HCFS-GNN: Semantic-Aligned Differentiable Heterogeneous Graph Anomaly Detection for Industrial Systems

1st Shiao Mi
Tianjin University of Science and Technology
Tianjin, China
m15698273520@163.com

2nd Xiangyu Wu
Beijing KeDong Electric Power Control System Co., Ltd.
Beijing, China
wuxiangyu@sgepri.sgcc.com.cn

3rd Suxiang Zhang
state grid information & Telecommunication center (Big Data center)
Beijing, China
zsuxiang@163.com

4th Yiying Zhang*
Tianjin University of Science and Technology
Tianjin, China
yiyingzhang@tust.edu.cn

5th Hongxin Zhao
Tianjin University of Science and Technology
Tianjin, China
18334897164@163.com

***Abstract*—Stealthy semantic decoupling attacks targeting control logic pose a severe threat to the integrity of Industrial Cyber-Physical Systems (ICPS). Existing homogeneous graph-based anomaly detection methods often struggle to capture the semantic conflicts between physical processes and control commands due to functional role confusion between sensors and actuators. To address this challenge, this paper proposes a Heterogeneous Control-Flow Sensitive Graph Neural Network (HCFS-GNN). First, we construct a heterogeneous bipartite control graph to explicitly decouple sensor observations from actuator actions at the topological level. Second, we introduce Symbolic Transfer Entropy (STE) combined with Gumbel-Softmax reparameterization to achieve lightweight real-time causal discovery and differentiable end-to-end structure optimization. Furthermore, a novel Control-Flow Sensitivity (CFS) metric is designed as a semantic high-pass filter to significantly amplify logical violation signals in critical control nodes. Extensive experiments on the SWaT and WADI benchmarks demonstrate that HCFS-GNN achieves F1-scores of 0.952 and 0.749, respectively. Notably, the proposed method achieves complete coverage of attack scenarios while maintaining high precision, and exhibits superior robustness under conditions of high noise and limited training data.**

Keywords—Industrial Cyber-Physical Systems, Anomaly Detection, Graph Neural Networks, Causal Discovery.

I. INTRODUCTION

Industrial Cyber-Physical Systems (ICPS), which deeply integrate computing, communication, and control technologies, have become the core driving force for critical infrastructures such as water treatment, power grids, and intelligent manufacturing [1,2]. However, it has also dismantled the traditional security boundaries of physical isolation, exposing core production networks directly to external threats [3]. In recent years, Advanced Persistent Threats (APTs) targeting industrial control systems, such as Stuxnet and Triton, have frequently occurred. The focus of these attacks has shifted from merely disrupting data availability to compromising the integrity of control logic in a stealthier manner [4]. Such attacks exploit physical inertia and control loop delays to tamper with actuator commands while simultaneously forging sensor feedback data. Under this mode, the statistical features of individual variables often remain within normal thresholds, causing traditional univariate detection systems to fail.

To address increasingly complex threats, anomaly detection methods have evolved from statistical learning to deep learning. Early methods like PCA and Isolation Forest [5,6], while computationally simple, struggle with high-dimensional non-linear data. With the development of deep learning, reconstruction-based or prediction-based models have become mainstream. For instance, LSTM-VAE [7] combines Long Short-Term Memory networks with Variational Autoencoders to capture temporal dependencies, while USAD [8] significantly enhances anomaly amplification through adversarial training mechanisms. However, these methods essentially treat multivariate time series as flat vectors, ignoring the complex topological structures and interaction mechanisms between variables. This results in significant topological blind spots when models face attacks that exploit logical vulnerabilities.

To bridge the gap in spatial modeling, Graph Neural Networks (GNNs) have been introduced into this field. Methods represented by GDN [9] and MTAD-GAT [10] attempt to explicitly model dependencies between variables via graph attention mechanisms. Despite performance improvements, existing GNN paradigms still face two theoretical defects when addressing ICPS security. The first is semantic mismatch [11]. Most methods build graph structures based on physical connections or statistical correlations. However, physical connections are not equivalent to logical control dependencies, and statistical correlation does not imply causation. This mismatch leads to the introduction of spurious edges and a failure to capture sparse control logic. The second defect is functional role confusion. Existing research widely ignores the essential difference between sensors and actuators in the control loop [12], treating them homogeneously. This approach triggers a semantic smoothing effect, where the sharp, deterministic logical features of actuators are masked by physical sensor noise, making it difficult for models to distinguish between normal physical fluctuations and malicious logical tampering.

To address these pain points, this paper proposes a Heterogeneous Control-Flow Sensitive Graph Neural Network (HCFS-GNN), shifting the focus of anomaly

detection from mere data association mining to deep control-flow semantic understanding. The main contributions of this paper are as follows:

1) We explicitly differentiate between sensor and actuator nodes at the modeling level, fundamentally resolving the problem of functional role confusion.

2) Utilizing Symbolic Transfer Entropy (STE) [13], we reduce the complexity of causal discovery to linear time and employ a dual defense mechanism to capture both logical structure anomalies and physical magnitude anomalies.

3) We introduce the Gumbel-Softmax reparameterization technique to achieve end-to-end optimization of the causal graph structure, removing reliance on static prior knowledge.

4) We design the Control-Flow Sensitive (CFS) metric as a semantic high-pass filter, forcing the model to focus on high-authority critical control nodes and effectively overcoming the semantic smoothing effect.

II. Problem Definition and Preliminaries

Before detailing the methodology, this chapter first provides a strict mathematical formalization for the ICPS data structure, the heterogeneous graph model, and the threat model focused on in this paper.

A. Symbol Definition and ICPS Semantic Modeling

Consider an ICPS containing N monitoring points. The state of the system at a discrete time step t is represented by a multivariate vector $\boldsymbol{x}_t \in \mathbb{R}^N$. To reflect the essential differences in function and semantics among system components, we discard the traditional homogeneous definition. Instead, we partition the node set $\mathcal{V}$ into two mutually exclusive subsets:

Sensor Set ($\mathcal{V}_S$):$\mathcal{V}_S = \{s_1, s_2, \dots, s_m\}$, containing m sensor nodes. Their observation values $\boldsymbol{x}_S^{(t)}$ are continuous variables constrained by physical mechanisms, representing the system's state or effect.

Actuator Set ($\mathcal{V}_A$): $\mathcal{V}_A = \{a_1, a_2, \dots, a_n\}$, containing n actuator nodes. Their state values $\boldsymbol{x}_A^{(t)}$ are typically discrete variables or setpoints constrained by control logic, representing the system's action or cause.

Clearly, $\mathcal{V} = \mathcal{V}_S \cup \mathcal{V}_A$, and $N = m + n$. Our goal is to train a model $f(\cdot)$ based on historical normal data sequences $\mathcal{X}_{train} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_T\}$. In the testing phase, this model will calculate an anomaly score $Score(t)$ in real-time to determine whether $\boldsymbol{x}_t$ is anomalous.

B. Dynamic Heterogeneous Bipartite Control Graph

To address the semantic mismatch problem, we model the system state as a dynamic directed heterogeneous graph.

Definition 1 (Heterogeneous Control-Flow Graph): At time t, the system topology is represented as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, \mathcal{A}_t, \mathcal{R})$.

$\mathcal{V}$: The set of nodes, containing a type mapping function $\phi: \mathcal{V} \to \{S, A\}$.

$\mathcal{E}_t$: The set of edges that evolve over time.

$\mathcal{A}_t$: The weighted adjacency matrix, where element $w_{uv}^{(t)}$ represents the causal strength from node u to node v.

$\mathcal{R}$: The set of edge types. This paper defines three edge types with clear control semantics $r \in \mathcal{R}$:

Sensing-Decision Edge ($S \to A$): Represents the logical dependency where a sensor state triggers an actuator action. This is a direct manifestation of control logic.

Execution-Physics Edge ($A \to S$): Represents the physical causality where an actuator action drives a change in the physical process. This is a manifestation of the physical process.

Interlock Sequence Edge ($A \to A$): Represents operational timing or mutual exclusion constraints between actuators (e.g., Valve Open followed by Motor Start).

As shown in Fig. 1, this heterogeneous graph definition forces the model to explicitly distinguish between different types of interactions during the feature learning phase. This effectively avoids confusing physical influences with logical instructions.

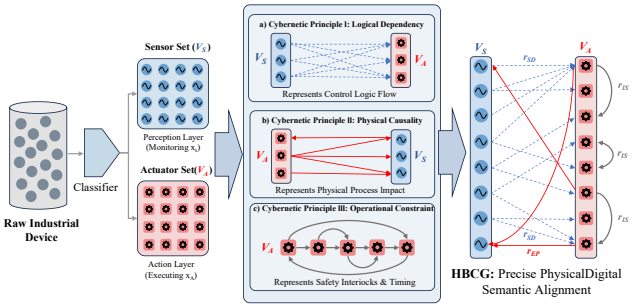

Fig. 1. Schematic illustration of the construction mechanism for the proposed Heterogeneous Bipartite Control Graph (HBCG). The mechanism first clarifies sensor set ($\mathcal{V}_S$) and actuator set ($\mathcal{V}_A$) through ontological separation, and then defines three specific semantic edges based on cybernetic principles (Sensing-Decision Logic r_{SD}, Execution-Physics Alignment r_{EP}, and Interlock Sequence r_{IS}).

C. Stealthy Semantic Decoupling Attack

This paper focuses on stealthy attacks targeting control logic, defined as Stealthy Control-Flow Attacks.

Definition 2 (Stealthy Semantic Decoupling Attack): An attacker tampers with the actuator state $x_a(t)$ to $x_a'(t)$ (logic attack) within a time interval $[t_s, t_e]$. To evade detection, they may simultaneously tamper with sensor data $x_s(t)$ to $x_s'(t)$ (physical spoofing). This attack is characterized by:

Statistical Stealthiness: The value of any single variable $x_i'(t)$ remains within the normal operating range and does not trigger univariate threshold alarms.

Logical Violation: The causal logic between $x_s'(t)$ and $x_a'(t)$ (as defined by $\mathcal{G}_t$) is destroyed. For example, an attacker might forcibly close a water pump but fabricate sensor data showing normal water pressure. In this case, although the water pressure data appears normal, it is physically incompatible with the "pump closed" state.

III. Methodology

This chapter details the proposed Heterogeneous Control-Flow Sensitive Graph Neural Network (HCFS-GNN)

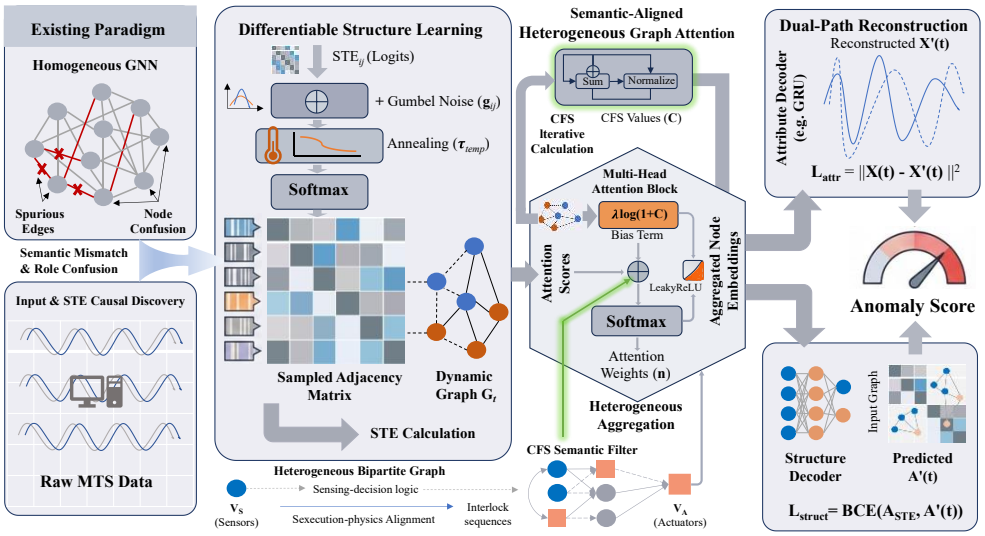

Fig. 2. Schematic diagram of the proposed HCFS-GNN method. The left panel illustrates how the heterogeneous bipartite graph modeling overcomes the limitations of traditional homogeneous GNNs. The central part details the process of differentiable structure learning via Gumbel-Softmax and how the CFS semantic filter injects logical constraints into the attention mechanism.

framework. As shown in Fig. 2, the architecture consists of four modules: real-time causal discovery based on Symbolic Transfer Entropy (STE), differentiable structure learning via Gumbel-Softmax, a heterogeneous graph attention network deeply coupled with Control-Flow Sensitivity (CFS), and a dual-path reconstruction mechanism.

A. STE-based Real-time Causal Discovery

To resolve the computational complexity crisis ($O(N^2)$) of traditional kNN-TE algorithms, we introduce Symbolic Transfer Entropy (STE). We further design a dual defense mechanism to address the potential loss of magnitude information inherent in symbolization.

Given a time series $X = \{x_1, x_2, \dots, x_T\}$, we first utilize Takens' embedding theorem to reconstruct the phase space. We define the embedding vector $X_t = (x_t, x_{t+\tau}, \dots, x_{t+(m-1)\tau})$, where m is the embedding dimension and τ is the time delay.

To address parameter sensitivity, we adopt an adaptive strategy, τ is determined by the first minimum of the Auto-Correlation Function (ACF) or Average Mutual Information (AMI) to adapt to different sensor dynamics; m is determined via the False Nearest Neighbors (FNN) method. Subsequently, each embedding vector X_t is mapped to a Permutation Symbol $\hat{x}_t$ based on the rank order of its elements.

Traditional full-graph causal inference requires computing TE values for all node pairs (i, j), resulting in a complexity of $O(N^2 \cdot T)$. In STE, the symbolization process itself is linear $O(T)$. To achieve true $O(N)$ complexity, we introduce a sparse neighborhood constraint, calculating STE only within a candidate neighborhood $\mathcal{N}_k$ defined by physical adjacency or historical strong correlations. This reduces the computational load from $N \times N$ to $N \times k$ (where $k \ll N$ is a constant).

The Symbolic Transfer Entropy from source node Y to target node X is defined as:

$$STE_{Y \to X} = \sum_{\hat{x}_{t+1}, \hat{x}_t, \hat{y}_t} p(\hat{x}_{t+1}, \hat{x}_t, \hat{y}_t) \log \frac{p(\hat{x}_{t+1} | \hat{x}_t, \hat{y}_t)}{p(\hat{x}_{t+1} | \hat{x}_t)} \quad (1)$$

While STE gains high noise resistance and computational efficiency through symbolization, it inevitably discards specific magnitude information. To mitigate this, we construct a Dual Defense Mechanism. The STE Structural Flow (First Defense) utilizes the causal skeleton to capture structural anomalies; for instance, sharp changes in STE values occur when sensor trends physically contradict actuator actions. Complementarily, the Attribute Reconstruction Flow (Second Defense), implemented in the decoder (see Section 3.4), minimizes $\|x - \hat{x}\|^2$ to capture physical anomalies.

B. Differentiable Structure Learning

Traditional graph construction often employs a hard threshold function $\mathbb{I}(STE > \theta)$ to determine edge existence. This operation is mathematically non-differentiable, preventing gradient backpropagation. To resolve this gradient blockage, we introduce the Gumbel-Softmax reparameterization.

It is crucial to define the optimization objective: we are not optimizing the STE calculation itself (which is a discrete preprocessing step), but rather the graph structure sampling distribution based on the STE prior. The neural network, guided by the final task loss, adjusts the sampling probability of edges. We treat the normalized STE_{ij} as unnormalized logits and inject Gumbel noise $g_{ij} \sim \text{Gumbel}(0,1)$ to introduce stochasticity. The connection weight $\hat{A}_{ij}$ (approximating the probability of edge existence) is calculated via the Softmax function:

$$\hat{A}_{ij} = \frac{\exp((\log(STE_{ij}) + g_{ij}) / \tau_{temp})}{\sum_k \exp((\log(STE_{ik}) + g_{ik}) / \tau_{temp})} \quad (2)$$

Here, τ_{temp} is the temperature coefficient. As $\tau_{temp} \to 0$, the distribution approaches a Categorical (One-hot) distribution. During training, we employ an annealing strategy, gradually reducing τ_{temp}. This renders the graph structure $\hat{\boldsymbol{A}}$ a learnable latent variable, bridging discrete statistical priors with continuous optimization spaces.

C. Semantic-Aligned Heterogeneous Graph Attention with CFS

To empower the model to identify critical assets and resolve the semantic smoothing issue, we propose the Control-Flow Sensitivity (CFS) metric and embed it as a high-pass filter into the heterogeneous graph attention network.

CFS is a dynamic centrality measure based on the heterogeneous causal graph. Drawing from PageRank, we posit that an actuator's importance depends on its inherent physical criticality (prior knowledge) and the sensitivity of nodes pointing to it. For an actuator node $a_i \in \mathcal{V}_A$, its CFS value $C(a_i)$ is iteratively calculated as:

$$C(a_i) = (1-\alpha) \cdot BaseCrit(a_i) + \alpha \sum_{j \in \mathcal{N}_{in}(a_i)} \frac{C(j) \cdot \hat{A}_{ji}}{OutDegree(j)} \quad (3)$$

where BaseCrit(a_i) is initialized based on Entity-Spatial Relationships (ESR), and $\hat{A}_{ji}$ is the dynamic weight generated by Gumbel-Softmax.

To thoroughly address semantic smoothing, we embed CFS as a bias term in the attention coefficient calculation:

$$\alpha_{ij}^{logic} = Softmax_j\left(LeakyReLU\left(\boldsymbol{a}_{logic}^T [\boldsymbol{h}_i \| \boldsymbol{h}_j]\right) + \lambda \log(1 + C(s_j))\right) \quad (4)$$

Physically, the term $\lambda \log(1 + C(s_j))$ acts as a "semantic high-pass filter". It allows discrete signals from high-CFS nodes (critical logic nodes) to "penetrate" the low-frequency physical background noise. Even if the numerical fluctuation of a critical sensor is minimal, the model forces a higher attention weight to verify its consistency if its CFS value is high.

D. Dual-Path Reconstruction

In response to the dual defense mechanism, the decoder is designed with two parallel paths. The attribute reconstruction path reconstructs the raw readings of sensors and actuators $\widehat{\boldsymbol{X}}^{(t)}$, with a loss function $\mathcal{L}_{\text{attr}} = \|\boldsymbol{X}^{(t)} - \widehat{\boldsymbol{X}}^{(t)}\|^2$, responsible for capturing magnitude anomalies. The structure reconstruction path reconstructs the current control logic connections $\widehat{\boldsymbol{A}}^{(t)}$, with a loss function $\mathcal{L}_{\text{struct}} = \text{BCE}(\mathbf{A}_{STE}^{(t)}, \widehat{\boldsymbol{A}}^{(t)})$, used to detect logical decoupling.

The final anomaly scoring function incorporates CFS as a logic amplifier:

$$\text{Score}(t) = \sum_{i \in \mathcal{V}} \left(1 + \gamma \cdot C(i)\right) \cdot \left((1-\beta)\|\boldsymbol{x}_i^{(t)} - \hat{\boldsymbol{x}}_i^{(t)}\|^2 + \beta \mathcal{L}_{\text{struct}}^{(i,t)}\right) \quad (5)$$

This ensures that for core control devices (high CFS), even minor logical deviations are significantly amplified by the factor $(1 + \gamma \cdot C(i))$, thereby triggering alarms.

IV. Experiments

A. Experimental Setup

1) Datasets: Experiments are conducted on two real-world datasets collected from large-scale testbeds, which are standard benchmarks for ICPS security research:

SWaT (Secure Water Treatment) [14]: A six-stage water treatment testbed. The dataset contains 7 days of normal operation data and 4 days of attack data, covering 51 sensors and actuators. It includes 36 distinct attack scenarios, ranging from single-point anomalies to complex multipoint control-logic violations.

WADI (Water Distribution) [15]: An extension of SWaT representing a water distribution network. It features 127 nodes and more complex coupling logic. The dataset includes 14 days of normal operation and 2 days of attack scenarios.

2) Baselines: The proposed HCFS-GNN is compared against three categories of methods:

Reconstruction/Prediction-based: LSTM-VAE [7], and USAD [8]. These methods rely on statistical deviations without explicit graph modeling.

Homogeneous GNNs: GDN [9] (Graph Deviation Network) and MTAD-GAT [10]. These represent the current mainstream graph-based approaches but suffer from the functional role confusion and semantic mismatch identified in this paper.

Domain-Specific: SA2 [12], a method that distinguishes sensors/actuators but lacks dynamic causal structure learning.

3) Evaluation Metrics: Standard Precision, Recall, and F1-Score are employed to quantify point-level detection accuracy. To overcome the bias of traditional point-based metrics in continuous attack intervals, the Time-series aware Precision and Recall (TaPR) metric is utilized. TaPR weights the score based on the overlap between the detection result and the ground truth attack interval, providing a more realistic reflection of the model's coverage of continuous anomalies. Finally, referencing the study in SA2, the Undetected Attack Rate (UAR) is adopted as an event-level metric. UAR measures the percentage of attack scenarios that are completely missed by the model.

B. Overall Detection Performance

Table I summarizes the complete test results of all models on the SWaT and WADI datasets, covering five key metrics.

As shown in Fig. 3, HCFS-GNN demonstrates superior performance across both datasets. On the SWaT dataset, in contrast to the significant undetected rate exhibited by the homogeneous graph model GDN, HCFS-GNN achieves complete coverage of all attack scenarios while maintaining high Precision, effectively validating the advantage of heterogeneous graph modeling in capturing logical anomalies. On the more complex WADI dataset, benefiting from STE-based dynamic causal structure learning, the proposed model significantly outperforms SA2, which is constrained by static topology, as well as traditional reconstruction methods in terms of F1-Score and Recall.

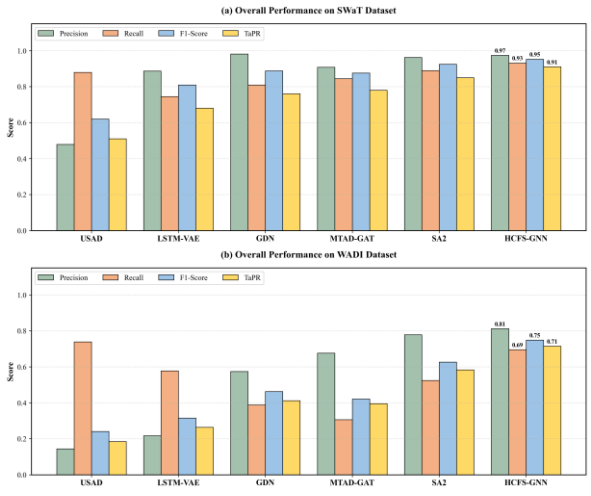

Fig. 3. Detection performance comparison results on SWaT and WADI datasets.

TABLE I. OVERALL PERFORMANCE COMPARISON ON SWAT AND WADI DATASETS

Model	SWaT					WADI				
	Pre	*Rec*	*F1*	*TaPR*	*UAR*	*Pre*	*Rec*	*F1*	*TaPR*	*UAR*
USAD	0.479	0.879	0.620	0.510	16.6%	0.144	0.739	0.241	0.185	23.0%
LSTM-VAE	0.886	0.745	0.809	0.680	50.0%	0.217	0.578	0.315	0.264	23.1%
GDN	0.981	0.809	0.887	0.760	45.8%	0.574	0.388	0.463	0.412	38.4%
MTAD-GAT	0.908	0.844	0.875	0.780	16.7%	0.676	0.306	0.421	0.395	23.1%
SA2	0.962	0.889	0.924	0.850	0.0%	0.778	0.524	0.626	0.583	0.0%
HCFS-GNN	**0.974**	**0.931**	**0.952**	**0.910**	**0.0%**	**0.812**	**0.695**	**0.749**	**0.715**	**0.0%**

C. *Robustness and Generalization Analysis*

To verify the applicability of the model in non-ideal industrial environments, the model's robustness to noise interference and its dependency on training data volume were further evaluated. In the noise robustness test, Gaussian noise of 5%, 10%, and 15% intensities was injected into the test set. As shown in Fig. 4, as noise intensity increased, the performance of purely data-driven models declined sharply, with an average F1-Score drop exceeding 15%. This is attributed to the tendency of these models to overfit high-frequency fluctuations in the raw data. In contrast, HCFS-GNN demonstrated exceptional stability; even under a 15% noise level, the F1-Score dropped by only approximately 3%.

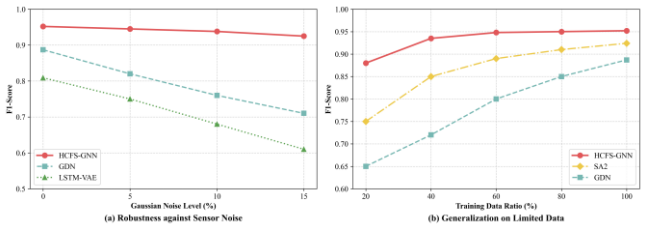

Fig. 4. Performance evaluation under non-ideal conditions. (a) Detection performance comparison under varying Gaussian noise levels. (b) Model generalization capability with different training data ratios.

In the data efficiency and generalization test, the model's performance was evaluated under conditions of limited training data (ranging from 20% to 100%). Results indicate that HCFS-GNN achieves over 95% of its final performance using only 40% of the training data, exhibiting a significantly faster convergence rate compared to all baseline models. In contrast, fully connected graph models such as GDN exhibit suboptimal performance with limited data, primarily due to the requirement for massive amounts of samples to prune spurious correlations.

D. *Ablation Study*

To verify the contribution of each core component within the framework, an ablation study with four variants was designed: removing the STE module (substituting with Pearson correlation), removing the heterogeneous structure (utilizing homogeneous GAT), removing Gumbel-Softmax (applying hard thresholds), and removing the CFS bias. As illustrated in Fig. 5 and Table II, results indicate that the exclusion of the STE module precipitated the most significant performance degradation (F1 declined from 0.952 to 0.865), accompanied by a surge in the undetected rate to 12.5%. This substantiates the cornerstone role of dynamic causal inference in capturing non-linear couplings. Furthermore, while removing the CFS bias resulted in a more moderate impact on F1, it led to an increase in the undetected rate to 8.3%.

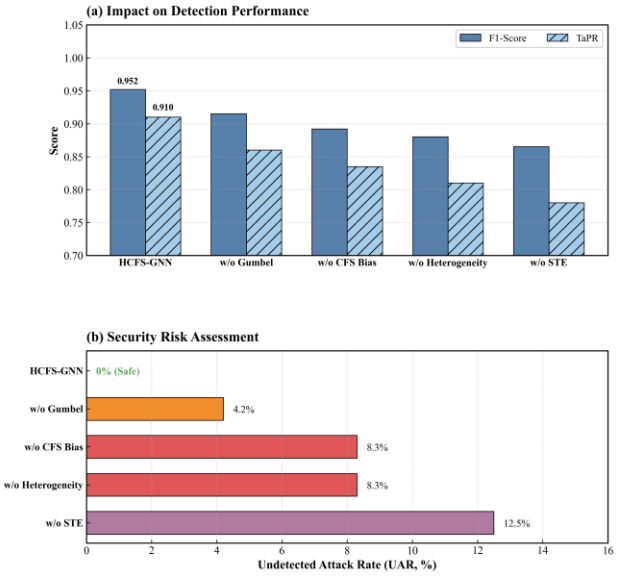

Fig. 5. Ablation Study Comparison Figure

TABLE II. ABLATION STUDY RESULTS

Variants	Pre	Rec	F1	TaPR	UAR
w/o STE	0.910	0.824	0.865	0.780	12.5%
w/o Heterogeneity	0.940	0.827	0.880	0.810	8.3%
w/o CFS	0.970	0.825	0.892	0.835	8.3%
w/o Gumbel-Softmax	0.965	0.870	0.915	0.860	4.2%
Full Model	**0.974**	**0.931**	**0.952**	**0.910**	**0.0%**

E. *Detection of Stealthy Attacks*

To intuitively demonstrate the capability of the model in detecting semantic decoupling attacks, an in-depth analysis of the P-101 Attack Case in the SWaT dataset was conducted. As

illustrated in Fig. 6, in this scenario, the attacker forces the main inlet pump (P-101) to the ON state while the tank level (LIT-101) remains low, intending to cause dry-run damage. Simultaneously, the LIT-101 data is spoofed via a man-in-the-middle attack to display a "normal level." Traditional models such as GDN failed to detect this anomaly, primarily because their learned attention weights were distributed uniformly among neighbor nodes. Since "normal level" and "pump on" are not historically mutually exclusive, statistical correlation masked the logical contradiction. In contrast, the attention heatmap of HCFS-GNN revealed a distinct pattern: at the onset of the attack, the attention weight assigned to the P-101 actuator node surged to over 0.85. This phenomenon is attributed to the STE module capturing the causal break between the P-101 action and the LIT-101 trend in real-time, resulting in a spike in structural reconstruction error. Concurrently, the CFS mechanism identified P-101 as the core actuating node within the current control flow, thereby compelling the attention mechanism to concentrate on this critical component.

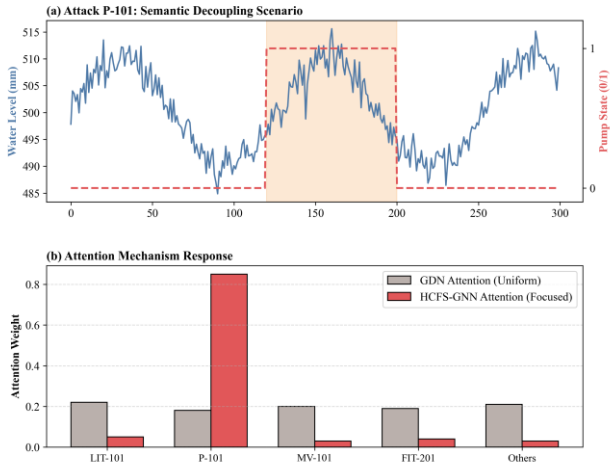

Fig. 6. Visualization of the stealthy semantic attack case (P-101). (a) The semantic decoupling scenario between pump P-101 and water level LIT-101. (b) Comparison of attention weight distributions during the attack, highlighting the logic conflict amplification mechanism of HCFS-GNN.

V. Conclusion

This paper proposes the HCFS-GNN framework to address stealthy semantic decoupling attacks targeting control logic in ICPS. By constructing a Heterogeneous Bipartite Control Graph combined with STE, this method effectively overcomes the semantic mismatch and functional role confusion defects inherent in traditional homogeneous Graph Neural Networks. The introduction of a differentiable Gumbel-Softmax mechanism achieves end-to-end optimization of the causal graph structure, while the CFS metric acts as a semantic high-pass filter, successfully preventing critical logical violations from being masked by physical noise. Experimental results on two datasets, SWaT and WADI, demonstrate that HCFS-GNN outperforms existing baselines and exhibiting exceptional robustness and generalization capabilities under high noise and limited training data conditions.

Future work will focus on further enhancing the scalability and adaptability of the model in ultra-large-scale complex industrial networks. We aim to explore lighter-weight causal inference mechanisms to handle larger node networks while maintaining linear time complexity, thereby supporting efficient deployment on resource-constrained edge devices. Furthermore, given the dynamic nature of industrial system operations, we plan to investigate strategies combining transfer learning or online learning. This will enable the model to rapidly adapt to new system configurations or operating conditions without the need for large-scale retraining, ensuring continuous and effective defense against evolving Advanced Persistent Threats.

References

[1] A. K. Verma, A. Navas-Fonseca, C. Burgos-Mellado and T. Dragičević, "Realigning Academic Cybersecurity Research With Industrial Needs in Cyber-Physical Systems," in IEEE Open Journal of the Industrial Electronics Society, vol. 6, pp. 1210-1247, 2025.

[2] J. Jin et al., "Cloud-Fog Automation: The New Paradigm Toward Autonomous Industrial Cyber-Physical Systems," in IEEE Journal on Selected Areas in Communications, vol. 43, no. 9, pp. 2917-2937, Sept. 2025.

[3] J. Xing, D. Ye and P. Li, "Stealthy Attacks With Historical Data on Distributed State Estimation," in IEEE Transactions on Information Forensics and Security, vol. 20, pp. 4541-4550, 2025.

[4] P. Griffioen, B. H. Krogh and B. Sinopoli, "Ensuring Resilience Against Stealthy Attacks on Cyber-Physical Systems," in IEEE Transactions on Automatic Control, vol. 69, no. 12, pp. 8234-8246, Dec. 2024.

[5] J. Zheng, Z. Yang and Z. Ge, "Deep Residual Principal Component Analysis as Feature Engineering for Industrial Data Analytics," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-10, 2024, Art no. 2523310.

[6] H. Xu, G. Pang, Y. Wang and Y. Wang, "Deep Isolation Forest for Anomaly Detection," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 12, pp. 12591-12604, 1 Dec. 2023.

[7] D. Park, Y. Hoshi and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based Variational Autoencoder," in IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 1544-1551, July 2018.

[8] J. Audibert, P. Michiardi, F. Guyard, S. Marti and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series." In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining , pp. 3395-3404, 2020.

[9] A. Deng and B.Hooi, "Graph Neural Network-Based Anomaly Detection in Multivariate Time Series," in Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4027-4035, 2021.

[10] H. Zhao et al., "Multivariate Time-Series Anomaly Detection via Graph Attention Network," 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 841-850.

[11] C. Lu et al., "Heterogeneous Data Fusion and Anomaly Detection in Industrial IoT Systems Using Spatio-Temporal Graph Neural Networks," 2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM), Chengdu, China, 2024, pp. 548-556.

[12] J. Cai, Z. Wei and J. Luo, "ICS Anomaly Detection Based on Sensor Patterns and Actuator Rules in Spatiotemporal Dependency," in IEEE Transactions on Industrial Informatics, vol. 20, no. 8, pp. 10647-10656, Aug. 2024.

[13] C. Rui, S. Liang, J.G. Wang, Y. Yao, J.R. Su, and L.L. Liu, "Lag-Specific Transfer Entropy for Root Cause Diagnosis and Delay Estimation in Industrial Sensor Networks," Sensors 25, no. 13: 3980, 2025.

[14] J. Goh, S. Adepu , K.N. Junejo and A. Mathur, "A Dataset to Support Research in the Design of Secure Water Treatment Systems," Critical Information Infrastructures Security. CRITIS 2016. Lecture Notes in Computer Science(), vol 10242. Springer, Cham, 2017.

[15] A. C. Mujeeb, V. R. Palleti, and A.P. Mathur. "WADI: a water distribution testbed for research in the design of secure cyber physical systems." Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks. 2017: 25-28.