FDIA Attack Detection Technique for Smart Grids Based on Graph Reconstruction and Spatio-Temporal Joint Modeling

Yuzhang Gao[1,a], Jing Xia[2*]

[1]Beijing Jiaotong University Beijing China
[2]Goldwind Science &Technology Co., Ltd., Beijing China

Abstract: With the widespread application of smart grids, the false data injection attack (FDIA) has become a major threat to power grid security. Traditional detection methods often have difficulty in effectively identifying such attacks, especially in complex environments. To address this problem, this paper proposes a spatial-temporal joint FDIA detection framework named HSGT-Net, which integrates the Hodge diffusion, Sparge sparsity mechanism, and time series modeling. This framework unifies the branch power and bus power data through a graph reconstruction mechanism, and combines Hodge multi-order diffusion with spatial feature modeling of Sparge-iTransformer, so it can effectively capture the local and global dependencies of the power grid. A GRU module is formulated to model time series features, with improved perception ability of dynamic attack features.

Experimental results show that the HSGT-Net performs better than traditional machine learning methods (such as SVM and KNN) and deep learning models (such as MLP, CNN, and GNN+LSTM) on both IEEE-30 and IEEE-57 standard test systems. Especially in large-scale power system scenarios, it can still maintain high accuracy and robustness. In addition, ablation experiments further verify the effectiveness of each module and prove the key role of joint modeling on spatial and temporal features in FDIA detection. In a word, the HSGT-Net demonstrates strong adaptability and computational efficiency in FDIA attack detection and has good engineering application prospects.

1. Introduction

In recent years, with the deep integration of energy Internet and new generation information as well as communication technology (ICT), the power system is constantly moving towards intelligent and digital transformation [1]. Smart grids improve resource allocation and operational efficiency through distributed energy integration, real-time monitoring, and flexible scheduling. Compared to traditional grids, they enable tight integration of physical infrastructure with communication, information processing, and data analytics, forming an interactive cyber-physical system [2]. However, Smart grids face network security threats, especially in state estimation and data collection. False data injection attacks (FDIAs) can bypass residual-based detection, distort state estimation, and lead to abnormal power flow and imbalance, posing serious risks to grid security [3].

Many FDIA detection methods have been proposed, ranging from early residual analysis and power flow-based approaches to machine learning models like SVM [4]. With the development of deep learning, CNNs have been used for spatial feature extraction [5], and GNN-LSTM models have been introduced for joint spatial–temporal learning [6]. However, most existing methods still focus on static spatial features and lack modeling of dynamic temporal evolution.

Generally, existing FDIA detection studies face several challenges: (1) Inconsistent modeling between branch flow and bus injection leads to weak feature fusion; (2) GNNs struggle to capture long-range or high-order spatial dependencies; (3) Temporal modeling is limited, and spatial-temporal joint learning remains insufficient; (4) Some deep models are computationally intensive, limiting scalability. To address these problems, this paper proposes HSGT-Net, a multi-module framework integrating Hodge diffusion, sparse mechanisms, and time series modeling to enhance detection accuracy and efficiency for large-scale applications.

The main work and innovations of this paper are as follows:

1) A unified graph reconstruction mechanism is proposed to achieve the fusion input of branch power and bus power data, ensuring the integrity and physical rationality of feature expression.

2) A spatial feature modeling method that integrates Hodge multi-order diffusion and Sparge-iTransformer is constructed, which can effectively capture the local and long-range spatial dependencies and reduce the computational complexity with sparse mechanisms.

[a]22120205@bjtu.edu.cn
[*] Corresponding author, xiajingbeijing@126.com

3) The GRU time series are used to achieve dynamic evolution modeling of spatial features, with improved dynamic perception of FDIA attack features.

4) Multiple experimental verifications are performed on the IEEE-30 bus and IEEE-57 bus standard test systems. The results show that the proposed HSGT-Net is superior to the typical methods in terms of detection accuracy, robustness and computational efficiency.

2. Model design

The HSGT-Net architecture consists of graph reconstruction, spatial feature extraction, temporal modeling, and output prediction, as illustrated in Figure 1. It first unifies branch and bus power data in the edge space. Hodge diffusion and Sparge-iTransformer capture local and global spatial features, while GRU models temporal dynamics. A fully connected layer then outputs the FDIA detection score.

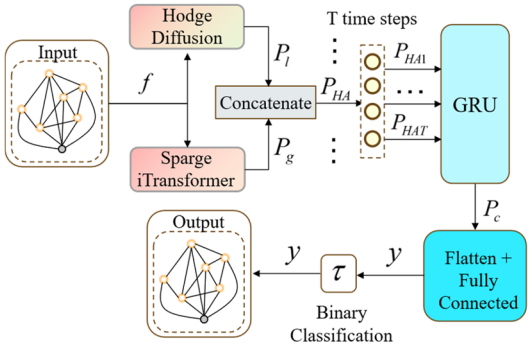

Figure 1. HSGT-Net Architecture for FDIA Detection.

2.1. Graph reconstruction

In FDIA detection, branch flow and bus injection power correspond to edge-space and node-space features. To unify them, a graph reconstruction mechanism is proposed, which introduces a super virtual node connected to all bus nodes via virtual edges. This forms a reconstructed graph $G_e = (V_e, E_e)$, where $V_e = V \cup \{N+1\}$, and $E_e = E \cup E_{\text{vir}}$, with $|E_{\text{vir}}| = N$. This design ensures spatial consistency between branch and bus power data, providing an integrated input feature f for subsequent modeling.

The input feature f combines branch flow power and bus injection power and ensures consistency with the original measurement data z used in state estimation. Ultimately, feature f serves as a unified and physically consistent foundation for the subsequent spatial feature extraction and temporal modeling.

2.2. Local spatial features

In the spatial feature extraction process, the Hodge diffusion [7] is applied to capture the local spatial dependencies in the power grid. This technique allows features to propagate through the graph based on the topology, which enables the model to account for the physical constraints of grid. The feature vector f is propagated through the graph using the diffusion operator L. The core operation of the Hodge diffusion process is as follows:

$$P_l = [f, Lf, L^2 f, \ldots, L^{E_e-1} f] \tag{1}$$

where f is the initial feature vector, L is the graph Laplacian operator that represents the graph's structure, E_e is the maximum diffusion step, and P_s is the set of spatial features obtained after applying multiple steps of diffusion.

The output feature P_l contains the multi-scale spatial information, which is then used as input for the subsequent temporal modeling stage.

2.3. Global spatial features

To address the limitations of local diffusion mechanisms in modeling long-range spatial dependencies, this paper proposes a Sparge-iTransformer module that integrates a dynamic sparsity mechanism with efficient attention operations. This module enables the global spatial modeling on power grid graphs with high accuracy and low computational overhead. The overall framework is illustrated in Figure 2.

The input edge-space features f are first linearly projected into the query, key, and value matrices:

$$Q = fW_Q, \quad K = fW_K, \quad V = fW_V \tag{2}$$

where W_Q, W_K, and W_V are learnable projection matrices.

Then, the mean pooling is applied to Q and K to obtain compressed feature representations $\overline{q_i}$ and $\overline{k_j}$, followed by the computation of initial relevance scores between edges:

$$a_{ij} = \overline{q_i} \times \overline{k_j} \tag{3}$$

To measure the information richness of each edge, the self-cosine similarity is further computed:

$$s_{self}^{q} = \text{cosine}(q_i, \overline{q}), \quad s_{self}^{k} = \text{cosine}(k_j, \overline{k}) \tag{4}$$

where $\overline{q}$ is the mean of the global query feature, $\overline{k}$ is the mean of the global key feature.

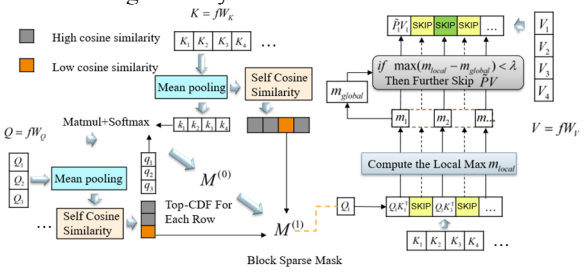

Figure 2. Sparge mechanism principle.

Based on the computed scores, a preliminary sparse mask is constructed using the Top-CDF strategy along with low self-similarity filtering [8]:

$$M_{ij}^{(1)} = \begin{cases} 1, & \text{if Top-CDF or low self-cosine similarity} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

To dynamically adjust for varying input distributions, a filtering mechanism based on the difference between local

and global maxima is introduced. If the local maximum m_{local} significantly deviates from the global maximum m_{global}, the corresponding entry is retained:

$$M_{ij}^{(2)} = \begin{cases} 1, & m_{local} - m_{global} \geq \lambda \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The final sparse mask is obtained by merging both parts:

$$M_{ij} = M_{ij}^{(1)} \cup M_{ij}^{(2)} \quad (7)$$

Guided by the sparse mask, attention scores are computed only for retained positions:

$$\delta_{ij} = \frac{QK^T}{\sqrt{d}} \times M_{ij} \quad (8)$$

The attention weights are then normalized as:

$$\alpha_{ij} = \frac{\exp(\delta_{ij})}{\sum_{k \in M_i = 1} \exp(\delta_{ik})} \quad (9)$$

The final global spatial feature is aggregated as:

$$P_g = \sum_j \alpha_{ij} V_j \quad (10)$$

This mechanism enables Sparge-iTransformer to accurately model long-range spatial dependencies while significantly reducing computational complexity, thus providing high-quality global spatial features for subsequent temporal modeling and FDIA detection.

2.4. Temporal feature extraction

To capture the dynamic temporal variations in the measurement data, a GRU-based module is employed to model the temporal dependencies of spatial features across multiple time steps. The input to the GRU is a sequence of fused spatial features over T time steps, denoted as $\{P_{HA1}, P_{HA2}, \ldots, P_{HAT}\}$, where each P_{HAt} is obtained by combining the local feature P_s and the global feature P_g. The GRU updates the hidden state through two gating mechanisms. First, the update gate Z_t and the reset gate R_t are computed as:

$$Z_t = \sigma(W_z P_{HAt} + U_z H_{t-1}) \quad (11)$$

$$R_t = \sigma(W_r P_{HAt} + U_r H_{t-1}) \quad (12)$$

Then, the candidate hidden state at time step t is calculated as:

$$\widetilde{H}_t = tanh\big(W_h P_{HAt} + U_h (R_t \odot H_{t-1})\big) \quad (13)$$

where W and U are weight parameters, $\odot$ represents the Hadamard element-by-element product, and $tanh$ is the hyperbolic tangent function.

Finally, the hidden state is updated using a gated interpolation between the previous state and the candidate state:

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \widetilde{H}_t \quad (14)$$

After processing all T steps, the hidden state at the final time step H_{Tf} is selected as the output of the temporal modeling, denoted as:

$$P_c = H_{Tf} \quad (15)$$

This temporal feature P_c encodes both historical dynamic information and spatial context across time, and is subsequently used as input to the classification module for FDIA attack detection.

2.5. Others

After extracting temporal features from the GRU module, the combined temporal–spatial feature is flattened and passed through a fully connected layer to produce the final detection score y using a Sigmoid activation. The model classifies a sample as an attack if $y \geq \tau$, where τ is a predefined threshold.

A joint loss function is used during training, consisting of binary cross-entropy loss $\mathcal{L}_d$ and consistency loss $\mathcal{L}_a$, which enforces temporal feature stability. The total loss is defined as:

$$\mathcal{L} = \mathcal{L}_d + \lambda \mathcal{L}_a \quad (16)$$

where λ balances the two components.

3. Experiment results

This paper uses the Pandapower to model the IEEE-30 bus and IEEE-57 bus power systems and constructs a temporal dataset with both normal and attack states.

3.1. Dataset and experimental environment

To simulate dynamic load and generation variations, each bus's load is perturbed using a normal distribution with mean $1 + k[s_t]$ $(k = 0.1)$ and standard deviation $\sigma_s = 0.03$, where $[s_t]$ is the normalized ERCOT load data [9]. The DC optimal power flow (DCOPF) calculation is performed to obtain the bus injection power and branch flow at each time step as the baseline measurements.

For FDIA attack simulation, false data is injected during some time steps $t \in T_a$. The number of attacked buses n is randomly chosen from $[2,6]$, and the attack strength is sampled from the attack amplitude bounds $[C_{\min}, C_{\max}] = [4,5]$. Perturbations are added to the measurements with Gaussian noise $(\sigma_n = 0.012)$. The final dataset consists of 120,000 time steps, with 110,000 for training and 10,000 for testing. The attack sample ratio is 10% in training and 50% in testing. Experiments are conducted on a Windows 10 system with an NVIDIA GeForce RTX 4090 GPU, 64GB RAM, and a 2.11GHz processor, using Python 3.7 and libraries such as Pandapower, PyTorch, and Scikit-learn.

3.2. Detection performance experiment:

To evaluate the FDIA detection performance of HSGT-Net, four metrics are used: accuracy, recall, precision, and F1-

Score. We compare HSGT-Net with SVM, KNN, MLP, CNN, and GNN+LSTM, as shown in Tables 1 and 2. In both IEEE-30 and IEEE-57 systems, HSGT-Net consistently achieves the highest accuracy and F1-Score, demonstrating strong adaptability across different system scales.

Table 1. Experimental Results. (IEEE-30bus)

Model performance comparison	IEEE-30bus					
	SVM	KNN	MLP	CNN	GNN+ LSTM	HSGT-Net
Accuracy	83.50	88.83	90.29	93.20	94.56	**97.11**
Recall	82.14	88.18	90.00	92.73	94.25	**97.86**
Precision	86.79	90.65	91.67	94.44	95.63	**96.83**
F1-Score	84.40	89.40	90.83	93.58	94.94	**97.35**

Table 2. Experimental Results. (IEEE-57bus)

Model performance comparison	IEEE-57bus					
	SVM	KNN	MLP	CNN	GNN+ LSTM	HSGT-Net
Accuracy	79.8	85.22	87.68	90.34	92.68	**95.93**
Recall	77.78	84.40	86.24	89.09	91.89	**96.46**
Precision	83.17	87.62	90.38	92.80	94.44	**96.11**
F1-Score	80.38	85.98	88.26	90.91	93.15	**96.28**

Traditional machine learning models struggle to capture complex attack patterns, while deep learning methods fail to fully exploit the spatial–temporal characteristics of power systems. By integrating Hodge diffusion, Sparge attention, and GRU-based temporal modeling, HSGT-Net significantly improves FDIA detection performance.

3.3. Robustness experiment

To evaluate robustness, HSGT-Net is tested on the IEEE-30 system under varying load disturbance factor k, noise standard deviation σ_n, and attack amplitude $[C_{\min}, C_{\max}]$, using accuracy and F1-Score as metrics.

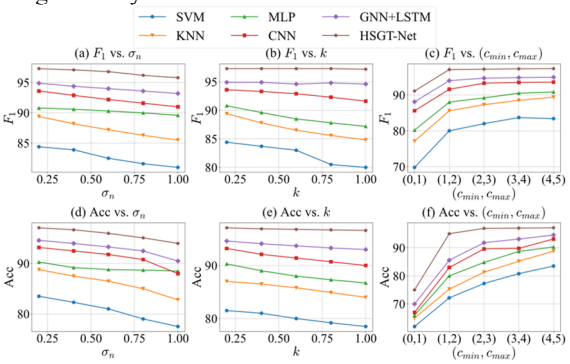

Figure 3. Detection performance comparison under various disturbance and attack conditions.

As shown in Figure 3, traditional models suffer significant performance degradation under increased disturbance and attack intensity, while HSGT-Net consistently achieves the highest accuracy and robustness, demonstrating strong adaptability and generalization in complex scenarios.

3.4. Computational cost experiment

To evaluate the training efficiency, we analyze the training time of each model on the IEEE-30 and IEEE-57 systems, as shown in Figure 4. Results show that the training time of SVM increases significantly with increasing system size, reaching 30 minutes on the IEEE-57 system, indicating poor scalability. In contrast, KNN remains relatively stable.

Among deep learning models, MLP and CNN show the fastest training speed, followed by GNN+LSTM. The HGT-Net has a higher computational cost due to the Transformer structure, with training time reaching 39 minutes per epoch on the IEEE-57 system.

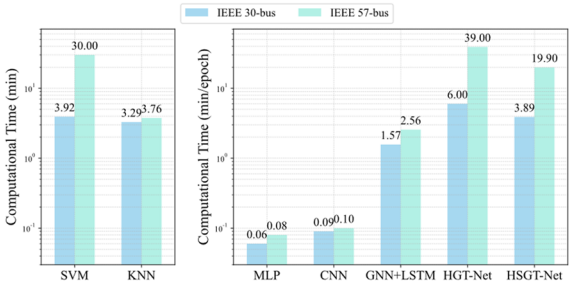

Figure 4. Comparison of computational time for different models on IEEE-30 and IEEE-57 systems.

In comparison, the proposed HSGT-Net reduces this cost to 19.90 minutes per epoch by using the Sparge module, achieving about 49% reduction. This demonstrates its training efficiency and suitability for large-scale applications.

While this section focuses on training efficiency, demonstrating HSGT-Net's reduced complexity via the Sparge module, a quantitative evaluation of inference latency and memory footprint on resource-constrained embedded devices (e.g., RTUs) was not conducted. Such analysis is crucial for assessing real-world deployment feasibility against operational requirements (like potential real-time constraints) and is planned for future work.

5. Conclusion

This paper proposes HSGT-Net, a model that jointly models spatial and temporal features to enhance FDIA detection accuracy and efficiency in smart grids. By integrating graph structures with temporal dynamics, the model addresses the limitations of traditional methods in robustness and computational cost under high-dimensional and complex scenarios. HSGT-Net demonstrates stable performance in large-scale systems and outperforms existing methods in accuracy, robustness, and efficiency. Future work will prioritize further enhancing model security by evaluating robustness against adaptive attacks targeting the Sparge mechanism's thresholds and by investigating the Hodge diffusion module's resilience to gradient masking, including the exploration of potential defense strategies.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant U2468203 and 62473038.

References

1. Chen, J., Wei, C., Wang, K., Yu, S., & Li, H. (2025) Cybersecurity of distributed energy resource systems in the smart grid: A survey. *Applied Energy*, 383: 125364.
2. Li, N., Wang, Q., Wang, Y., Liu, Y., Zhang, H., & Wang, J. (2025) Enhancing Detection of False Data Injection Attacks in Smart Grid Using Spectral Graph Neural Network. *IEEE Transactions on Industrial Informatics*, in press.
3. Luo, X., Zhang, Y., Liu, Y., & Li, H. (2020) Interval observer-based detection and localization against false data injection attack in smart grids. *IEEE Internet of Things Journal*, 8: 657–671.
4. Xiong, X., Li, X., Hu, Z., & Wang, T. (2022) Detection of false data injection attack in power information physical system based on SVM–GAB algorithm. *Energy Reports*, 8: 1156–1164.
5. Moayyed, H., Ghassemian, M., & Jiang, J. (2021) Image processing based approach for false data injection attacks detection in power systems. *IEEE Access*, 10: 12412–12420.
6. Kuo, P., Li, J., & Lin, Y. (2024) GNN-LSTM-based fusion model for structural dynamic responses prediction. *Engineering Structures*, 306: 117733.
7. Xia, W., Huang, Y., Li, G., Wu, J., Zhang, Y., & Liu, J. (2024) Locational detection of false data injection attacks in the edge space via hodge graph neural network for smart grids. *IEEE Transactions on Smart Grid*, in press
8. Zhang, J., et al. (2025) Spargeattn: Accurate sparse attention accelerating any model inference. *arXiv*, 2502.18137.
9. Electr. Reliability Council of Texas Corp. (2020) Backcasted (Actual) Load Profiles—Historical. [Online]. Available: http://www.ercot.com/mktinfo/loadprofile/alp/.