which is shifted by the graph shift operators, and weighted by the trainable parameter matrix $\boldsymbol{W}_{lk} \in \mathbb{R}^{F_{l-1} \times F_l}$ [38]. Thus, the dimension of the features at each node is converted from F_{l-1} to F_l, such that the high-dimensional features could be extracted.

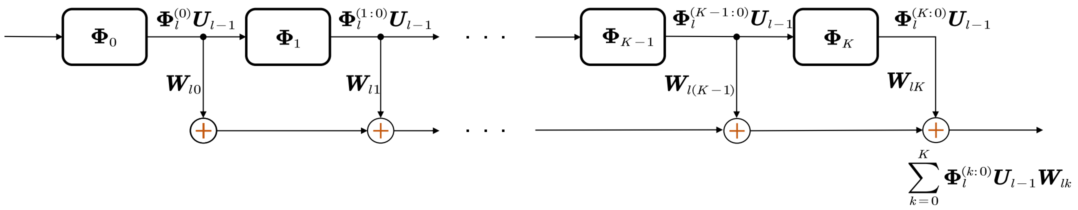


Fig. 3. Graph filter in the graph convolutional attention layer.

With the nonlinear activation function $\sigma(\cdot)$ applied elementwise, the output $\boldsymbol{U}_l \in \mathbb{R}^{N\times F_l}$ of the graph convolution at the graph convolutional attention layer l is given by

$$\boldsymbol{U}_l = \sigma\left(\sum_{k=0}^{K}\boldsymbol{\Phi}_l^{(k:0)}\boldsymbol{U}_{l-1}\boldsymbol{W}_{lk} + \mathbf{1}\otimes\boldsymbol{b}_l^{\mathrm{T}}\right) \tag{8}$$

where $\boldsymbol{b}_l \in \mathbb{R}^{F_l}$ represents the bias term of layer l [38], with $\mathbf{1}$ denoting the all-ones vector. We select herein the PReLU nonlinearity [47]

$$\mathrm{PReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases} \tag{9}$$

as the nonlinear activation function $\sigma(\cdot)$ with the trainable negative input slope parameter α initialized as 0.2. For simplicity, we apply the same graph shift operator $\boldsymbol{\Phi}_l$ in the graph filter, to shift the input feature at each graph convolutional attention layer l. Thus, $\boldsymbol{\Phi}_l^{(k:0)}$ is given by the powers of the graph shift operator, i.e., $\boldsymbol{\Phi}_l^{(k:0)} = \boldsymbol{\Phi}_l^k$. As indicated by (8), with the graph shift operator $\boldsymbol{\Phi}$, the nodal feature $[\boldsymbol{U}_{l-1}]_{i,:}$ at each node i is propagated across the graph $\mathcal{G}$, and the output nodal feature $[\boldsymbol{U}_l]_{i,:}$ at each node i is an aggregation of the neighboring nodal features, such that the spatial proximity between the features on the neighboring buses could be extracted from the data residing on the underlying graph. Moreover, for graph filters of different orders k, different nodal features could be rationally tackled with the elastically assigned parameter matrices $\boldsymbol{W}_{lk}$, $k = 0, 1, \ldots, K$.

It is noteworthy that with the preset graph shift operator $\boldsymbol{\Phi}_l$ (e.g., select the unweighted adjacent matrix or the admittance matrix of the corresponding grid for $\boldsymbol{\Phi}_l$), we could develop FDIA detectors based on the architecture of spatial Graph Neural Networks.

However, practically, the appropriate graph shift operator $\boldsymbol{\Phi}_l$ is generally unknown, and its weights cannot be suitably preset, either [38]. Alternatively, we consider herein leveraging the graph attention mechanism [42] to obtain the unknown graph shift operator $\boldsymbol{\Phi}_l$. First, given the attention mechanism $\mathcal{A}$, the attention coefficient of node i is determined by importance of node $j \in \mathcal{N}_i$ with respect to node i, and is formally given by

$$e_{l,i,j} = \mathcal{A}\Big(\big[\boldsymbol{U}_{l-1}\boldsymbol{B}_l\big]_{i,:}, \big[\boldsymbol{U}_{l-1}\boldsymbol{B}_l\big]_{j,:}\Big) \tag{10}$$

with the trainable parameter matrix $\boldsymbol{B}_l \in \mathbb{R}^{F_{l-1}\times F_l}$. With the attention mechanism $\mathcal{A}$ composed of a single-layer feedforward neural network [38], [42], the attention coefficient $e_{l,i,j}$ can be given by

$$e_{l,i,j} = \sigma\left(\boldsymbol{m}_l^{\mathrm{T}}\Big[\big[\boldsymbol{U}_{l-1}\boldsymbol{B}_l\big]_{i,:}||\big[\boldsymbol{U}_{l-1}\boldsymbol{B}_l\big]_{j,:}\Big]^{\mathrm{T}}\right) \tag{11}$$

with the trainable parameter vector $\boldsymbol{m}_l \in \mathbb{R}^{2F_l}$. We select the PReLU nonlinearity as the nonlinear activation function $\sigma(\cdot)$. With the concatenation operator $||$, the features at each node i are concatenated with those at node j. Note that the attention coefficient $e_{l,i,j}$ represents the importance of the features of node j to node i, and is relevant only to the features at both node i and its neighboring node j. Specifically, $e_{l,i,j}$ would be assigned with a bigger value if node j is more important to node i. Correspondingly, the features at node j would impact more notably on the filtered features at node i. We could obtain the attention coefficient through optimizing the loss function (14) below so as to enhance the locational detection performance, as detailed hereinafter.

With the softmax function, each entry $[\boldsymbol{\Phi}_l]_{i,j}$ (corresponding to the edge connecting bus j with bus i) of the graph shift operator can be normalized by

$$[\boldsymbol{\Phi}_l]_{i,j} = \mathrm{softmax}\big(e_{l,i,j}\big) = \frac{\exp\big(e_{l,i,j}\big)}{\sum_{p\in\mathcal{N}_i}\exp\big(e_{l,i,p}\big)} \tag{12}$$

for connected buses i and j, or $i = j$; and $[\boldsymbol{\Phi}]_{i,j} = 0$, otherwise. Take for example, the procedure of the attention mechanism processing is illustrated in Fig. 4. Each (i, j)th entry of the graph shift operator is the most relevant to the nodal feature $[\boldsymbol{U}_{l-1}]_{i,:}$ of node (bus) i and $[\boldsymbol{U}_{l-1}]_{j,:}$ of node (bus) j, and would be assigned with a larger value if node j is more important to node i [38], [42], such that the locational detection performance would be enhanced. In light of (8), the computational complexity of each graph convolutional attention layer l is given by $\mathcal{O}(F_{l-1}(NF_l + KNF_l + (E + N)))$. It is noteworthy that the dimensions of the trainable parameters, including $\boldsymbol{W}_{lk}, \boldsymbol{b}_l, \boldsymbol{m}_l$, and $\boldsymbol{B}_l$, are all irrelevant to the number of buses in the power system, owing to the graph attention mechanism. Hence, each graph convolutional attention layer l could be directly applied to power grids in the testing phase potentially distinct from those in the training phase.

We further employ the multihead attention mechanism [42] to improve the efficacy of the locational detection of our FDIA detector. As illustrated in Fig. 2, each hidden graph convolutional attention layer l is composed of H independent attention mechanisms (H heads). The output features of each graph convolutional attention layer l are obtained with the output features of H independent graph convolutions (8). The output features of each graph convolutional attention layer l is formally given by the concatenation of the output features

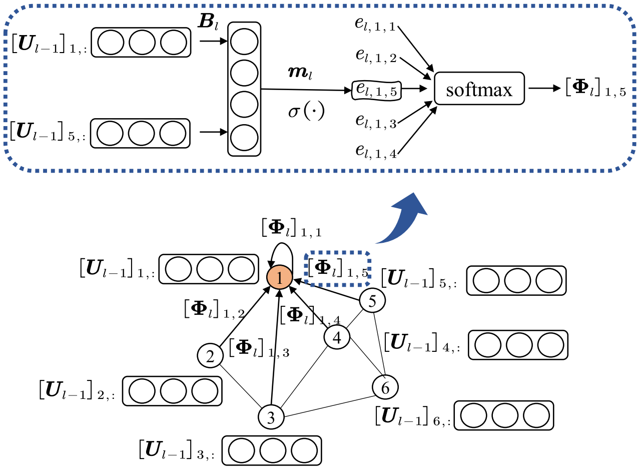

$[\boldsymbol{U}_{l-1}]_{1,:}$
$\boldsymbol{B}_l$
$[\boldsymbol{U}_{l-1}]_{5,:}$
$\boldsymbol{m}_l$
$\sigma(\cdot)$
$e_{l,1,1}$
$e_{l,1,2}$
$e_{l,1,5}$
$e_{l,1,3}$
$e_{l,1,4}$
softmax
$[\boldsymbol{\Phi}_l]_{1,5}$
$[\boldsymbol{\Phi}_l]_{1,1}$
$[\boldsymbol{U}_{l-1}]_{1,:}$
1
$[\boldsymbol{\Phi}_l]_{1,5}$
$[\boldsymbol{U}_{l-1}]_{5,:}$
$[\boldsymbol{\Phi}_l]_{1,2}$
$[\boldsymbol{\Phi}_l]_{1,4}$
5
4
$[\boldsymbol{U}_{l-1}]_{4,:}$
2
$[\boldsymbol{\Phi}_l]_{1,3}$
$[\boldsymbol{U}_{l-1}]_{2,:}$
6
3
$[\boldsymbol{U}_{l-1}]_{6,:}$
$[\boldsymbol{U}_{l-1}]_{3,:}$

Fig. 4. Graph attention mechanism for the graph convolutional attention layer l, with $F_{l-1} = 3$, $F_l = 4$. The graph topology consists of six nodes, each of which is indexed by the number at the node.

$\boldsymbol{U}_l^{(h)}$ of each independent dth graph convolution (8), namely, $\boldsymbol{U}_l = \|_{h=1}^{H}\boldsymbol{U}_l^{(h)}$. Especially, the output features of the last graph convolutional attention layer is obtained by averaging the output features of H independent graph convolution operation (8), i.e., $\boldsymbol{U}_l = (1/H)\sum_{h=1}^{H}\boldsymbol{U}_l^{(h)}$.

Remark 1: It is noteworthy that the NN architecture in the proposed GCAT-based detector is different from the conventional GCNs [40] and graph attention networks [42]. If $K = 1$ in (8) and $\boldsymbol{\Phi}$ is selected as the Laplacian matrix, the graph convolutional attention layer degenerates to that of the conventional GCNs [40]. On the other hand, if $K = 1$ and $\boldsymbol{B}_l = \boldsymbol{W}_{l,0} = \boldsymbol{W}_{l,1}$, the graph convolutional attention layer degenerates to that of graph attention networks [42].

D. Other Implementation Issues

We now brief the remainder of the implementation issues of the proposed GCAT-based FDIA detector, including both the flatten layer and dense (fully connected hidden) layer.

As illustrated in Fig. 2, the output features $\boldsymbol{U}_l \in \mathbb{R}^{N\times F_l}$ of the last graph convolutional attention layer is merged into one single vector $\boldsymbol{u}_f \in \mathbb{R}^{NF_l}$ at the flatten layer. With $\boldsymbol{u}_f$ fed into the dense layer, the output $\boldsymbol{y} \triangleq [y_1, y_2, \ldots, y_N]^{\mathrm{T}} \in \mathbb{R}^N$ can thus be given by

$$\boldsymbol{y} = \sigma\left(\boldsymbol{W}_D\boldsymbol{u}_f + \boldsymbol{b}_D\right) \tag{13}$$

where each entry y_i represents the probability of the corresponding bus i being attacked, and $\boldsymbol{W}_D \in \mathbb{R}^{N\times NF_l}$ and $\boldsymbol{b}_D \in \mathbb{R}^N$, respectively, denote the trainable parameter matrix and the bias of the dense layer [48]. We select $\mathrm{sigmoid}(x) = 1/(1+\exp(x))$ as the activation function $\sigma(\cdot)$ applied elementwise at the dense layer.

We consider herein optimizing all the free unknown parameters, including $\boldsymbol{W}_D, \boldsymbol{b}_D, \boldsymbol{W}_{lk}, \boldsymbol{b}_l, \boldsymbol{m}_l$, and $\boldsymbol{B}_l$ for each l and k, via minimizing the cross-entropy loss function

$$\mathcal{L}(\boldsymbol{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_{i,o}\log(y_i) + \left(1 - y_{i,o}\right)\log(1 - y_i)\right) \tag{14}$$

with the Adam optimizer [48]. The above function could measure the difference between the actual output (classified class probability) and the label, indicating whether each bus is actually attacked or not. Specifically, each entry $y_{i,o}$ of the label $\boldsymbol{y}_o \triangleq [y_{1,o}, y_{2,o}, \ldots, y_{N,o}]^{\mathrm{T}} \in \mathbb{R}^N$ equals either 1 or 0, correspondingly indicating that bus i is attacked or not.

With the aforementioned free parameters optimized, the output of the proposed GCAT-based detector is given by $\hat{\boldsymbol{y}} \in \mathbb{R}^N$, each entry $\hat{y}_i$ of which is determined by the binary decision

$$y_i \gtrless \tau \tag{15}$$

with the threshold $\tau \in (0, 1)$. The hypothesis that bus i is under attack would be accepted if $y_i > \tau$. Accordingly, $\hat{y}_i$ is set to be 1. Otherwise, $\hat{y}_i$ is set to be 0. Detectors would be generally more sensitive with a lower threshold τ. In contrast, as each output probability y_i of the proposed detector prior to the binary decision would be rather close to 1 or 0, the detector is generally less insensitive with respect to the threshold, such that it is capable of preferably discriminating between the presence and absence of FDIAs. Without loss of generality, we set τ to be 0.5 in the following practice unless otherwise specified. The superior discrimination (a.k.a. separability) capability [8] of the proposed detector is experimentally demonstrated by varying the threshold τ hereinafter. Obviously, $\|\hat{\boldsymbol{y}}\|_\infty = 1$ indicates the presence of FDIAs, with $\|\cdot\|_\infty$ denoting the infinite norm of the vector; otherwise, the power system is free of FDIAs. Correspondingly, in the former scenario, we could locate each attacked bus i, as its corresponding entry $\hat{y}_i$ equals 1.

IV. Case Study

We now evaluate the locational detection performance of the proposed GCAT-based detector, following the elaboration of the data set generation.

A. Data Set Generation

We first generate data sets by using Pandapower [49] for the test cases in the IEEE 14-bus power system [50], generally utilized to justify the performance of FDIA detectors [8], [17], [18]. Specifically, we generate the power grid data from the DC optimal power flow analysis implemented in Pandapower.

Considering a realistic power system operation, we herein fit the fluctuation of both the load and generated power values by utilizing the real-world load profiles [41]. We standardize the time series data of the length T from the "BUSHILF_SCENT" profile selected from the ERCOT's 15 min interval backcasted actual load profile [51] to be a vector $\boldsymbol{s} \in \mathbb{R}^{\mathrm{T}}$ of zero mean and unit variance. The state of the power grids would be impacted by either the load or generated power values scaled with the scaling parameter ς_i for each bus i in Pandapower. Each scaling parameter ς_i is assigned with a sample from the normal distribution $\mathcal{N}(1+q[\boldsymbol{s}]_t, \sigma_s^2)$ at each time instance $t = 1, \ldots, T$, where q represents the intensity of fluctuations of the load and the generated power values, and σ_s^2 represents the noise power of the load and the generated power values. Apparently, larger values of q correspond to more drastic changes of the load and generated power values. Whenas

$q = 0$, the load and generated power values of the buses fluctuate around the base load and the generated power values, as the data generation procedure in [8], [26]. We herein set $q = 0.1$, $\sigma_s = 0.03$, unless stated otherwise.

We next perform the DC optimal power flow analysis in Pandapower to generate data in the absence of attacks. We collect both the system state estimates $\hat{\boldsymbol{x}}^{(t)}$ and the physical measurements $\boldsymbol{z}^{(t)}$ of all buses at all time instances.

We now generate data in the presence of FDIAs, based on the data generated above. We select herein $M = N$ such that the FDIAs could bypass the BDDs [10]. Because real-world attackers would generally choose partial buses as attacked target, motivated by economic or other objectives, and then inject false data into physical measurements to mislead power grid operators into accepting a compromised system state estimate as anticipated by the attackers. Due to the lack of prior knowledge about the attack, for brevity, for each attacked time instance, we consider that the number n of the attacked buses[2] follows the discrete uniform $\mathcal{U}(n_{\min}, n_{\max})$ [8], with $n_{\min} = 2$ and $n_{\max} = 5$. Each attacked bus is also randomly selected with equal probability from the buses except for the slack bus [17]. The attacked bus set is denoted by $\mathcal{V}_a^{(t)}$ for each attacked time instance $t \in \mathcal{T}_a$, where each attacked time instance t from the attacked time instances set $\mathcal{T}_a$ is randomly selected with the equal probability from all-time instances. For brevity, the error $[\boldsymbol{c}^{(t)}]_i$ following the uniform distribution $\mathcal{U}(c_{\min}, c_{\max})$ of each attacked bus i for each attacked time instance $t \in \mathcal{T}_a$ can represent the aggression of FDIAs, where $c_{\min} = 0$ and $c_{\max} = 2$.

In the presence of FDIAs, we obtain the noisy system state data set $\mathcal{X}$ consisting of the system state estimates as well as the power injection data set $\mathcal{Z}$ consisting of the power injection measurements, on all the buses at all time instances. We set $\sigma_n = 0.01$ for all instants $t = 1, \ldots, T$. Unless stated otherwise, each data set contains 120 000 time instances, and is divided into the sets of training and testing, respectively, composed of 110 000 and 10 000 time instances. We set $f_a = 0.1$, the proportion of the attacked time instances in all-time instances, for the training set, and $f_a = 0.5$ for the testing set. The data set generation steps are summarized in Algorithm 1.

B. Implementation Issues and Performance Metrics

All the implementations are conducted in Python 3.7 using Pandapower [49], PyTorch [52], and sklearn [53] on Intel Core i7-11700 CPU with NVIDIA GeForce RTX 3080 Ti GPU. For benchmark methods, we compare our proposed method with the DT [28], GNB [24], SVM [25], Multilabel KNN [27], MLP [16], CNN [8], and GCN [41].

The DT segregates the attributes into classes based on their respective input features to predict the class labels [28]. The Multilabel KNN employs maximum a posteriori to determine the class labels, via the proximity calculation based on statistical information derived from the label sets [27]. Based on the assumption that all the features are independent of one another, and obeying the Gaussian distribution, the GNB classifier belongs to the family of simple probabilistic classifiers [24]. The SVM is a maximum margin classifier maximizing the margin and constructing a hyperplane in a high-dimensional space [25]. Both the GNB and SVM are appropriate for single-label multiclass problems. We herein solve our multilabel problem by implementing the GNB and the SVM based on the LP method, which converts the multilabel classification problem to the multiclass classification by exploiting the interlabel correlations [23].

Algorithm 1: Data Set Generation

```
Input: IEEE bus system, $\boldsymbol{s}$, $\mathcal{T}_a$, $\mathcal{V}_a^{(t)}$
Output: Dataset $\mathcal{X}$, $\mathcal{Z}$
Initialization: $q$, $\sigma_s$, $\sigma_n$, $n_{\min}$, $n_{\max}$ $c_{\min}$, $c_{\max}$, $\boldsymbol{c}^{(t)} = \boldsymbol{0}$
  for each instant $t$
$f_a$ for training set, $f_a$ for testing set
the number $T$ of the instances $T$
Base Data:
for each time instant $t$ till $T$ do
    for each bus $i$ do
        $\varsigma_i \sim \mathcal{N}(1 + q[\boldsymbol{s}]_t, \sigma_s^2)$.
    end
    Through Pandapower, $\boldsymbol{x}^{(t)}$, $\boldsymbol{z}^{(t)}$ are obtained by
      performing the DC power flow analysis.
end
Attack Implementation:
for each attacked time instant $t \in \mathcal{T}_a$ do
    for each attacked bus $i \in \mathcal{V}_a^{(t)}$ do
        $[\boldsymbol{c}^{(t)}]_i \sim \mathcal{U}(c_{\min}, c_{\max})$.
    end
    $\hat{\boldsymbol{x}}^{(t)} \leftarrow \hat{\boldsymbol{x}}^{(t)} = \hat{\boldsymbol{x}}_o^{(t)} + \boldsymbol{c}^{(t)}$.
    $\boldsymbol{z}^{(t)} \leftarrow \boldsymbol{z}^{(t)} = \boldsymbol{z}_o^{(t)} + \boldsymbol{H}\boldsymbol{c}^{(t)}$.
end
Data Collection:
for each time instance $t$ till $T$ and each bus $i$ do
    $[\hat{\boldsymbol{x}}^{(t)}]_i = [\hat{\boldsymbol{x}}^{(t)}]_i + n_x$, $n_x \sim \mathcal{N}(0, \sigma_n^2)$,
    $[\boldsymbol{z}^{(t)}]_i = [\boldsymbol{z}^{(t)}]_i + n_z$, $n_z \sim \mathcal{N}(0, \sigma_n^2)$,
    $\mathcal{X} \leftarrow \{\mathcal{X}, [\hat{\boldsymbol{x}}^{(t)}]_i\}$, $\mathcal{Z} \leftarrow \{\mathcal{Z}, [\boldsymbol{z}^{(t)}]_i\}$.
end
```

To evaluate the locational detection performance, we employ the Recall and Precision [54] as the performance metrics

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

where true positive (TP), false positive (FP), and false negative (FN), respectively, denote the counts of the samples that the attacked bus is correctly classified as attacked, the normal bus is incorrectly classified as attacked, and the attacked bus is incorrectly classified as normal.

[2] It is verified through our exhaustive experiments that the performance of the detectors in our experiment would scarcely be influenced by $n_{\min}$.

TABLE I
LOCATIONAL DETECTION PERFORMANCE FOR $\mathcal{Z}$ OF THE IEEE 14-BUS POWER SYSTEM

Model	Precision(%)	Recall(%)	F_1(%)	PDR(%)	LDR(%)
DT	98.59	40.77	57.69	51.80	93.61
GNB	87.88	82.72	85.22	72.14	96.93
SVM	87.88	69.08	79.42	67.12	96.17
KNN	95.96	71.52	81.96	66.05	96.03
MLP	**99.11**	86.31	92.26	81.77	98.45
CNN	97.73	87.76	92.48	82.10	98.49
GCN	98.80	88.40	93.31	83.54	98.64
GCAT	98.65	**92.93**	**95.70**	**88.77**	**99.11**

Taking both the Recall and Precision and into consideration, we calculate the F_1 score [54]

$$F_1 = \frac{2\text{Recall} \times \text{Precision}}{\text{Recall+Precision}}. \tag{18}$$

To evaluate the FDIAs detection performance, we also utilize the presence detection rate (PDR), which is the probability that the power system is correctly classified as attacked or not. Furthermore, we utilize the locational detection rate (LDR) [8]

$$\text{LDR} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{19}$$

to evaluate the bus-wise accuracy of locational detection for FDIAs, where true negative (TN) denotes the count of the samples that the normal buses are correctly classified as normal for all time instances. Additionally, we also utilize the FP rate (FPR) [54]

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{FN}} \tag{20}$$

to evaluate the locational detection performance. The performance metrics for each experiment hereinafter are obtained by averaging five independent trials for each model.

We employ the random search strategy to select the hyperparameters, such as the number of layers, the number of units of the hidden feature, and the number of independent attention mechanisms, by utilizing the F_1 score to evaluate the locational detection performance.

C. Locational Detection Performance

We now evaluate the locational detection performance of our GCAT-based detector, incorporated into the proposed architecture illustrated in Fig. 2. Specifically, following [41], we select the admittance matrix of the corresponding grid as the weighted adjacency matrix for the normalized Laplacian matrix in the GCN method, unless stated otherwise.

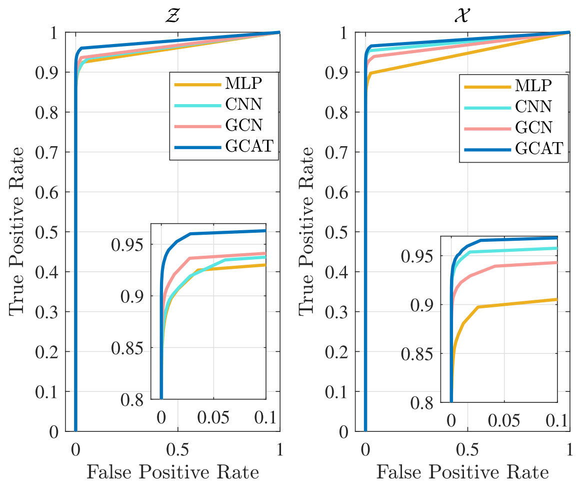


Fig. 5. ROC curves.

TABLE II
LOCATIONAL DETECTION PERFORMANCE FOR $\mathcal{X}$ OF THE IEEE 14-BUS POWER SYSTEM

Model	Precision(%)	Recall(%)	F_1(%)	PDR(%)	LDR(%)
DT	97.09	10.92	19.63	50.15	90.44
GNB	19.91	45.63	27.72	1.19	74.57
SVM	92.41	46.05	61.47	57.25	93.83
KNN	91.70	60.64	73.01	59.47	95.21
MLP	**99.65**	81.71	89.76	77.43	98.01
CNN	99.11	92.85	95.88	89.08	99.15
GCN	99.41	90.23	94.60	86.13	98.90
GCAT	98.53	**93.91**	**96.16**	**89.95**	**99.20**

As demonstrated in Tables I and II, the proposed GCAT model is observed to outperform both the CNN and MLP models, as well as the conventional machine-learning models, with higher Recall, F_1 score, PDR, and LDR, using either the data set $\mathcal{Z}$ or data set $\mathcal{X}$, as incorporating the underlying graph structure of the power grid into the GCAT model could reveal the spatial proximity among the features on neighboring buses from the non-Euclidean power grid data. Moreover, owing to the graph attention mechanism that would elastically assign different weights to each neighboring node, the performance of the GCAT-based detector is observed to be superior to that of the GCN-based detector, for either data sets. It is also observed that with the system state estimates as the input feature, the GCAT-based FDIA detector, as well as both the GCN-based and CNN-based FDIA detectors exhibit superior locational detection performance, in contrast to those with the power injection measurement as the input feature. Whereas the conventional approaches (the DT, GNB, SVM, KNN) are observed to behave unsatisfactorily for either data sets.

We further evaluate the metrics of the TP rate (TPR $\triangleq$ [TP/(TP + FN)]) and FPR, respectively. Notice that the probability that each bus i is under attack is indicated by the corresponding entry of the GCAT-, GCN-, CNN-, and MLP-based FDIA detectors prior to the binary decision. Given the threshold $\tau = 0.5$, we could locate the attacked bus(es) with the quantified probabilities in view of (15). Whereas the threshold ranging from 0 to 1, $\tau \in (0, 1)$, trades off between the TP rate and the FP rate, so that a higher threshold would result in a lower TPR and a higher FPR. Specifically, we have TPR $= 1$ and FPR $= 1$ for $\tau = 0$, whereas TPR $= 0$ and FPR $= 0$ for $\tau = 1$. The TP rate versus the FP rate

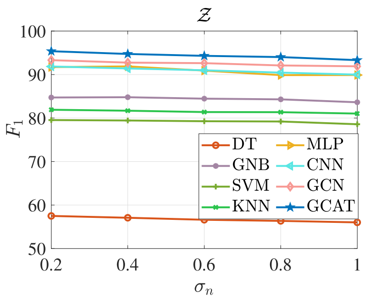


Fig. 6. F_1 versus σ_n for $\mathcal{Z}$.

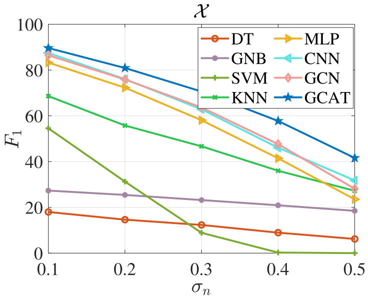


Fig. 7. F_1 versus σ_n for $\mathcal{X}$.

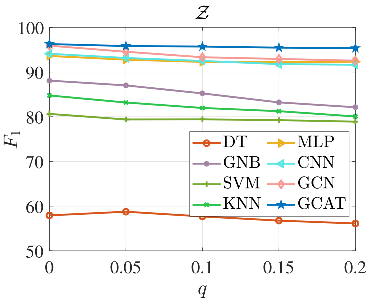


Fig. 8. F_1 versus q for $\mathcal{Z}$.

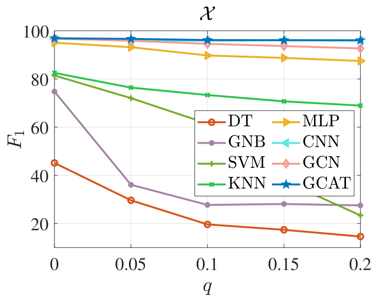


Fig. 9. F_1 versus q for $\mathcal{X}$.

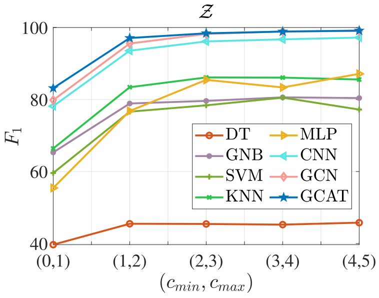


Fig. 10. F_1 versus $(c_{\min}, c_{\max})$ for $\mathcal{X}$.

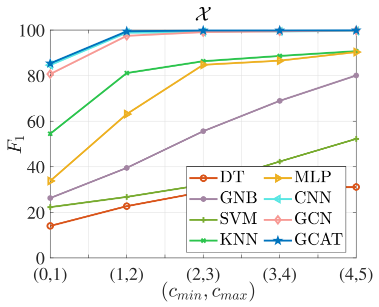


Fig. 11. F_1 versus $(c_{\min}, c_{\max})$ for $\mathcal{X}$.

could be visualized with the receiver operating characteristics (ROCs) curve [54]. Whereas the area under a ROC curve (AUC) could be utilized to measure the separability of FDIA detectors [8], [54].

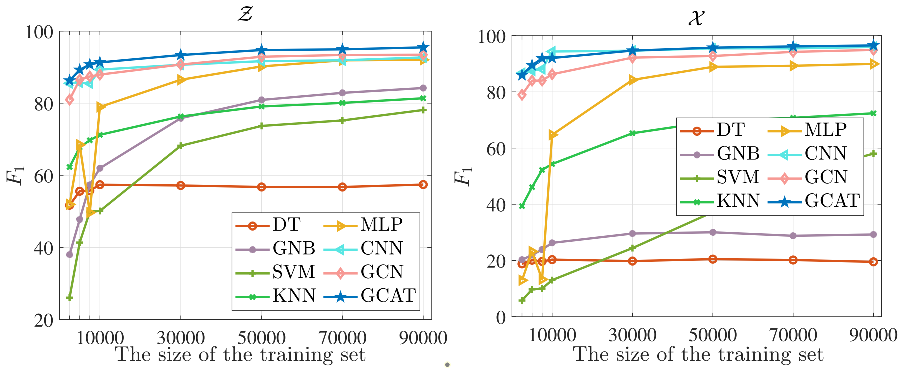
We, respectively, plot in Fig. 5 the ROC curves of the proposed GCAT-based FDIA detector as well as the FDIA detectors based on the GCN, CNN, and MLP, given the threshold ranging from 0 to 1 with the interval 0.05. We could observe that the GCAT-based detector exhibits superior separability and accurate locational detection performance to its competitors, whatever using the data set $\mathcal{Z}$ or data set $\mathcal{X}$.

D. Robustness Validation

We now evaluate the respective impacts of additive noises during data acquisition, the fluctuation intensities of the load and generated power values, and the aggressiveness of attacks on the locational detection performance of the proposed GCAT-based detector.

We first assess the impact of additive noises during data acquisition, by varying the standard deviation σ_n of the noise for each bus. We can observe in Figs. 8 and 7 that for different noise powers, the proposed GCAT-based FDIA detector outperforms the other competitors, with higher F_1 scores, for either the system state estimates or the power injection measurements as the input feature. Moreover, the performance of the FDIA detectors with the power injection measurements as the input feature is observed to be insensitive to the noise power change. In contrast, we could readily observe in Fig. 7 the performance level down of the FDIA detectors utilizing the system state estimates as the input feature with respect to the increasing noise power.

We further validate the impact of the fluctuation intensity of the load and generated power values on buses, for the parameter q ranging from 0 to 0.2. It can be observed from Figs. 8 and 9 that, by using either data set $\mathcal{Z}$ or data set $\mathcal{X}$, the superior margins between the deep-learning-based FDIA detectors and the conventional machine-learning-based FDIA detectors are more pronounced as q increases. The proposed GCAT-based detector is observed to outperform the other competitors, for either data set $\mathcal{Z}$ or data set $\mathcal{X}$. Moreover, the performance of the proposed GCAT-based detector is observed to be scarcely impacted by the drastic change of the load and generation values. Especially, with q increasing, the performance degradation of the DT, GNB and SVM-based



(a) (b)

Fig. 12. F_1 score with different sizes of the training set. (a) Power injection measurements as the input feature. (b) System state estimates as the input feature.

TABLE III
PERFORMANCE WITH THE TWO INPUT FEATURES

Model	Precision(%)	Recall(%)	F_1(%)	PDR(%)	LDR(%)
DT	99.16	41.07	58.08	51.86	93.66
GNB	74.82	82.97	78.69	61.96	95.20
SVM	94.26	63.07	75.57	62.56	95.64
KNN	94.75	66.52	78.17	62.46	96.03
MLP	99.14	89.02	93.81	84.84	98.74
CNN	98.84	93.85	96.28	89.97	99.22
GCN	**99.42**	93.51	96.38	90.28	99.25
GCAT	99.00	**94.45**	**96.67**	**91.03**	**99.30**

TABLE IV
LOCATIONAL DETECTION PERFORMANCE FOR $\mathcal{Z}$ OF THE IEEE 39-BUS POWER SYSTEM

Model	Precision(%)	Recall(%)	F_1(%)	PDR(%)	LDR(%)
DT	90.55	20.60	33.56	50.27	96.87
GNB	51.73	47.13	49.32	43.54	96.28
SVM	77.65	41.52	54.11	54.57	97.29
KNN	91.73	50.36	65.02	56.11	97.92
MLP	**98.74**	82.00	89.56	77.12	99.27
CNN	98.60	87.40	92.66	82.44	99.47
GCN	97.34	89.09	93.03	83.33	99.49
GCAT	97.60	**92.42**	**94.93**	**87.09**	**99.62**

TABLE V
LOCATIONAL DETECTION PERFORMANCE FOR $\mathcal{X}$ OF THE IEEE 39-BUS POWER SYSTEM

Model	Precision(%)	Recall(%)	F_1(%)	PDR(%)	LDR(%)
DT	95.97	3.02	5.85	50.00	96.27
GNB	7.37	12.59	9.30	0.05	90.56
SVM	53.90	2.12	4.08	50.02	96.17
KNN	91.73	50.36	65.02	56.11	97.92
MLP	**99.56**	76.63	86.60	73.15	99.09
CNN	99.33	87.11	92.82	83.05	99.48
GCN	97.08	82.98	89.48	77.07	99.25
GCAT	98.23	**92.08**	**95.05**	**87.41**	**99.63**

FDIA detectors with the power injection measurements $\mathcal{Z}$ as the input feature is observed to be smaller than that of the same detectors with the system state estimates as the input feature.

We then investigate the impact of the aggressiveness of FDIAs for different ranges of $(c_{\min}, c_{\max})$. Particularly, we consider that the random attack error on a bus is either negative or positive. As demonstrated in Figs. 10 and 11, the performance of all the competing detectors improves, with either the power injection measurements or the system state estimates as the input feature, as the aggression of the voltage phase angle increases. The deep-learning-based FDIA detectors are observed to achieve excellent performance with the F_1 score close to 100%, as the aggression of the voltage phase angle is greater than 2°. We could also observe that the proposed GCAT-based detector apparently outperforms the other competitors, as it is more sensitive even to the slight voltage phase angle aggressions ($[\boldsymbol{c}]_i < 1$).

E. Training Set Size

In this experiment, we evaluate the impact of the size of the training set on the performance of the proposed GCAT-based detector. We also set $f_a = 0.1$ for the training set, and $f_a = 0.5$ for the testing set, including 10 000 time instances. Other conditions are the same as those presented in Section IV-C.

As shown in Fig. 12, with the increase of the training set size, the performance of the FDIA detectors is observed to be enhanced, for either the power injection measurements or system state estimates as input feature. We can observe that even with only a limited number of labeled data, the proposed GCAT-based FDIA detector could accurately detect and locate the FDIAs, especially with the power injection measurements as the input feature, in contrast to the other FDIA detectors.

F. Two Features as Input

We herein also consider both the power injections and the system state as the input features, and the result is given in Table III.

TABLE VI
PERFORMANCE FOR THE GCAT-BASED FDIA DETECTOR IN THE IEEE 14-BUS POWER SYSTEM

Dataset	**testing system**	Precision (%)	Recall (%)	F_1(%)	PDR (%)	LDR (%)	FPR(%)
$\mathcal{Z}$	original system	94.22	82.54	88.00	74.45	97.59	0.61
$\mathcal{Z}$	15-bus system	76.26	78.63	77.43	53.58	95.44	2.70
$\mathcal{Z}$	13-bus system	90.42	81.57	85.77	72.94	97.28	0.97
$\mathcal{X}$	original system	97.98	86.22	91.72	81.20	98.34	0.21
$\mathcal{X}$	15-bus system	64.56	86.70	74.01	29.62	93.95	5.25
$\mathcal{X}$	13-bus system	54.07	87.61	66.87	35.95	91.27	8.32

TABLE VII
PERFORMANCE FOR THE GCAT-BASED FDIA DETECTOR IN THE IEEE 39-BUS POWER SYSTEM

Dataset	**testing system**	Precision (%)	Recall (%)	F_1(%)	PDR (%)	LDR (%)	FPR(%)
$\mathcal{Z}$	original system	92.22	77.54	84.25	70.21	98.89	0.26
$\mathcal{Z}$	40-bus system	90.10	72.28	80.21	64.92	98.66	0.31
$\mathcal{Z}$	38-bus system	83.87	50.47	63.02	55.33	97.78	0.38
$\mathcal{X}$	original system	94.05	69.40	79.86	65.68	98.66	0.17
$\mathcal{X}$	40-bus system	60.13	58.21	59.15	34.08	96.98	1.51
$\mathcal{X}$	38-bus system	73.29	51.14	60.24	45.82	97.47	0.73

It is observed that taking both the power injections and the system states rather than either of them as the input features would yield enhanced performance of the detectors, as shown in Table III, in contrast to the results in Tables I and II. Furthermore, the proposed GCAT-based detector is observed to outperform the other competitors.

G. IEEE 39-Bus System

Next, we consider validating the proposed detectors with another system of a larger scale. Specifically, we compare the locational detection performance of the detectors in the IEEE 39-bus system, respectively, taking the power injection measurements and the system state estimates as the input feature.

As illustrated in Tables IV and V, in the IEEE 39-bus system, the performance of the proposed GCAT-based FDIA detector is still superior to other competitors, with higher Recall, F_1 score, PDR, and LDR, using either the data set $\mathcal{Z}$ or data set $\mathcal{X}$. This is reasonable because the proposed GCAT approach models the inherent underlying graph structures of power grids and spatially correlated measurement data residing on the buses of the system.

H. Performance Validation With Different Topologies

We now consider the challenging scenarios, where in the testing phase either a new bus joins the original smart grids or one bus failure occurs, in both the IEEE 14 and 39-bus systems. The specific topologies for the two scenarios in the IEEE 14 and 39-bus systems are shown in Fig. 13. In the first scenario, the bus indexed by 15 (40) joins the original IEEE 14-bus (39-bus) power system. In the second scenario, the bus indexed by 12 (36) and the transmission lines connected to this bus are removed from the original IEEE 14-bus (39-bus) power system. We generate the two corresponding testing sets for the aforementioned scenarios. Each testing set consists of 10 000 time instances.

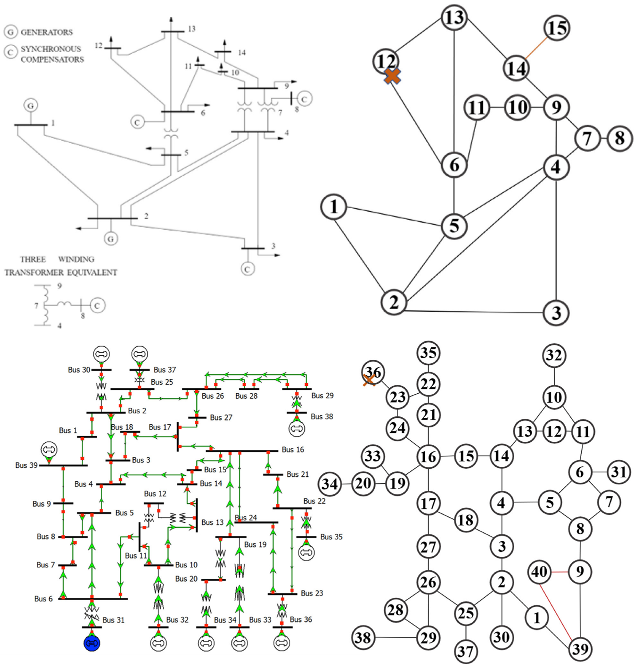


Fig. 13. IEEE 14-bus system [50], IEEE 39-bus system [55], and their corresponding topologies. In the first scenario, bus 15 (40) joins the original IEEE 14-bus (39-bus) power system, yielding a 15-bus (40-bus) system. In the second scenario, bus 12 (36) and the transmission lines connected to this bus are removed from the original IEEE 14-bus (39-bus) power system, yielding a 13-bus (38-bus) system.

To cope with the above situations, we remove both the flatten layer and dense layer in our GCAT model, such that the dimensions of the training parameters of the layers in the resulting simplified GCAT-based detector are irrelevant to the number of nodes. As demonstrated in Tables VI and VII, compared to the original GCAT-based locational detection model, for the data set with the power injection measurements as the input feature, the precision, recall, F_1, PDR, and LDR of the simplified GCAT-based detector decrease by 4.43%, 10.34%, 7.7%, 14.32%, and 1.52% in the IEEE 14-bus power system, and 5.38%, 14.88%, 10.68%, 16.88%, and 0.73% in the IEEE 39-bus power system, respectively. Furthermore, for the data set with the system state estimates as the input feature, the precision, recall, F_1, PDR, and LDR of the simplified GCAT-based detector reduces by 0.55%, 7.69%, 4.44%, 8.75%, and 0.86% in the IEEE 14-bus power system, and 4.18%, 22.68%, 15.19%, 21.73%, and 0.97% in the IEEE 39-bus power system, respectively.

It is noteworthy that the FDIA detectors based on the GCN, CNN, and MLP cannot be directly applied to the scenarios where the power grid topology in the testing phase differs from that in the training phase, as the dimensions of the trainable parameters in the these models are intimately associated with the number of buses (nodes). In contrast, on account of the graph attention mechanism, the simplified GCAT-based FDIA detector could be applied to such challenging scenarios. Moreover, the simplified GCAT-based detector exhibits preferable performance of the location of attacked buses (Precision, Recall, F_1, LDR, and FPR), but slightly worse performance of the detection of the attacked system (PDR), as the accuracy of the detection attacked system requires that all the attacked buses are correctly detected. We can also observe that the simplified GCAT-based locational detector with the power injection measurements as the input feature outperforms that with the system state estimates as the input feature.

V. Conclusion

In this article, utilizing either power injection measurements or system state estimates, we develop a locational detector for FDIAs based on the GCAT, which could be integrated noninvasively into smart grids. By exploiting the graph attention mechanism to extract spatial proximity between the features on neighboring buses from the non-Euclidean graph data of the power grids, the proposed GCAT-based FDIA detector could not only utilize the inherent graph structure of power grids but also adapt the unknown graph shift operators. The superior locational detection performance of the proposed detector is validated through illustrative simulations. We also validate the accurate locational detection performance of the proposed detector even with only a limited number of labeled data. Moreover, the feasibility of the proposed detector to the scenarios where the power grid topology in the training phase differs from that in the testing phase is also experimentally validated.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[2] X. Yu and Y. Xue, "Smart grids: A cyber-physical systems perspective," *Proc. IEEE*, vol. 104, no. 5, pp. 1058–1070, May 2016.

[3] M. Orlando et al., "A smart meter infrastructure for smart grid IoT applications," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 12529–12541, Jul. 2022.

[4] M. Thomas and J. McDonald, *Power System SCADA and Smart Grids*. Boca Raton, FL, USA: CRC Press, Dec. 2017.

[5] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*, vol. 24. Boca Raton, FL, USA: CRC Press, 2004.

[6] H. Lee, Tushar, B. Cui, A. Mallikeswaran, P. Banerjee, and A. K. Srivastava, "A review of synchrophasor applications in smart electric grid," *WIREs Energy Environ.*, vol. 6, no. 3, p. e223, May 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/wene.223

[7] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Informatics*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[8] S. Wang, S. Bi, and Y.-J. A. Zhang, "Locational detection of the false data injection attack in a smart grid: A multilabel classification approach," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8218–8227, Sep. 2020.

[9] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power App. Syst.*, vol. 94, no. 2, pp. 329–337, Mar. 1975.

[10] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–33, May 2011. [Online]. Available: https://doi.org/10.1145/1952982.1952995

[11] 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.

[12] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, "A machine-learning-based technique for false data injection attacks detection in industrial IoT," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8462–8471, Sep. 2020.

[13] N. N. Tran, H. R. Pota, Q. N. Tran, and J. Hu, "Designing constraint-based false data-injection attacks against the unbalanced distribution smart grids," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9422–9435, Jun. 2021.

[14] J. Tian, B. Wang, J. Li, Z. Wang, B. Ma, and M. Ozay, "Exploring targeted and stealthy false data injection attacks via adversarial machine learning," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 14116–14125, Aug. 2022.

[15] 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.

[16] A. Tabakhpour and M. M. A. Abdelaziz, "Neural network model for false data detection in power system state estimation," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, 2019, pp. 1–5.

[17] E. Drayer and T. Routtenberg, "Detection of false data injection attacks in power systems with graph fourier transform," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Anaheim, CA, USA, 2018, pp. 890–894.

[18] E. Drayer and T. Routtenberg, "Detection of false data injection attacks in smart grids based on graph signal processing," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1886–1896, Jun. 2020.

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

[20] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

[21] M.-L. Zhang, Y.-K. Li, and X.-Y. Liu, and X. Geng, "Binary relevance for multi-label learning: An overview," *Front. Comput. Sci.*, vol. 12, no. 2, pp. 191–202, Apr. 2018. [Online]. Available: https://doi.org/10.1007/s11704-017-7031-7

[22] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Mach. Learn.*, vol. 85, no. 3, pp. 333–359, Dec. 2011. [Online]. Available: https://doi.org/10.1007/s10994-011-5256-5

[23] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification," in *Machine Learning: ECML*. Berlin, Germany: Springer, 2007, pp. 406–417, doi: 10.1007/978-3-540-74958-5_38.

[24] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inform. Sci.*, vol. 44, no. 1, pp. 48–59, 2018.

[25] D. A. Pisner and D. M. Schnyer, "Chapter 6—Support vector machine," in *Machine Learning*. New York, NY, USA: Academic, 2020, pp. 101–121. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128157398000067

[26] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.

[27] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0031320307000027

[28] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.

[29] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.

[30] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.

[31] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.

[32] R. Moslemi, A. Mesbahi, and J. M. Velni, "A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4930–4941, Sep. 2018.

[33] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2218–2234, May 2020.

[34] Z. Liu and J. Zhou, "Introduction to graph neural networks," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 14, no. 2, pp. 1–127, Mar. 2020. [Online]. Available: https://www.morganclaypool.com/doi/10.2200/S00980ED1V01Y202001AIM045

[35] R. Ramakrishna and A. Scaglione, "Grid-graph signal processing (grid-GSP): A graph signal processing framework for the power grid," *IEEE Trans. Signal Process.*, vol. 69, pp. 2725–2739, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9415125/

[36] W. Xia, D. He, and J. Chen, "On the PMU placement optimization for the detection of false data injection attacks," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3794–3797, Sep. 2023.

[37] D. Wang, X. Wang, L. Jin, and Y. Zhang, "Detection of power grid disturbances and cyber-attacks based on machine learning," *J. Inf. Security Appl.*, vol. 46, pp. 42–52, Jun. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214212618305866

[38] E. Isufi, F. Gama, and A. Ribeiro, "EdgeNets: Edge varying graph neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7457–7473, Nov. 2022.

[39] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.

[40] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," Feb. 2017, *arXiv:1609.02907*.

[41] O. Boyaci et al., "Graph neural networks based detection of stealth false data injection attacks in smart grids," *IEEE Syst. J.*, vol. 16, no. 2, pp. 2946–2957, Jun. 2022.

[42] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. 6th Int. Conf. Learn. Represent.*, Feb. 2018, pp. 1–12.

[43] 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. [Online]. Available: http://ieeexplore.ieee.org/document/6257525/

[44] S. Tufail, I. Parvez, S. Batool, and A. Sarwat, "A survey on cyber-security challenges, detection, and mitigation techniques for the smart grid," *Energies*, vol. 14, no. 18, p. 5894, Sep. 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/18/5894

[45] A. Gavili and X.-P. Zhang, "On the shift operator, graph frequency, and optimal filtering in graph signal processing," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6303–6318, Dec. 2017.

[46] W. Xia, J. Chen, and L. Yu, "Distributed adaptive multi-task learning based on partially observed graph signals," *IEEE Trans. Signal Inf. Process. Net.*, vol. 7, pp. 522–538, 2021.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1026–1034.

[48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: The MIT Press, 2016.

[49] L. Thurner et al., "Pandapower—An open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6510–6521, Nov. 2018.

[50] R. Christie. "Power systems test case archive." 1993. [Online]. Available: http://labs.ece.uw.edu/pstca/pf14/pg_tca14bus.htm

[51] ERCOT. "Backcasted (actual) load profiles-historical." 2020. [Online]. Available: http://www.ercot.com/mktinfo/loadprofile/alp/

[52] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32. Red Hook, NY, USA: Curran Assoc., Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[53] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, Nov. 2011.

[54] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X

[55] J. Snodgrass et al., "Case study of enhancing the MATPOWER polish electric grid," in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, 2022, pp. 1–6.

Wei Xia received the B.S. degree in communication engineering and the M.S. and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2002, 2005, and 2008, respectively.

From 2009 to 2017, he was with the School of Electronic Engineering, UESTC. From March 2015 to March 2016, he had been on a 12-month sabbatical leave as a Visiting Scholar with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA. Since 2018, he has been with the School of Information and Communication Engineering, UESTC, where he is an Associate Professor. Since August 2023, he also has been with the School of Computer Science and Technology (School of Cyberspace Security), Xinjiang University, Ürümqi, China. His general research interests include statistical signal processing, adaptive signal processing, radar signal processing, and advanced implementation techniques of signal processing algorithms.

Deming He received the B.E. degree from Hohai University, Nanjing, China, in 2021. He is currently pursuing the M.S. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China.

His current research interests include smart grid security and emitter localization technology.

Lisha Yu received the B.S. degree from Harbin Institute of Technology University, Weihai, China, in 2020, and the M.S. degree from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2023.

She is currently with China Satellite Network Exploration Company Ltd., Chongqing, China. Her research interests include smart grid security, distributed graph signal processing, and beamforming technology.