Electric Power Systems Research 238 (2025) 111149

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Spatio-temporal graph attention network-based detection of FDIA from smart meter data at geographically hierarchical levels

Md Abul Hasnat [a,c,*], Harsh Anand [a], Mazdak Tootkaboni [b], Negin Alemazkoor [c]

[a] Department of Systems & Information Engineering, University of Virginia, 151 Engineer's Way, Charlottesville, 22904, VA, USA
[b] Department of Civil & Environmental Engineering, University of Massachusetts Dartmouth, 285 Old Westport Road, Dartmouth, 02747-2300, MA, USA
[c] Department of Civil & Environmental Engineering, University of Virginia, 151 Engineer's Way, Charlottesville, 22904, VA, USA

ARTICLE INFO

Keywords:
Smart meter
Graph neural network
Graph attention
FDIA
Anomaly detection
Time series

ABSTRACT

The power consumption data from residential households collected by smart meters exhibit a diverse pattern temporally and among themselves. It is challenging to distinguish between regular consumer behavior and injected falsified measurements into the data stream with the intent of energy theft or compromising the security of the associated measurement infrastructure. This work identifies the challenges of detecting falsified measurements in smart meter data aggregated at geographically hierarchical levels and proposes a novel graph attention network (GAT)-based unsupervised learning framework to detect false data injection attacks (FDIA) from the moving statistics of the power consumption data in real-time, namely MOVSTAT-GAT. The proposed technique is capable of detecting falsified measurements at both 9-digit and 5-digit ZIP code labels in an unsupervised manner, solely from smart meter power consumption data with no additional meters. Moreover, the proposed technique offers a visualization technique to assist the operator in identifying the localization characteristics of the attack and proposes an automated localization strategy for localized FDIAs. Experiments suggest the effectiveness of the proposed framework, especially for localized FDIA or external anomalies, such as power outages and denial-of-service (DoS). Additionally, a detailed discussion regarding the implementation of MOVSTAT-GAT in the industrial environment has been provided.

1. Introduction

Smart meters, installed at customer sites within the electrical distribution network, play a pivotal role by recording and transmitting data related to each consumer's electricity consumption [1–3]. This data, whether used independently or in conjunction with other distribution system information, enables the automation and intelligent management of various tasks essential to the operation, security, and reliability of the distribution system at the utility level. Key applications include forecasting load demands, adjusting loads to prevent outages, automated billing, pricing, demand response programs, and the management of daily and critical peak shifts—all dependent on the quality and integrity of smart meter data [4]. Additionally, when combined with sub-station level measurements, smart meter data supports a range of grid security-related applications [5,6] and is crucial for analyzing electricity consumer behavior [7]. This analysis is vital for long-term grid planning [8,9] and to address issues like energy poverty, highlighting the need for accurate data. However, the integrity of these data can be compromised by the injection of false measurements, either to reduce electricity bills (i.e., *energy theft*) or to disrupt utility services by adversaries.

To address the vulnerabilities in smart meter data and enhance the reliability required for critical applications, developing robust detection mechanisms for false data injection attacks (FDIA) is crucial. FDIA [10,11] involves the injection of false measurements that distort the actual data, potentially compromising various operational and planning activities. Here, we present a general mathematical framework for FDIA, focusing on smart meter power consumption data, similar to the attack models discussed in [12,13]. By analyzing the signatures of these FDIAs in power consumption time series, our framework aims to distinctly differentiate between these malicious manipulations and anomalous, yet honest consumer behaviors. This distinction is vital for utilities striving to maintain system integrity and reliability, underscoring the significant challenges posed by sophisticated data attacks.

The detection mechanisms for FDIA and energy theft in smart meter data depend significantly on the resources, strategies, and planning of utility systems. These mechanisms can operate at the user level, utilizing direct meter readings, or at the system level, using aggregated readings from multiple meters within a utility [14]. At the user level, detection is challenging mainly due to the lack of temporal correlation

* Corresponding author.
E-mail address: jau2et@virginia.edu (M.A. Hasnat).

https://doi.org/10.1016/j.epsr.2024.111149
Received 25 May 2024; Received in revised form 11 July 2024; Accepted 2 October 2024
Available online 11 October 2024

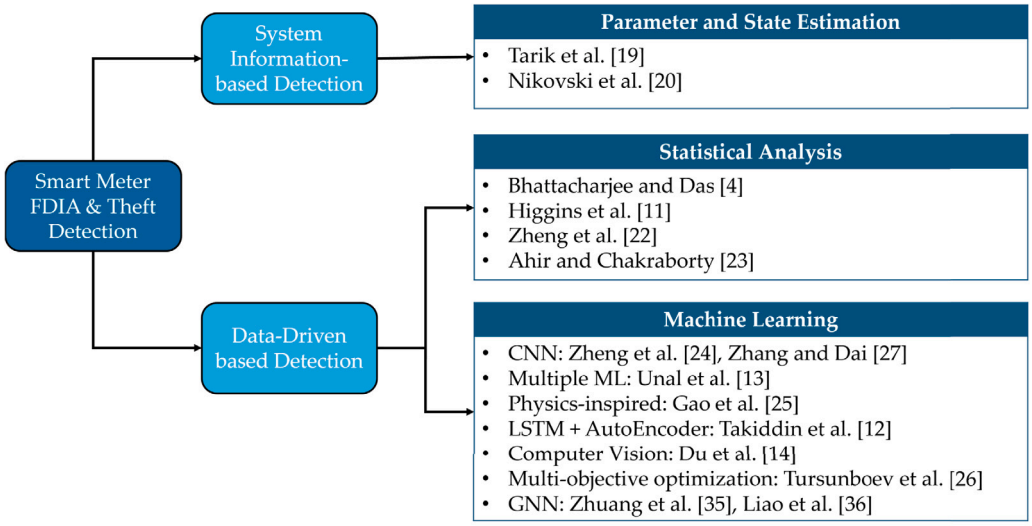


Fig. 1. An up-to-date review of smart meter FDIA and theft detection techniques focusing on machine learning methods.

and meter-to-meter correlation resulting from high stochasticity related to the consumer behavior pattern. Moreover, the abundance of fluctuations and peaks [15], especially in the more granular data, creates additional challenges for abrupt change detection-based techniques to detect FDIAs from a single smart meter time series. Conversely, system- or utility-level detection is more feasible due to the presence of spatio-temporal correlations among aggregated data, which can sometimes be enhanced with the aid of additional measurement devices at the system level. This level of detection is particularly effective in identifying organized attacks [4] and is more manageable from a systems management perspective. However, a notable limitation at this level is that household-level changes often have an insignificant reflection on the aggregated smart meter time series, potentially obscuring subtle anomalies. This paper presents an analysis of both the realities and challenges associated with FDIA detection at the user and system levels, considering the trade-offs and benefits of various strategies from spatio-temporal correlation perspectives.

1.1. Related work

The literature on energy theft and FDIA detection in electrical distributed systems is extensive [16–18] and can be broadly categorized into two main approaches: (1) techniques that rely on grid information such as distribution system states, topology, and system parameters [19,20], and (2) techniques that primarily utilize power consumption data from users, sometimes in combination with grid information [21]. Although the first category typically offers higher accuracy, the second category is gaining increasing attention from researchers for several reasons. Firstly, techniques in the first category often require comprehensive information about system parameters and states, which may not always be available. Secondly, these methods are system-specific and generally require significant modifications to be adapted to new systems. Thirdly, the widespread deployment of smart meters in households and factories provides granular power consumption data, facilitating data-driven analysis for theft detection. Consequently, this section will focus exclusively on the data-driven detection of energy theft and FDIA using smart meter data.

Further, the methods utilizing data-driven techniques for FDIA or energy theft detection can be categorized into two groups: statistical analysis-based detection that requires no training and training-based detection using machine learning. Among statistical analysis-based detection, Bhattacharjee and Das [4] proposed a two-step statistical technique to detect data falsification. In the first step, falsification is identified by the ratio of harmonic to arithmetic mean for a certain duration, which is confirmed in the second step by comparing the ratio with a safe margin. However, this technique requires a longer FDIA duration to detect anomalies effectively. Zheng et al. [22] utilized a combination of an information theory-based technique and a clustering-based technique to detect energy theft from granular power consumption data. However, their technique requires observer meters to determine the non-technical loss. Ahir and Chakraborty [23] analyze the shapes of the power consumption time series corresponding to different contexts (e.g. weekends or weekdays, seasons, etc.) to develop a theft detection technique combining dynamic time warping and k-nearest neighbor which can provide anomaly scores for each detected case. Higgins et al. [11] suggest clustering of the smart meter power consumption data using statistical features and subsequently using an incentive-weighted anomaly detection technique to detect FDIA from the power consumption data.

The second category of methods, i.e., training-based detection using machine learning, dominates the recent literature. For instance, Zheng et al. [24] proposed a wide and deep convolutional neural network for energy theft detection. This technique requires the labels of the theft data to formulate the detection as a supervised learning problem, and longer theft durations are required to be detected. The authors in Ünal et al. [13] achieved promising accuracy in detecting FDIA in smart meter data by combining several machine learning, deep learning, and parallel computing techniques. However, this technique also requires the observer meter to obtain non-technical loss. Gao et al. [25] proposes a physically inspired data-driven model to detect energy theft. They developed a modified linear model to capture the relationship between the amount of electricity used and the voltage magnitudes recorded by the smart meters and utilized the regression residual to detect theft. Takiddin et al. [12] proposes an LSTM-based auto-encoder to capture the temporal as well as meter-to-meter dynamics among the smart meter data with an hourly temporal resolution to design an unsupervised learning-based framework for the detection of energy theft. Du et al. [14] proposed converting the 1-D power consumption time series to 2-D image data for obtaining visually distinguishable features to detect FDIA using a computer vision-based approach. Tursunboev et al. [26] proposed to integrate an evolutionary multi-objective optimization to maximize precision and recall in a hybrid deep learning method to detect energy theft under a supervised learning framework. Zhang and Dan [27] presented an explainable feature extraction framework along with a convolutional neural network and attention-based technique for energy theft detection.

The spatio-temporal analysis, particularly the application of graph neural networks (GNNs) on smart meter data is very limited. Given

the ability of GNNs to capture the complex interaction among the system components through data, they have recently been applied to enhance the robustness and accuracy of anomaly and event detection in various sectors, e.g. power grid [28–30], internet-of-things [31], biomedical engineering [32], and social and financial networks [33,34]. For smart meter data, a couple of very recent works utilized spatio-temporal graph neural networks for energy theft detection under supervised learning frameworks using labeled theft data. For instance, Zhuang et al. [35] and Liao et al. [36] proposed GNN-based theft detection techniques using supervised learning frameworks. A chart on the recent smart meter FDIA and energy theft literature has been provided in Fig. 1 for the convenience of readers to follow the above discussions. In this work, we present a spatio-temporal analysis of smart meter power consumption data at different geographic hierarchical levels and propose a graph attention network (GAT)–based FDIA detection technique in a multi-variate time-series setting. Unlike [36], the proposed technique considers a geographic entity (household or 9-digit ZIP code depending on the geographic hierarchy) as the graph vertices, therefore, facilitating direct location identification along with detection of the attack in real-time. Moreover, unlike the recent works in [26,35,36] the proposed framework is completely unsupervised and therefore does not require labeled theft or corrupted data to train the model. Additionally, this work recognizes the effectiveness of the proposed work in [4] to utilize moving statistics (AM/HM ratio) to detect energy theft. Nevertheless, in our experiments, with more types of FDIA definitions and with smaller duration attacks, a generalized training-based unsupervised approach is required to detect FDIAs with precision i.e., distinguishing them from irregular yet honest customer behaviors. Therefore, for the first time, the moving statistics-based feature extraction technique (MOVSTAT) has been incorporated into the GAT-based analysis to extract important information from the multivariate smart meter data for FDIA detection.

1.2. Challenges, limitations, and gaps in the literature

The deployment of smart meters has enabled the application of data-driven, particularly machine learning-based methods, to effectively detect FDIAs and energy thefts in the electric system. These methods operate without the need for knowledge of system states, topologies, and parameters. Despite these advancements, several research challenges persist regarding the precise detection and location identification of FDIAs with low false alarm rates. Firstly, a number of data-driven techniques assume the presence of observer meters alongside household smart meters to calculate non-technical losses, a scenario that is often impractical for utility companies [22]. Secondly, FDIA detection is frequently approached as a supervised learning problem requiring labeled data for the corrupted states [24,26], yet in realistic settings, labeled theft data is rarely available. Such supervised classification of theft and honest data sets limits the ability to detect new types of theft or FDIA not present in the training datasets. Thirdly, methods that treat FDIA detection as moving statistics-based abrupt change detection often suffer from high false positives, particularly when data are granular and the theft duration is short [4]. Fourthly, while many supervised learning techniques can classify users as honest or corrupt in an offline setting, the literature lacks approaches for real-time identification of attack locations post-detection. Thus, this limits utility companies to comprehend the holistic depiction of the attackers' strategy. Finally, reported detection accuracies in the studies vary widely due to different attack models and a range of attack parameters. Moreover, the same attack parameters may have varying impacts depending on the statistics of the power consumption data and the utility system's characteristics, such as the number of households or meters. Therefore, it is crucial to analyze the sensitivity of model parameters to different accuracy metrics and establish relationships between model parameters and attack parameters to adapt detection models to utility systems of various sizes and geographic hierarchies.

1.3. Main contributions

This work addresses the listed shortcomings by proposing the MOVSTAT-GAT technique, an innovative approach for data-driven detection of energy theft and FDIA from high-resolution smart meter data. This technique eliminates the need for an observer meter, operates under an unsupervised learning paradigm that does not require labeled corrupted data, and is capable of detecting an FDIA almost instantaneously at its onset, making it suitable for identifying short-duration attacks. MOVSTAT-GAT can detect FDIAs using both individual and aggregated smart meter data across different geographic levels: using multivariate time series corresponding to individual meter power consumption at 9-digit ZIP code levels, and using multivariate time series corresponding to individual 9-digit aggregated power consumption at 5-digit ZIP code levels. Here, GNN, specifically the graph attention network (GAT) is utilized to capture the spatial correlation among smart meters or different 9-digit ZIP code aggregated time series within a 5-digit ZIP code. The key contributions of our work are outlined as follows:

- The spatio-temporal aspects of the smart meter data at different aggregation levels (user level, 9-digit, and 5-digit ZIP code levels) are analyzed to understand FDIA detection challenges at different hierarchical levels with the existing techniques.
- We propose MOVSTAT-GAT, an FDIA detection technique for multivariate time series, combining moving statistics-based feature extraction and graph attention networks. To the best of our knowledge, it is the first unsupervised real-time FDIA and energy theft detection method using a spatio-temporal graph neural network of any kind. Our technique offers excellent detection accuracy, especially for small-magnitude FDIAs, and facilitates the localization of clustered attacks. Moreover, MOVSTAT-GAT does not require the assumption of any additional meters apart from the household smart meters.
- The proposed framework supplements a visualization tool for utility providers, alongside an automated scheme to aid human operators in assessing FDIA characteristics such as localization, spread, duration, and intensity of the attacks.
- A detailed discussion on implementing the proposed technique by electric utility companies covers model parameter tuning for different locations and geographic levels, considering utility strategies and resources. It also addresses scalability, robustness to new attack types, and the relationship between model and attack parameters in the context of FDIA detection.

2. Data and attack models

2.1. Mathematical representation of power consumption data

The power consumption signal $p(n,t)$ represents the consumption of electric power at the time t by the nth household captured by the nth smart meter, where $t \in \mathcal{T}$ and $n \in \mathcal{N}_i$. $\mathcal{T}$ is the set representing the total duration of data collection and $\mathcal{N}_i$ is the set of all smart meters mounted on the households located at the ith 9– digit ZIP codes. The cardinality of the set, $|\mathcal{N}_i| = N_i$, represents the total number of smart meters, i.e., households located at the ith 9– digit ZIP code. The aggregated signal at the 9– digit ZIP code level is defined as:

$$p_N(m,t) = \sum_{n=1}^{N_m} p(n,t), \tag{1}$$

where, $p_N(m,t)$ is the aggregated power consumption signal the mth 9– digit ZIP code, and $m \in \mathcal{M}_j$. Here, the set $\mathcal{M}_j$ is the set of all 9– digit ZIP codes within the jth 5– digit ZIP code. The aggregated signal at the 5– digit ZIP code level is defined as:

$$p_M(l,t) = \sum_{m=1}^{M_l} p_N(m,t), \tag{2}$$

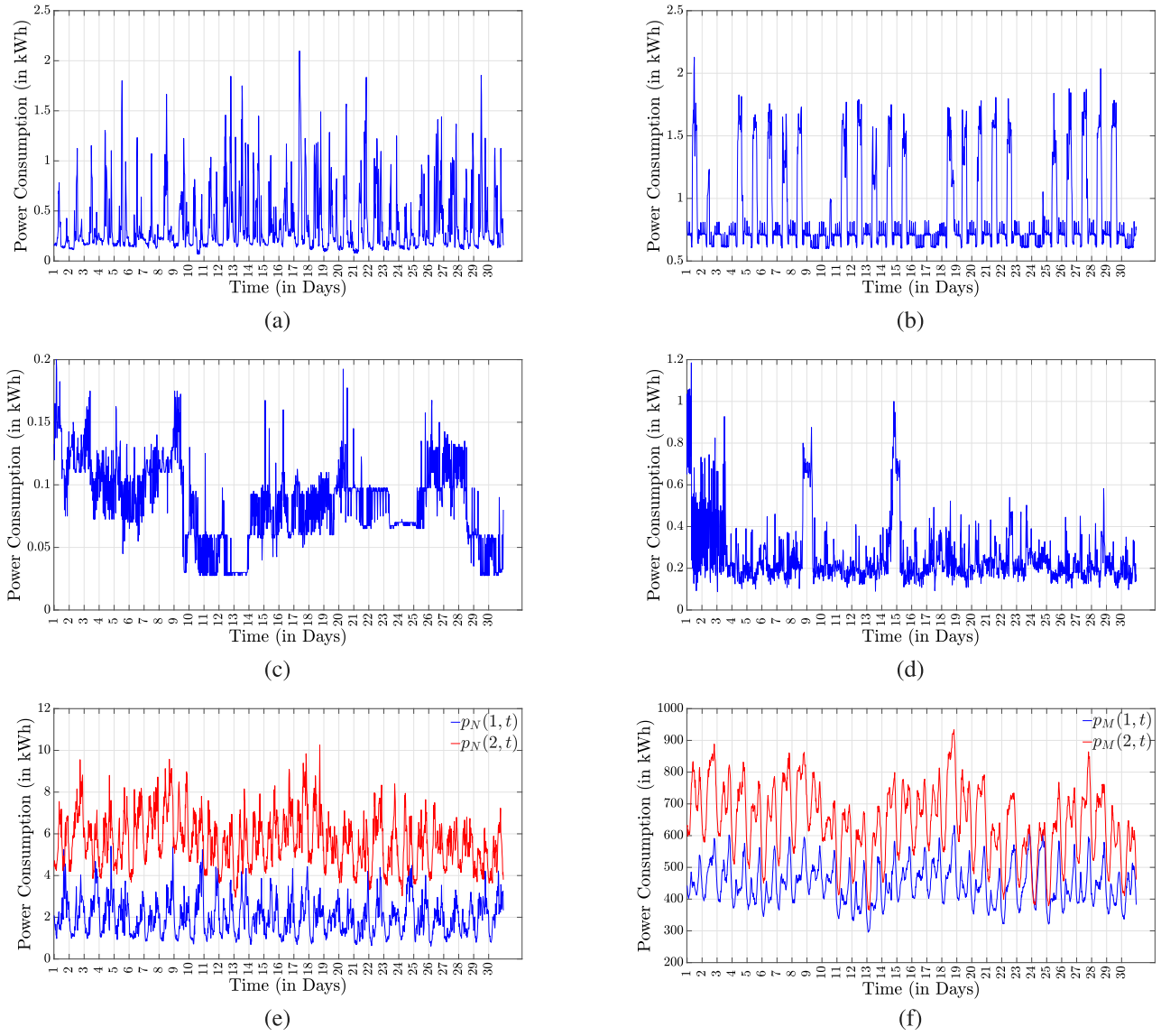


Fig. 2. Household power consumption time series measured by smart meters, (a)–(d): four randomly selected households exhibiting varying patterns: (a) mostly daily periodicity, (b) both daily and weekly periodicity, (c) highly irregular temporal behavior, and (d) daily and weekly periodicity in the later part of the month. **Aggregated power consumption time series (e)–(f):** at two random (e) 9-digit ZIP codes and (f) 5-digit ZIP codes. Household and ZIP code numbers are arbitrary and have no connection with the actual addresses.

where, $M_j = |\mathcal{M}_j|$. Fig. 2 illustrates the power consumption data recorded in April 2019 from four different households using four smart meters within a single 9– digit ZIP code. The power consumption patterns over the month in the four households are very different, although each household is likely to have temporal patterns to some extent. For instance, the daily and weekly periodicity is observed in most of the households (easily observable in Fig. 2(b), but also present in Fig. 2(a) and most of the latter portion of 2(d). However, this daily and weekly periodicity varies extensively among the houses due to various factors, e.g., the number of members in the household, the electrical appliances used by a particular household, the work schedule, and the lifestyle of the household members. Moreover, for some of the households, the irregular portions of the data consumption are frequent and may contain peaks or high fluctuations, as in Fig. 2(c). Therefore, the meter-to-meter correlation is very irregular and difficult to capture. This issue poses a challenge in utilizing meter-to-meter correlation in the spatio-temporal analysis of smart meter data in detecting FDIA. However, learning the node embeddings (i.e., vector representation of each node/user) and dynamic graphs from the time series eases these challenges to some extent and enables utilizing the minimal correlation with only a few (smart meters) nodes.

The aggregated time series at two different 9-digit ZIP code levels have been illustrated in Fig. 2(e) indicating a more regular behavior than a single smart meter data in terms of periodicity and smoothness in the time domain. Moreover, better similarity in terms of shape and fluctuation pattern can be observed among the aggregated time series at different 9-digit ZIP code levels than among the time series corresponding to individual meters. Aggregated time series at the 5-digit ZIP code level shows an even better periodicity, smoothness, and inter-time series correlation among themselves (Fig. 2(f)). However, changes in the smart meter readings which are small in magnitude or pertaining to a small number of meters in a localized geographical

area are not well reflected in aggregated time series. Through our experiments, we show that localized attacks are better detected by MOVSTAT-GAT than by single-variate (LSTM-based) analysis.

2.2. False data injection attack (FDIA) models

This section presents the mathematical approach for modeling five types of FDIAs and their physical significance in smart meter infrastructures and utilities. A generic definition of FDIA on smart meter time series at the household level is introduced, with the various types considered as special cases. Similar attacks have been previously discussed in the smart meters FDIA literature [12,13]. Let us consider a set of households, $\mathcal{N}_{A_i} \subset \mathcal{N}_i$ in the ith 9-digit ZIP code, which is under FDIA within the time interval $[t_{start}, t_{end}] \subset \mathcal{T}$. $\gamma = \frac{|\mathcal{N}_{A_i}|}{|\mathcal{N}_i|}$ signifies the fraction of compromised smart meters within the 9-digit ZIP code. The time series associated with the attacked meters within the attack interval can be expressed as:

$$p(n_A, t) \equiv c(t), \quad for \quad t_{start} \le t \le t_{end}, \quad and \quad n_A \in \mathcal{N}_{A_i}. \tag{3}$$

$c(t)$ is the generic corrupted signal that can be defined to model different types of FDIA and thefts in smart meter data, as discussed here.

2.2.1. Type I FDIA

The first type of FDIA considers reducing the power consumption of the meter by subtracting a constant value, δ, from the power consumption value of all the compromised meters: $c(t) = p(n_A, t) - \delta$, where $\delta \in \mathbb{R}$ can be termed as the magnitude of the FDIA. For $\delta > 0$, this type of FDIA signifies energy theft, since a reduced amount of power consumption lowers the electricity bills. However, adversaries can also launch FDIA with $\delta < 0$, purposely seeking to harm the reputation of utilities.

2.2.2. Type II FDIA

The second type of FDIA is a modified and more realistic version of the first one. The Type I FDIA can be easily detectable for large $\delta > 0$ for the resulting negative value of $p(n_A, t)$. Type II FDIA replaces the negative values $p(n_A, t)$ with zeros to avoid easy detection by the utilities. Mathematically, Type II FDIA can be described as: $c(t) = \max\{p(n_A, t) - \delta, 0\}$. Type II FDIA can also be seen as a form of energy theft.

2.2.3. Type III FDIA

Type III FDIA involves scaling the value of the power consumption by a scalar factor $\beta \in \mathbb{R}$: $c(t) = \beta p(n_A, t), \quad \beta \ge 0.$, Similar to Type I, it can represent energy theft for $\beta < 1$ by lowering energy consumption, or it can purposely damage the reputation of the utilities by increasing energy consumption ($\beta > 1$).

2.2.4. Type IV FDIA

Type IV FDIA is mathematically represented as $c(t) = 0$. Although it can be seen as a special case of Type III FDIA with $\beta = 0$, it can be associated with various physical conditions of the smart metering infrastructure. It can indicate a naive strategy for energy theft, causing smart meter readings to zero, a denial-of-service (DoS) attack from cyber-attacks or physical disconnection or damage to the meters, or a power outage affecting multiple households. Detecting this type of FDIA is crucial for utility providers.

2.2.5. Type V FDIA

Type V FDIA represents a modified and more sophisticated version of energy theft which limits the power consumption smart meter reading of a household to a certain power consumption value, λ: $c(t) = \min\{p(n_A, t), \lambda\}$.

The detection performance of the smart meters FDIAs at both hierarchical levels depends on the type of the FDIAs as well as the parameters associated with the FDIA models, e.g. $\gamma, \delta, \beta, \lambda$. This dependence of detection performance on attack type and parameters is analyzed in Section 4.

2.3. Random and clustered attacks

FDIAs can be launched in the system either randomly or in a clustered way, depending on adversaries' location, resources, intention, and strategy. Clustered attacks at the 9-digit and 5-digit ZIP code levels involve adjacent smart meters or meters in adjacent ZIP codes, respectively. Detecting and localizing these clustered attacks is crucial for utilities as they indicate localized adversaries, or power outages or DoS in specific geographic areas.

3. FDIA detection technique using MOVSTAT-GAT

This section details FDIA detection techniques (shown in Fig. 3) at both the 5-digit and 9-digit ZIP code levels. We propose GNN-based FDIA detection using moving statistics-based features from multivariate power consumption time series (MOVSTAT-GAT) for both hierarchical levels. At the 9-digit ZIP code level, detection considers all individual meter power consumption time series, $p(n, t)$, while at the 5-digit ZIP code level, it considers aggregated power consumption time series, $p_N(l, t)$, from the 9-digit levels. The following subsections present the detailed steps of the MOVSTAT-GAT technique using a generic multivariate time series $x(n, t)$, where $x(n, t) \equiv p(q, t)$ for user-level data in 9-digit ZIP codes, or $x(n, t) \equiv p_N(q, t)$ for aggregated data at 5-digit ZIP codes. Specializations for each hierarchical level are noted as needed.

3.1. Prepossessing using periodicity removal

Smart meter data shows weekly and daily periodicity, varying with household consumption patterns. This periodicity is more prominent in aggregated data (e.g., at the 9-digit ZIP code level) because of the averaging effect. To emphasize the fluctuation due to FDIA by separating them from the periodic fluctuations, these periodicities from the time series: $x'(n, t) = x(n, t) - x_w(n, t) - x_d(n, t)$, where, $x_w(n, t)$ and $x_d(n, t)$ represent weekly and daily fluctuations, respectively, which are estimated from historical data. It is worth mentioning that since the temporal duration of the dataset used in this work is only one month, the seasonal (yearly) periodicity could not be removed which could increase detection accuracy for longer-duration data sets.

3.2. Extracting moving statistics-based Features (MOVSTAT)

Instead of applying GAT to the raw periodicity-removed power consumption data, we propose a moving statistics-based transformation (MOVSTAT) to reduce the stochasticity and uncertainty, improving the distinction between genuine and corrupted data. The transformed time series corresponding to the nth component (i.e., nth smart meter data or aggregated time series corresponding to the nth 9-digit ZIP code), is described by the following equation: $y(n, t) = \mathcal{T}_{t_d}^{STAT}(x'(n, t))$, where $\mathcal{T}_{t_d}^{STAT}(.)$ is the generic moving statistics operator that calculates the temporal statistics of a time series within the interval $[t_d, t]$. For example, $\mathcal{T}_{t_d}^{AM}(x'(n, t))$ calculates the temporal arithmetic mean of the time series $x'(n, t)$ within the interval $[t_d, t]$, while $\mathcal{T}_{t_d}^{SD}(x'(n, t))$ calculates the temporal standard deviation of time-series $x'(n, t)$ within that interval. The effects of choosing the proper moving statistics have been discussed in Sections 4.7.1 and 5.3.3. The presented results in this article are generated using $\mathcal{F}_{13}^{SD}$ unless mentioned otherwise.

3.3. GAT-based prediction of $y(n, t)$

In this work, we primarily adopted the graph attention-based time-series prediction techniques from [37]. However, for the detection stage, we replaced these techniques with a proposed modified method described in Section 3.4. To predict the transformed time series, i.e, time-varying moving statistics-based features, $y(n, t)$ at any time t, the past w samples are used:

$$\hat{y}(n, t) = \mathcal{F}[y(n, t-1), y(n, t-2), \ldots, y(n, t-w+1)] = \mathcal{F}[\mathbf{Y}],$$

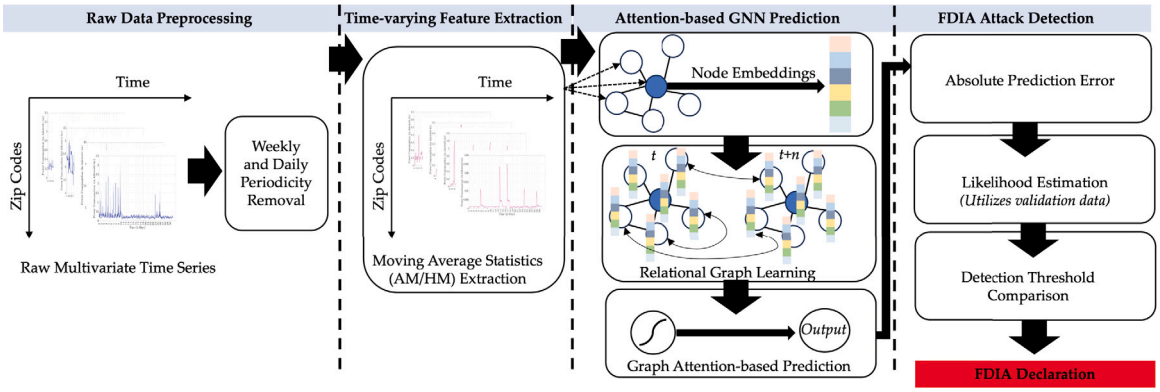


Fig. 3. Schematic diagram of the proposed MOVSTAT-GAT technique for FDIA detection.

where, $\mathbf{Y}$ is a $N \times w$ matrix containing the values of $y(n,t-1), y(n,t-2), \ldots, y(n,t-w+1)$ and $\mathscr{F}$ is the prediction operator. For implementing the prediction method using GNN, the associated graph, $\mathcal{G}$ is learned from the data (i.e., $y(n,t)\forall n,t$) themselves. The graph, $\mathcal{G}$ is defined by the binary adjacency matrix, $\mathbf{A}$ with elements: $A_{i,j} = 1$, if $\mathbf{v}_i.\mathbf{v}_j \in \mathscr{M}_i^K(\mathbf{v}_i.\mathbf{v}_j)$, and 0, otherwise, where, $\mathbf{v}_k$ is $d-$ dimensional embedding vector of node k. The embedding vector of all the nodes, $n \in \mathcal{N}$ is defined as: $\mathbf{v}_k = \mathscr{E}(\mathbf{Y}), \forall k \in \mathcal{N}$, The embedding operator $\mathscr{E}$ is a function that learns the embedding of each node from the data, i.e. $\mathbf{Y}$, and $\mathscr{M}_i^K(\mathbf{v}_i.\mathbf{v}_j)$ is the set of top K values of $\mathbf{v}_i.\mathbf{v}_j$ for the ith node.

The data matrix, $\mathbf{Y}$ is then passed through a linear system characterized by $\mathbf{W_{Lin}}$ and subsequently concatenated with the embedding vectors: $\mathbf{G} = \left[\mathbf{V} \;\; \mathbf{W_{Lin}}\mathbf{Y}^{\mathbf{T}}\right]^{\mathbf{T}}$, where, $\mathbf{G}$ is a $2d \times N$ matrix, which is called the gate of a GAT [37], $\mathbf{W_{Lin}}$ is a $d \times w$ matrix with learnable entries, $\mathbf{V}$ is a $d \times N$ containing the embedding vectors $\mathbf{v}_k$ in its columns. The query ($\mathbf{q}$) and the key ($\boldsymbol{\kappa}$) of the attention network are calculated, respectively, as: $\mathbf{q} = \mathbf{a}_{\mathbf{q}}^{\mathbf{T}}\mathbf{G}$, $\boldsymbol{\kappa} = \mathbf{a}_{\boldsymbol{\kappa}}^{\mathbf{T}}\mathbf{G}$,where, $\mathbf{a_q}$ and $\mathbf{a}_{\boldsymbol{\kappa}}$ are both $2d \times 1$ vector with learnable entries, and $\left[\mathbf{a}_{\mathbf{q}}^{\mathbf{T}} \;\; \mathbf{a}_{\boldsymbol{\kappa}}^{\mathbf{T}}\right]^{\mathbf{T}}$ is called the attention weight matrix. The graph attention scores are calculated as:

$$\alpha_{ij} = \text{SOFTMAX}_i\left(\text{LeakyReLu}\left(\left(\mathbf{Q}+\mathbf{K}^{\mathbf{T}}\right) \bigotimes \mathbf{A}\right)\right)$$

where $\mathbf{Q}$ and $\mathbf{K}$ are two $N \times N$ matrices having $\mathbf{q}$ and $\boldsymbol{\kappa}$ in each of the columns, respectively, and $\bigotimes$ denote element-wise matrix multiplication of two matrices. $LeakyReLu$ is the non-linear activation function operator [38] and $SOFTMAX_i$ signifies applying the Softmax operator along the ith row of the matrix. The attention scores are then used to obtain the aggregated representation of $k-$ th node as:

$$\mathbf{z}_k = \text{ReLU}\left(\sum_{j \in k \cup \mathcal{N}\mathcal{B}_1(k)} \alpha_{kj}\mathbf{w}_1\mathbf{Y}\right)$$

The matrix, $\mathbf{Z}$ contains the $\mathbf{z}_k$ in the columns. The forecast values at each node k at time t is obtained as: $\hat{y}(n,t) = \mathscr{P}\left[\mathbf{z}_k \cdot \mathbf{v}_k\right]$, where $\mathscr{P}$ is an operator with learnable parameters, $\mathbf{W}_p$ implemented by a fully connected network with hidden layers.

3.4. Declaration of FDIA

Once $\hat{y}(n,t)$ is predicted for any instant, t, the normalized absolute error of prediction can be calculated as: $e(n,t) = |\hat{y}(n,t) - y(n,t)|$. From the prediction error, an FDIA is declared to be detected at time, t, if the likelihood of the normalized absolute error falls below a certain likelihood threshold, θ_l, described by:

$$f_{1,2,\ldots N}(e(1,t), e(2,t), \ldots e(N,t)) < \theta_l \tag{4}$$

where $f_{1,2,\ldots N}(\zeta_1, \zeta_2, \ldots \zeta_N)$ is the joint probability distribution function of $e(n,t)$ which should be estimated from the historical data. However, in this work, the marginal distributions $f_n(\zeta)$ for each n are suggested to be considered as independent normal distributions with means μ_{e_n} and standard deviation σ_{e_n}, which are estimated from the validation data. Therefore, Eq. (4) can be approximated as:

$$\sum_{n=1}^{N}\left|\frac{e(n,t)-\mu_{e_n}}{\sigma_{e_n}}\right| > \theta, \tag{5}$$

θ is the threshold of detection. $\hat{e}(n,t) = \left|\frac{e(n,t)-\mu_{e_n}}{\sigma_{e_n}}\right|$ is named as the estimated normalized absolute error. The dependency of detection performance on θ has been discussed in 4.7.2.

3.5. Location identification of clustered attacks

Once MOVSTAT-GAT detects an FDIA, the technique supplements identifying the locational characteristics of the injected FDIA by visualizing estimated normalized absolute error, $\hat{e}(n,t)$ as an image. The clustered attack, which is usually important from the location identification perspective, can be distinguished from the image pixels' intensities. A location within the attack cluster can be identified as:

$$n_{loc} = arg\max_n \hat{e}(n,t) = arg\max_n \left|\frac{e(n,t)-\mu_{e_n}}{\sigma_{e_n}}\right| \tag{6}$$

4. Simulation and analysis of results

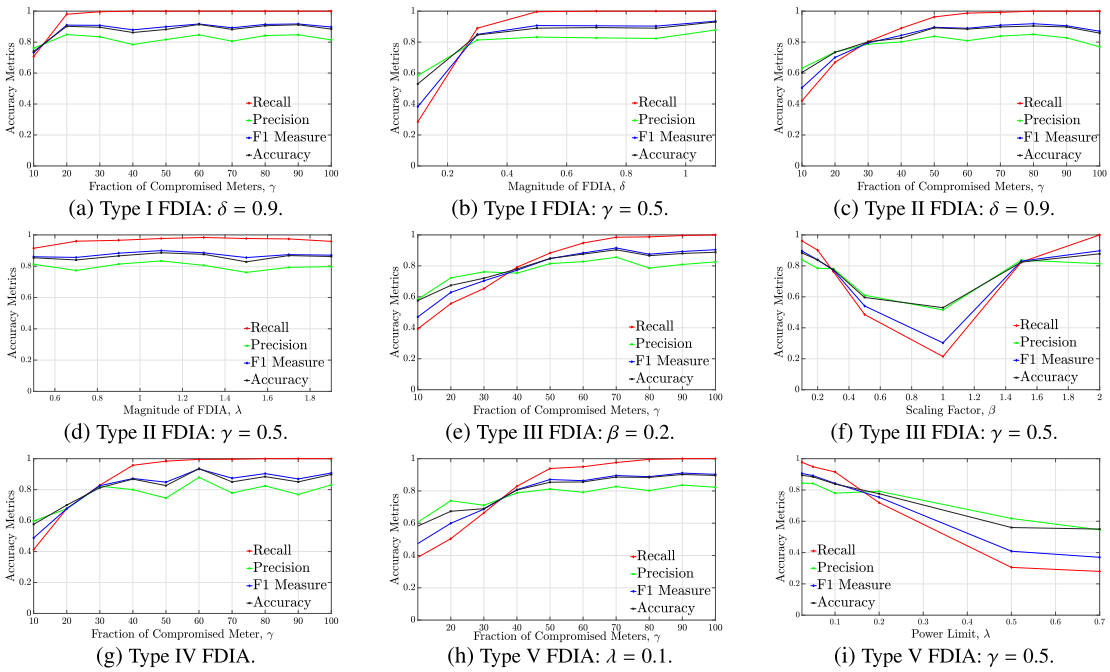
4.1. Data description and experiment details

Data from 12,571 smart meters with a temporal resolution of 30 minutes for one month have been considered to analyze the effectiveness of the proposed framework. These data are collected from 565 9-digit ZIP codes within six 5-digit ZIP codes under ComEd [39]. Of the total 1,440 time instances of data, 70% were used for training, 10% for validation, and 20% for testing. This insufficiency of temporal data for training poses additional challenges in FDIA detection. A detailed discussion of the effect of data availability is provided in Section 5.

For analyzing the detection performance of MOVSTAT-GAT, 1000 scenarios per 9-digit ZIP code (or 5-digit ZIP code) for each type of attack are generated. The scenarios include attack (FDIA) scenarios or non-attack scenarios with equal probabilities. The starting time of the FDIAs is selected from all possible time instances from the test data set with uniform probability.

4.2. Model training

Most model parameters are fixed for detection at both geographic levels and ZIP codes. The dimension, d of the embedding vector, $\mathbf{v_k}$ is set to 48, and the linear process considers 6 past samples (w). $\mathscr{P}$ is implemented with a fully connected neural network with two layers of 256 hidden units each. The learning rate is 0.001, batch size is 64, and training runs for a maximum of 500 epochs. Sensitivity to graph



(a) Type I FDIA: $\delta = 0.9$. (b) Type I FDIA: $\gamma = 0.5$. (c) Type II FDIA: $\delta = 0.9$.

(d) Type II FDIA: $\gamma = 0.5$. (e) Type III FDIA: $\beta = 0.2$. (f) Type III FDIA: $\gamma = 0.5$.

(g) Type IV FDIA. (h) Type V FDIA: $\lambda = 0.1$. (i) Type V FDIA: $\gamma = 0.5$.

Fig. 4. Detection performance at one 9-digit ZIP code using MOVSTAT-GAT for different types of attacks, and variation of performance over different attack parameters.

sparsity (K), moving statistics length (t_d), and detection threshold (θ) are discussed in the following subsections. The training time of the model depends on the size and sparsity of the graph and the geographic hierarchy of detection. For example, training at a 9-digit ZIP code with around 25 households requires 11.05 min for 500 epochs with $K = 10$ whereas at a 5-digit ZIP code with around 60, a 9-digit ZIP code requires 6.2 min on an NVIDIA RTX 4090 GPU. A more detailed discussion of the parameter dependency of the training time is discussed in Section 5.3.2.

4.3. Performance evaluation matrix

For evaluating detection performance, we use recall (true positive rate), precision (1-false positive rate), accuracy (mean of recall and precision), and F1 score (harmonic mean of recall and precision). The choice of metric depends on the implementation scenario. Typically, recall is prioritized over precision in cases like power outages or meter DoS, even if it increases false positives. We present the ROC curve to illustrate the trade-off between higher detection rates and lower false positives through adjustments to the detection threshold, θ. Location accuracy for clustered attacks is defined as correctly locating any component of the cluster, whether at the smart meter level for 9-digit ZIP codes or 9-digit ZIP codes at the 5-digit level.

4.4. Detection performance at 9-digit ZIP code level

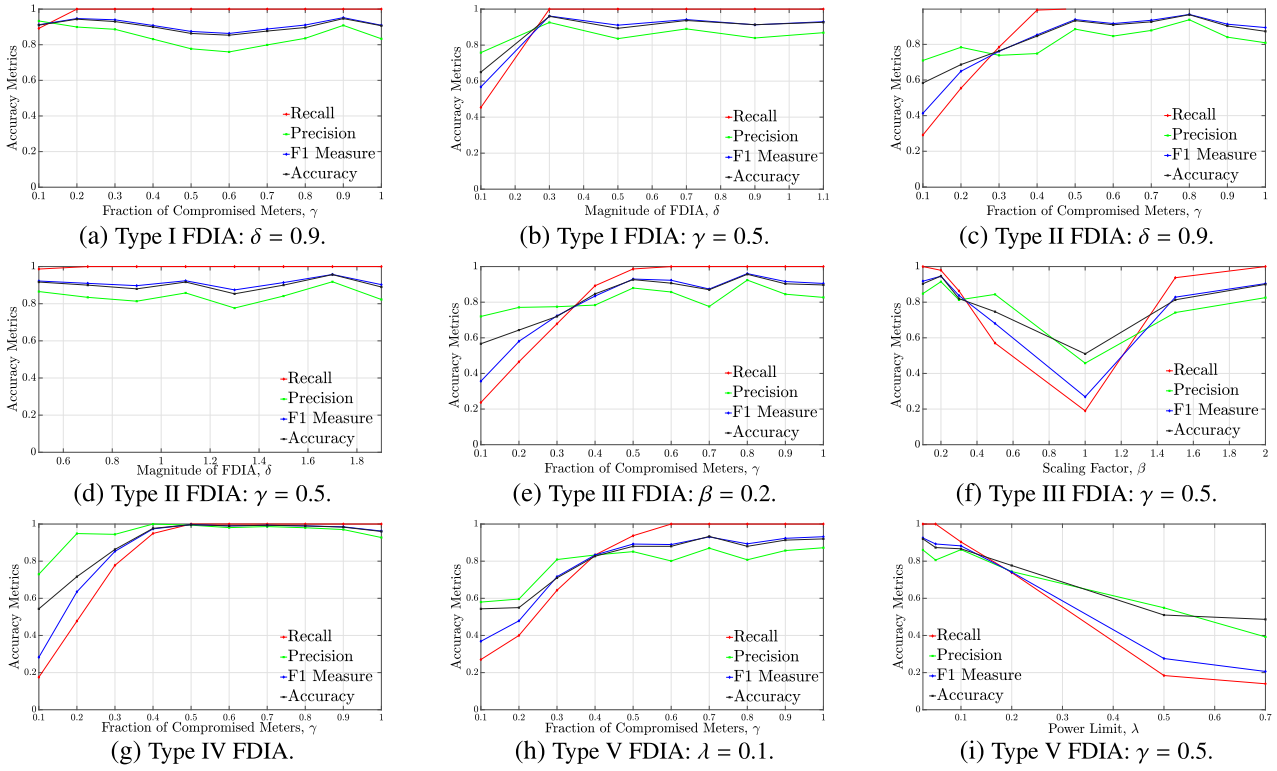
Fig. 4 represents the performance of the detection of the five types of FDIA in a single 9-digit ZIP code. Our extensive simulations show that for most of the 9-digit ZIP codes, the obtained performance of detection is similar, therefore the presented results are representative of detection performance at each 9-digit ZIP code. For each of the five types of attacks, the accuracy metrics are evaluated against the fraction (γ) of total compromised smart meters, and other attack parameters related to that specific type of FDIA. From the detection perspective, Type I and II are the simplest types of FDIAs that can be easily detected for larger changes in values (i.e, large δ, $|\beta| \gg 1$, and $|\beta| \ll 1$ and becomes challenging to detect for smaller changes in values (i.e., small δ and $|\beta - 1| \gg 1$, and $|\beta| \ll 1$. For Type IV attack, which introduces zero (0) values in the smart meter readings, the challenge of accurate detection lies in the fact that a reading of zero (0) power consumption for a while, can be mistaken for typical behavior of consumers not using electricity because of their absence in the household. Therefore, Type IV FDIA is challenging to detect when a small fraction of meters are compromised. Type II attacks show less sensitivity to the magnitude of false data, δ, however, very small and very large values of δ pose more challenges for detection. A small value of δ creates negligible changes in the corresponding time series value similar to the Type I FDIA and reduces detection rate (recall) while a larger value of δ, unlike Type I FDIA, introduces many zero (0) values in the meter reading which presents the same challenges as Type IV attacks. For Type V FDIA which limits the apparent power consumption in the meter, the detection accuracy lowers with the limit increase. For all five types of FDIAs, the recall makes the accuracy of detection increase with the fraction of compromised meters, γ as expected while the false positive remains the same.

4.5. Detection performance at 5-digit ZIP code level

The detection accuracy at the 5-digit ZIP code level for the same FDIAs is shown in Fig. 5. The performance variation with the fraction of compromised smart meters, γ, and the attack parameters, β, δ, and λ are similar to the results at the 9-digit levels. The results in Fig. 5 are for scattered random attacks at smart meters in various 9-digit ZIP codes under the 5-digit ZIP code. For clustered attacks, in which all the compromised meters are from one or more adjacent ZIP codes the detection accuracy is significantly higher. For example, as illustrated in Fig. 6 in clustered Type IV attack, the proposed MOVSTAT-GAT technique achieves a detection accuracy of 96.5%, for $\gamma = 0.0836$ which is below 60% for scattered attacks.

4.6. Ability to localize FDIAs

Once an FDIA is detected using MOVSTAT-GAT, for further investigation of the detected FDIA, the utility operators must identify the location of the FDIA in the system. However, precise detection is challenging for scattered attacks on smart meters located in different ZIP codes. Moreover, location identification in these cases is relatively less important from the monitoring perspective of utilities. On the



(a) Type I FDIA: $\delta = 0.9$. (b) Type I FDIA: $\gamma = 0.5$. (c) Type II FDIA: $\delta = 0.9$.

(d) Type II FDIA: $\gamma = 0.5$. (e) Type III FDIA: $\beta = 0.2$. (f) Type III FDIA: $\gamma = 0.5$.

(g) Type IV FDIA. (h) Type V FDIA: $\lambda = 0.1$. (i) Type V FDIA: $\gamma = 0.5$.

Fig. 5. Detection Performance at one 5-digit ZIP code using MOVSTAT-GAT for different types of attacks, and variation of performance over different parameters.

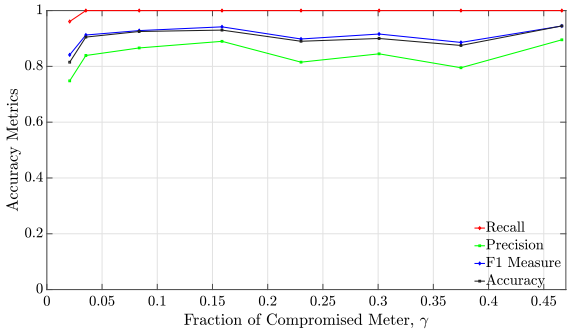


Fig. 6. Clustered FDIA: Detection performance accuracy in clustered FDIA is significantly higher than the scattered FDIAs. Results show detection performance at the 5-digit ZIP code level for Type IV attacks.

other hand, clustered attacks are localized among several nearby 9-digit ZIP codes, and location identification in those cases is important from the system monitoring perspective. For example, Type IV attacks in several adjacent 9-digit ZIP codes can indicate a power outage in a geographic location which the utility should address. Fig. 7 illustrates monitoring of estimated normalized absolute error, $\hat{e}(n,t)$ as an image at (a) 9-digit ZIP code level and (b) 5-digit ZIP code levels. The clustered FDIAs can be identified from the localized pattern of higher intensity pixels during the attack location along smart meter # 5 in Fig. 7(a) and along ZIP code # 13-14 in Fig. 7(b) by observing estimated normalized absolute error, $\hat{e}(n,t)$ as images. Visual inspection of this image just after the detection of the FDIA can provide important information about the FDIA location. Here, the proposed technique suggests automated identification of any member of the cluster using Eq. (6). The automated location accuracy for different types of FDIA at the 5-digit ZIP code level is shown in Fig. 7(c). The results show that the proposed automated location identification technique is capable of locating power outages (Type 4 FDIA) even at a single 9-digit ZIP code with good accuracy. Expectedly, the accuracy of locating at the household level is challenging. For example, the locating accuracy of Type I FDIA at a single smart meter with $\delta = 1.1$ is 69.4%.

4.7. Performance sensitivity to various factors

4.7.1. Selection of moving statistics

In this work, the effectiveness of several moving temporal statistics for transforming the time series in the MOVSTAT technique has been tested, including arithmetic mean (AM), harmonic mean (HM), the ratio of harmonic and arithmetic mean $\left(\frac{HM}{AM}\right)$, and standard deviation (SD). As discussed in Section 2, although anomalies are better reflected on HM or $\frac{HM}{AM}$, in conjunction with our training-based model, they are not very effective features due to the large number of peaks associated with behavioral anomalies in the training time series data (Fig. 8(a)). Both AM and SD as moving statistic features provide lower training error and better accuracy in FDIA detection. However, our experiments show SD provides slightly better detection accuracy than AM and due to the temporal smoothing effect, AM involves slightly greater detection delay. Considering all these facts, the standard deviation within a temporal window of 13 samples (computed by $\mathscr{F}_{13}^{SD}$ operator) is chosen as the preferred time-varying moving statistics feature. Therefore, the presented results are generated using $\mathscr{F}_{13}^{SD}$ unless mentioned otherwise.

4.7.2. Threshold for detection, θ

The detection performance, characterized by the trade-off between the recall and precision (i.e., true positive rate and false positive rate) is governed by the choice of θ. Fig. 8(b) illustrates the receiver–operator characteristics (ROC) curves associated with the detection of different types of attacks at a particular 9-digit ZIP code, where the θ has been varied within the range $100 \leq \theta \leq 1000$ to demonstrate the recall–precision trade-off. The choice of θ at each 9-digit ZIP code level or 5-digit ZIP code level can be made by the operators in the utility depending on their resources, policy, and associated infrastructures.

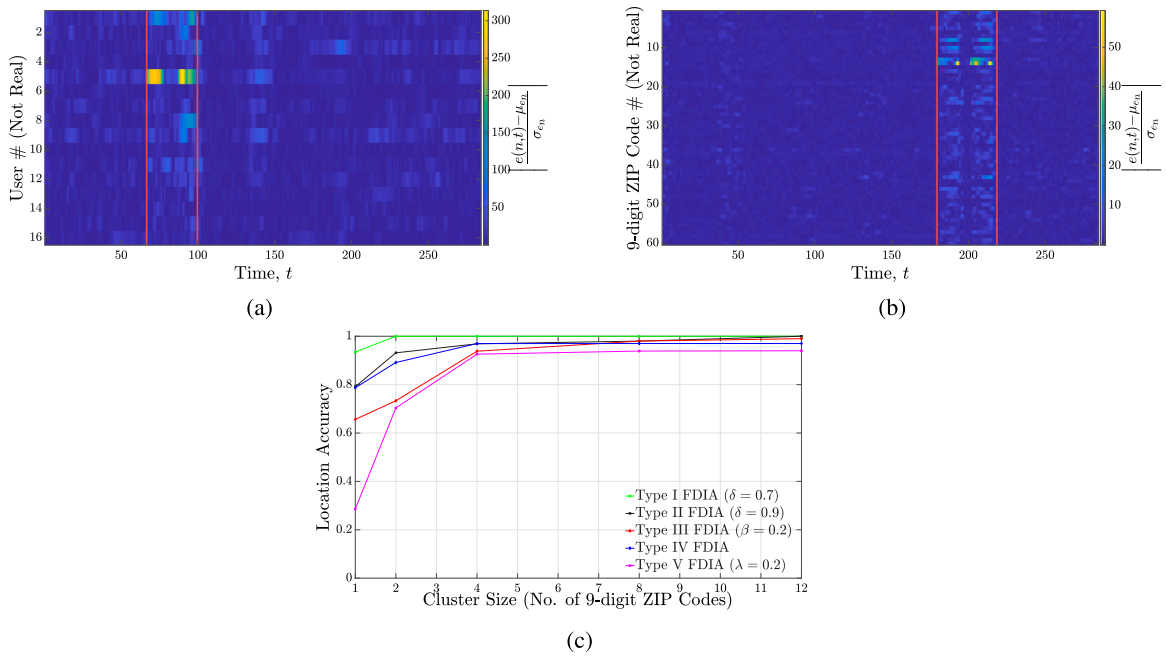


Fig. 7. Localization: (a)–(b) Location identification using visualization tool: (a) compromised smart meter # 5 within a 9-digit ZIP code, (b) ZIP code # 13-14 where the compromised smart meters are located within a 5-digit ZIP code. The vertical pair of red lines indicate the attack duration. (c) Automated location performance of clustered FDIA after detection at a 5-digit level.

Usually, in situations where human operators are engaged in monitoring the detection process, or where automated decisions are not promptly executed in response to real-time FDIA detection, the choice of θ is made in a way to emphasize the recall rather than the precision. A more detailed discussion on the choice of threshold, θ, in the power utility context is presented in Section 5.3.1.

4.7.3. Sensitivity to MOVSTAT filter length

Fig. 8(c) and (d) illustrates the detection performance sensitivity to the length of the moving statistics filter, t_d. A larger t_d facilitates the detection of attacks at a small number of meters, i.e., small γ, at the cost of precision and hardware implementation cost. The strategies of selecting t_d in the industry setting at discussed in Section 5.3.3.

4.8. Comparison

To compare the performance of the proposed technique, two benchmark methods have been considered. The first one involves using an auto-encoder-based prediction of multi-variate smart meter data to detect FDIA in an unsupervised manner. Since the detection of FDIAs with regular intensity and spreads is easier to detect from the aggregated data, a univariate LSTM-based detection method has been considered as the second benchmark technique. The reason for selecting these two methods as benchmarks is that, auto-encoder is a state-of-the-art baseline structure being extensively used in the recent literature for anomaly detection, whereas LSTM has been considered the most widely used technique to handle sequence and time-series data to date.

Our extensive experiments show that the performance of MOVSTAT-GAT, auto-encoder, and LSTM-based techniques are comparable for wide-spread attacks (larger γ); however, the main advantage of the proposed MOVSTAT-GAT lies in detecting FDIAs for very small values of γ, even at a single household or a single 9-digit ZIP code, depending on the geographic hierarchy of detection. In particular, detecting FDIAs in a very small number of households clustered in a geographical locality from the 5-digit ZIP code level can be very crucial from the utility point of view to identify coordinated energy theft, meter DoS, or power outages in a certain geographic area. Fig. 9(a) illustrates the effectiveness of the proposed MOVSTAT-GAT technique over the benchmarks at a 9-digit ZIP code level in case of Type I attack ($\delta = 2, \gamma = 0.0417$) in a single household. MOVSTAT-GAT achieves a superior balance between the true positive rate and the false positive rate compared to benchmark techniques. Fig. 9(b) shows the out-performance of MOVSTAT-GAT over the benchmarks for detecting Type IV attack at the households of one single 9-digit ZIP code out of 60 within a certain 5-digit ZIP code (i.e., $\gamma = 0.0260$). The second example can be a representation of a power outage in a certain locality, which is very important for detection from a higher geographic level. Due to the high variance of detection rate in this range of attack intensity creates some non-monotony in the ROC curves for all the methods, although the out-performance of the proposed MOVSTAT-GAT is evident from the trend.

It must be noted that, in addition to detection, the proposed MOVSTAT-GAT technique can locate the attack after detection with good accuracy, as shown in Fig. 7(c). Moreover, the representation of $\hat{e}(n,t)$ as an image provides a visualization tool to the utility operators to identify the localization behavior of the FDIAs. In addition, subtle changes in the intensity pattern of, $\hat{e}$ make an expert operator aware of even smaller magnitude attacks classified as false negatives by automated detection based on the threshold.

5. Practical applicability

One of the key aspects of our proposed method is its adaptability for real-world implementation. Here, we discuss the MOVSTAT-GAT implementation strategies, focusing on detecting and locating various FDIAs at different hierarchical levels. We emphasize adjusting model parameters and thresholds to optimize performance while considering the potential human intervention by utility operators.

5.1. Visualization tool for utility operators

The visualization tool, shown in Fig. 7, complements our automated detecting and locating mechanism, aiding utility operators in service monitoring through visualization [40]. Continuous sliding frames of images, like those in Fig. 7(a) and (b), help operators identify changes in pixel intensity related to specific smart meters or 9-digit ZIP codes,

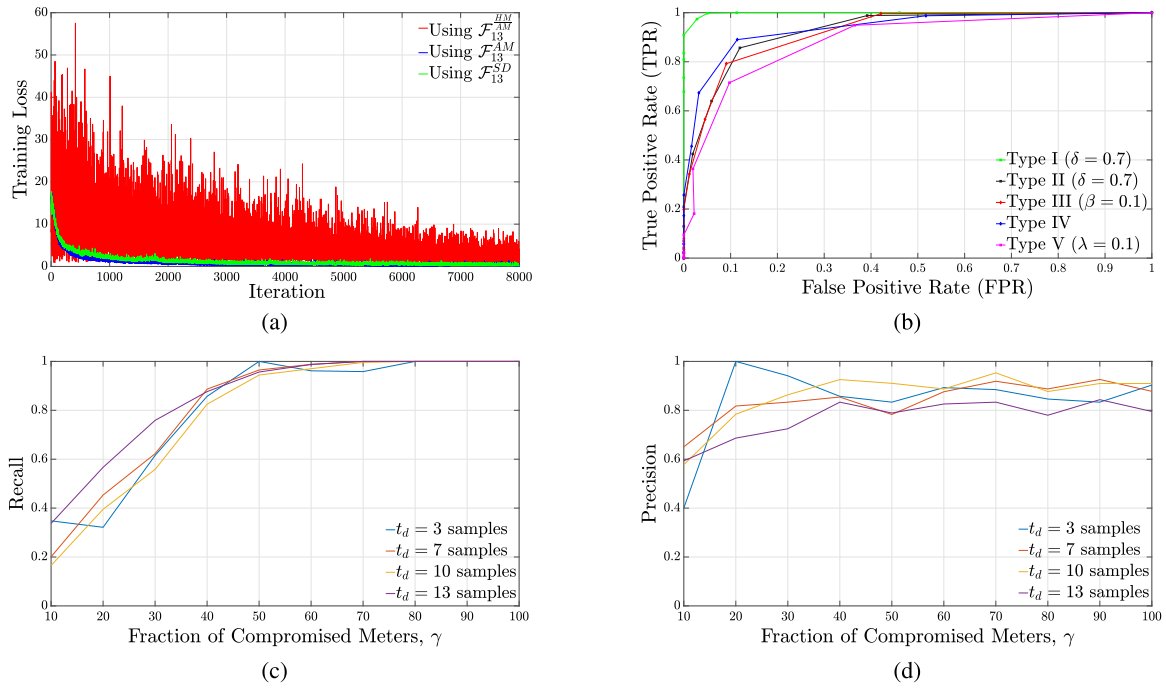


Fig. 8. Sensitivity Analysis: (a) Training with different moving statistics. (b) ROC curve represents the dependency of detection performance on the likelihood threshold for different types of FDIAs, $\gamma = 0.5$ for all attacks. (c)–(d): Sensitivity to moving statistics filter length at 9-digit ZIP code levels, t_d: (c) recall, (d) precision.

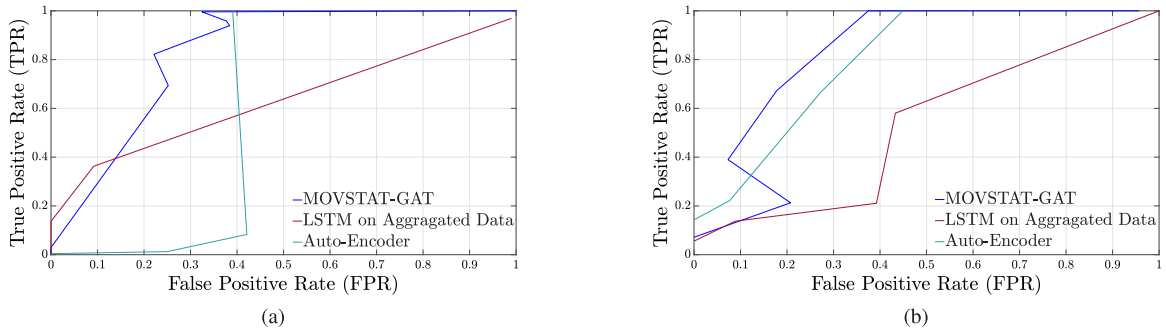


Fig. 9. Comparison of detection performance using ROC curves: (a) detection performance at 9-digit ZIP code for Type I attack at one single household ($\gamma = 0.041$), (b) detection performance of Type IV attack at 5-digit ZIP code for clustered attack ($\gamma = 0.017$).

indicating potential FDIAs at those locations. Integrating this tool with the automated system offers several advantages. For small γ, i.e., attacks targeting a small number of smart meters, the system may have a high false negative rate (low recall). While small-scale attacks might be ignored due to minimal financial loss, they can be crucial in scenarios like localized outages or in understanding attacker strategies. The visualization tool helps operators identify these false negatives. Operators can also reduce false positives by analyzing the pixel intensity, location, and duration of detected FDIAs, refining the detection system's accuracy. Additionally, operators with detailed geographical knowledge can supplement the automated system, especially during multiple attacks or complex interconnectivity, to precisely locate attacks and understand the broader context, including attackers' strategies and resources.

5.2. Scope of real-time detection

The MOVSTAT-GAT's automatic detection mechanism excels in real-time FDIA detection, with over 90% of attacks detected instantly or within $1-2$ samples of onset. This makes it ideal for real-time system monitoring and decision-making in industrial environments. It enables prompt responses to attacks, power outages, or meter DoS incidents at specific ZIP codes. Timely detection and precise localization of power outages, especially due to extreme weather or grid issues, are crucial for immediate restoration. Our technique, complemented by the visualization tool, effectively detects power outages in real-time using only household power consumption data, without needing additional grid topology or electrical attribute information.

5.3. Model parameter tuning for utility application

Here we discuss optimal model parameter tuning in realistic scenarios based on the sensitivity analysis in Section 4.7. Proper tuning, aligned with domain knowledge and dataset specifics, can enhance detection and locating performance. The influence of attack parameters and utility company priorities is also considered.

5.3.1. Detection threshold flexibility

The detection threshold, θ, is the only model parameter needing adjustment when transferring the detection mechanism between ZIP codes. Adjusting θ optimizes performance based on utility company objectives and resources. The precision–recall trade-off related to θ is shown in Fig. 8(b) and discussed in Section 4.7.2. In practice, several factors influence θ selection. Utility operators may choose a smaller θ to detect FDIAs, theft, DoS, or outages in small localities, tolerating

some false alarms, or a larger θ to ignore certain attacks based on strategy, resources, or prior knowledge. Operators can also use multiple thresholds with the same model, make decisions on the basis of alarms raised by the different thresholds, the visualization tool described in 5.1, and domain knowledge.

5.3.2. Handling graph sparsity

Graph sparsity, denoted by K, indicates connectivity among graph vertices, reflecting spatial correlations among nodes like smart meters or ZIP codes. In direct smart meter data detection at the 9-digit ZIP code level, K can impact performance, especially for low attack intensities or few targeted smart meters (δ and γ small, λ large, and β close to 1). When using aggregated meter readings at the 9-digit ZIP code level for detection at the 5-digit level, selecting K is more flexible due to adjustments in the detection threshold, θ. This flexibility is due to reduced uncertainty in household-level data when aggregated, leading to stronger correlations with neighboring nodes. Our experiments showed setting K around 10 (8–12) for 9-digit aggregated data provides good performance across scenarios. For direct smart meter data detection, K values between 5 and 10 generally yield optimal results, using household consumption correlations within 9-digit ZIP codes. Setting $K = 10$ universally, with appropriate θ adjustments, yields satisfactory results, as shown in Figs. 4 and 5. Human intervention can enhance performance based on utility company objectives and resources. Minor fluctuations in K do not significantly impact MOVSTAT-GAT performance, as attention weights accommodate these adjustments.

5.3.3. Flexibility in choice of moving statistics filter length and other parameters

The MOVSTAT filter length, t_d, impacts the detection of low-intensity or small-spread attacks, as shown in Fig. 8(c)–(d). Very small t_d values (e.g., $t_d = 3$) result in low detection rates for low-intensity attacks, while very large t_d values can slightly affect precision during large attacks and increase hardware implementation costs. Our experiments suggest setting t_d between 7 and 13 samples to balance performance and practicality, ensuring effective detection across various scenarios avoiding false positives related to anomalous consumer behaviors. For other remaining model parameters, it is advisable to maintain consistent values across all ZIP codes, as outlined in Section 4.2. This includes keeping the sizes of the temporal window, $\mathbf{w_l}$, the embedding vector, $\mathbf{v_k}$, and the number of layers and hidden units in neural networks, $\mathbf{W_p}$, fixed. This approach ensures uniformity and simplifies implementation for utility operators across different areas.

5.4. Robustness to new attack types

As an unsupervised learning method, MOVSTAT-GAT learns from the inherent probability distribution of honest data during the training process. Any significant deviation triggers false data detection, making the model more robust to new FDIAs than supervised learning frameworks. Attackers must inject data that closely resembles honest data to bypass detection, posing a trade-off with the attacks' reward. Additionally, being unsupervised, MOVSTAT-GAT naturally avoids challenges related to data imbalance from theft or scarce injected data.

5.5. Scalability and extensibility of MOVSTAT-GAT

The proposed framework, implementable in real-time using smart meter data, scales with the number of spatial variables or graph vertices. We tested its effectiveness at two levels: 9-digit ZIP codes (20–30 vertices representing households) and 5-digit ZIP codes (30 to several hundred vertices). GNNs, due to their inherent sparsity, are more scalable than other neural networks. Thus, the technique is extendable to higher geographic levels (e.g., county, state). For extensions to thousands of vertices, besides adjusting the sparsity parameter, K, using a static graph adjacency matrix can enhance scalability, especially if vertex correlations are not significantly time-varying, which is a common scenario observed at higher geographic levels.

5.6. Limitations

While the proposed MOVSTAT-GAT technique excels in detection performance, scalability across different system sizes and geographic hierarchies, and robustness to various utility system scenarios, it still has a few limitations. First, the technique is not capable of classifying the specific types of FDIAs on smart meters, which could further assist the utility operators in understanding the attackers' motives and strategies. Moreover, since the smart meters in our study provide only power consumption readings, detection of events (e.g., various types of faults, power quality, and inverter-related events) that are dependent on the measurement of voltage or other electrical attributes, are not detectable under this framework. Finally, since the current model is trained using data for a single month, it cannot capture the seasonal behavior of household power consumption. This can be addressed by incorporating multi-year data that would enhance model performance.

6. Conclusions and future works

This work proposes MOVSTAT-GAT, a spatio-temporal graph attention network-based unsupervised technique to analyze smart meter data at different geographic hierarchical levels to detect and locate energy theft, power outage, meter DoS, and meter-reading alterations within a general FDIA framework. MOVSTAT-GAT achieves excellent detection and locating performance for regular-intensity attacks. However, by leveraging the advantage of the moving-statistics filter and the attention on the dynamic graph to emphasize the changes during the attack onset, its outperformance compared to baseline state-of-the-art techniques is more significant for small-intensity and localized attacks. Specifically, the experiments highlight the effectiveness of MOVSTAT-GAT for detecting very localized attacks, where the fraction of compromised meters is as small as 1% of the total meters. This enables the operators to address disruptions or adversaries at a certain geographic location within the utility service area. The comprehensive experiments evaluate the effectiveness of the proposed technique under different conditions, considering performance sensitivity to various parameters and factors relevant to real-world industry implementation.

There are a few directions in which the proposed framework can be extended by overcoming its limitations. Future work can expand MOVSTAT-GAT to classify different types of FDIAs (e.g., energy theft, meter-reading altering, power outage, meter DoS, etc.), aiding utility operators to distinguish between the issues and understand attackers' strategies. Additionally, in case of the availability of voltage measurement data from smart meters, future work can incorporate other distribution system service interruptions such as faults, voltage stability, and power quality issues. Finally, training the model using multi-year power consumption data would enable the model to capture the seasonal patterns within the data to enhance model performance and robustness.

CRediT authorship contribution statement

Md Abul Hasnat: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Harsh Anand:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis, Data curation. **Mazdak Tootkaboni:** Writing – review & editing, Validation, Supervision, Investigation, Funding acquisition. **Negin Alemazkoor:** Writing – review & editing, Validation, Supervision, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Negin Alemazkoor and Mazdak Tootkaboni reports financial support

was provided by National Science Foundation. Negin Alemazkoor and Mazdak Tootkaboni reports financial support was provided by Office of Naval Research. If there are other authors they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

NA and MT acknowledge financial support from the Office of Naval Research, United States under grant N00014-22-1-2012, and the National Science Foundation, United States under grants CMMI-1826155.

References

[1] Y. Wang, Q. Chen, T. Hong, C. Kang, Review of smart meter data analytics: Applications, methodologies, and challenges, IEEE Trans. Smart Grid 10 (3) (2018) 3125–3148.

[2] S. Mitra, B. Chakraborty, P. Mitra, Smart meter data analytics applications for secure, reliable and robust grid system: Survey and future directions, Energy 289 (2024) 129920, http://dx.doi.org/10.1016/j.energy.2023.129920.

[3] A. Hoogsteyn, M. Vanin, A. Koirala, D. Van Hertem, Low voltage customer phase identification methods based on smart meter data, Electr. Power Syst. Res. 212 (2022) 108524, http://dx.doi.org/10.1016/j.epsr.2022.108524.

[4] S. Bhattacharjee, S.K. Das, Detection and forensics against stealthy data falsification in smart metering infrastructure, IEEE Trans. Dependable Secure Comput. 18 (1) (2018) 356–371.

[5] Z. Soltani, M. Khorsand, Real-time topology detection and state estimation in distribution systems using micro-pmu and smart meter data, IEEE Syst. J. 16 (3) (2022) 3554–3565.

[6] M.A. Ravaglio, L.F.R. Toledo, S.L. Santos, L.R. Gamboa, D.B. Dahlke, J.A. Teixeira, E.T. Yano, A.P. Silva, O. Kim, M.G. Antunes, Detection and location of high impedance faults in delta 13.8kV distribution networks, Electr. Power Syst. Res. 230 (2024) 110291, http://dx.doi.org/10.1016/j.epsr.2024.110291.

[7] S. Nandkeolyar, P.K. Ray, Identifying households with electrical vehicle for demand response participation, Electr. Power Syst. Res. 208 (2022) 107909, http://dx.doi.org/10.1016/j.epsr.2022.107909.

[8] H. Anand, M. Darayi, Power network component vulnerability analysis: A machine learning approach, Procedia Comput. Sci. 185 (2021) 73–80.

[9] H. Anand, M. Darayi, A probabilistic approach to modeling power network component importance considering economic impacts, in: IIE Annual Conf. Proceedings, Institute of Industrial and Systems Engineers (IISE), 2021, pp. 1010–1015.

[10] Y. Li, X. Wei, Y. Li, Z. Dong, M. Shahidehpour, Detection of false data injection attacks in smart grid: A secure federated deep learning approach, IEEE Trans. Smart Grid 13 (6) (2022) 4862–4872.

[11] M. Higgins, B. Stephen, D. Wallom, Incentive-weighted anomaly detection for false data injection attacks against smart meter load profiles, 2023, arXiv preprint arXiv:2301.10628.

[12] A. Takiddin, M. Ismail, U. Zafar, E. Serpedin, Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids, IEEE Syst. J. 16 (3) (2022) 4106–4117.

[13] F. Ünal, A. Almalaq, S. Ekici, P. Glauner, Big data-driven detection of false data injection attacks in smart meters, IEEE Access 9 (2021) 144313–144326.

[14] Z. Du, Z. Yan, Y. Xu, A dimensional augmentation-based data-driven method for detecting false data injection in smart meters, IEEE Trans. Smart Grid (2023).

[15] H. Anand, R. Nateghi, N. Alemazkoor, Bottom-up forecasting: Applications and limitations in load forecasting using smart-meter data, Data-Centric Eng. 4 (2023) e14.

[16] T. Ahmad, H. Chen, J. Wang, Y. Guo, Review of various modeling techniques for the detection of electricity theft in smart grid environment, Renew. Sustain. Energy Rev. 82 (2018) 2916–2933.

[17] M. Ahmed, A. Khan, M. Ahmed, M. Tahir, G. Jeon, G. Fortino, F. Piccialli, Energy theft detection in smart grids: Taxonomy, comparative analysis, challenges, and future research directions, IEEE/CAA J. Autom. Sin. 9 (4) (2022) 578–600, http://dx.doi.org/10.1109/JAS.2022.105404.

[18] E. Stracqualursi, A. Rosato, G. Di Lorenzo, M. Panella, R. Araneo, Systematic review of energy theft practices and autonomous detection through artificial intelligence methods, Renew. Sustain. Energy Rev. 184 (2023) 113544, http://dx.doi.org/10.1016/j.rser.2023.113544.

[19] M. Tariq, H.V. Poor, Electricity theft detection and localization in grid-tied microgrids, IEEE Trans. Smart Grid 9 (3) (2018) 1920–1929, http://dx.doi.org/10.1109/TSG.2016.2602660.

[20] D.N. Nikovski, Z. Wang, Method for detecting power theft in a power distribution system, 2014, US 13/770, 460.

[21] A. Althobaiti, A. Jindal, A.K. Marnerides, U. Roedig, Energy theft in smart grids: A survey on data-driven attack strategies and detection methods, IEEE Access 9 (2021) 159291–159312, http://dx.doi.org/10.1109/ACCESS.2021.3131220.

[22] K. Zheng, Q. Chen, Y. Wang, C. Kang, Q. Xia, A novel combined data-driven approach for electricity theft detection, IEEE Trans. Ind. Inform. 15 (3) (2018) 1809–1819.

[23] R.K. Ahir, B. Chakraborty, Pattern-based and context-aware electricity theft detection in smart grid, Sustain. Energy Grids Netw. 32 (2022) 100833.

[24] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, Y. Zhou, Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids, IEEE Trans. Ind. Inform. 14 (4) (2017) 1606–1615.

[25] Y. Gao, B. Foggo, N. Yu, A physically inspired data-driven model for electricity theft detection with smart meter data, IEEE Trans. Ind. Inform. 15 (9) (2019) 5076–5088.

[26] J. Tursunboev, V. Palakonda, J.-M. Kang, Multi-objective evolutionary hybrid deep learning for energy theft detection, Appl. Energy 363 (2024) 122847, http://dx.doi.org/10.1016/j.apenergy.2024.122847.

[27] W. Zhang, Y. Dai, A multiscale electricity theft detection model based on feature engineering, Big Data Res. (2024) 100457, http://dx.doi.org/10.1016/j.bdr.2024.100457.

[28] B.L.H. Nguyen, T.V. Vu, T.-T. Nguyen, M. Panwar, R. Hovsapian, Spatial-temporal recurrent graph neural networks for fault diagnostics in power distribution systems, IEEE Access 11 (2023) 46039–46050, http://dx.doi.org/10.1109/ACCESS.2023.3273292.

[29] S.H. Haghshenas, M.A. Hasnat, M. Naeini, A temporal graph neural network for cyber attack detection and localization in smart grids, in: 2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, ISGT, 2023, pp. 1–5, http://dx.doi.org/10.1109/ISGT51731.2023.10066446.

[30] Z. Qu, Y. Dong, Y. Li, S. Song, T. Jiang, M. Li, Q. Wang, L. Wang, X. Bo, J. Zang, et al., Localization of dummy data injection attacks in power systems considering incomplete topological information: A spatio-temporal graph wavelet convolutional neural network approach, Appl. Energy 360 (2024) 122736.

[31] H. Nguyen, R. Kashef, TS-IDS: Traffic-aware self-supervised learning for IoT network intrusion detection, Knowl.-Based Syst. 279 (2023) 110966, http://dx.doi.org/10.1016/j.knosys.2023.110966.

[32] A. Rahmani, A. Venkitaraman, P. Frossard, A meta-gnn approach to personalized seizure detection and classification, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2023, pp. 1–5.

[33] J. Chen, Q. Chen, F. Jiang, X. Guo, K. Sha, Y. Wang, SCN_gnn: A GNN-based fraud detection algorithm combining strong node and graph topology information, Expert Syst. Appl. 237 (2024) 121643.

[34] H.T. Phan, N.T. Nguyen, D. Hwang, Fake news detection: A survey of graph neural network methods, Appl. Soft Comput. (2023) 110235.

[35] W. Zhuang, W. Jiang, M. Xia, J. Liu, Dynamic generative residual graph convolutional neural networks for electricity theft detection, IEEE Access 12 (2024) 42737–42750, http://dx.doi.org/10.1109/ACCESS.2024.3379201.

[36] W. Liao, R. Zhu, Z. Yang, K. Liu, B. Zhang, S. Zhu, B. Feng, Electricity theft detection using dynamic graph construction and graph attention network, IEEE Trans. Indus. Inform. 20 (4) (2024) 5074–5086, http://dx.doi.org/10.1109/TII.2023.3331131.

[37] A. Deng, B. Hooi, Graph neural network-based anomaly detection in multivariate time series, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 5, 2021, pp. 4027–4035, http://dx.doi.org/10.1609/aaai.v35i5.16523.

[38] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, 2015, arXiv preprint arXiv:1505.00853.

[39] Commonwealth Edison Company, ComEd - An Exelon Company, 2024, https://www.comed.com. (Accessed 14 March 2024).

[40] POWER Magazine, How utilities can achieve automated condition-based grid maintenance to boost safety and grid reliability, 2024, POWER Magazine.