Physics-Informed Neural Network-Based Anomaly Detection for Digital Substations Under Multi-IED Compromise

Kuchan Park, Junho Hong, Wencong Su
Department of Electrical and Computer Engineering
University of Michigan-Dearborn
Dearborn, MI, USA
kuchan@umich.edu, jhwr@umich.edu, wencong@umich.edu

Akila Herath, Chen-Ching Liu
Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, VA, USA
akilaasansana@vt.edu, ccliu@vt.edu

***Abstract*—The digitalization of power systems has accelerated the deployment of Intelligent Electronic Devices (IEDs) in substations, significantly enhancing monitoring, protection, and control capabilities. At the same time, the attack surface for cyber threats has expanded, and coordinated attacks on multiple IEDs can severely compromise the integrity of measurement data. Conventional Byzantine or voting-based frameworks assume that at least $2f + 2$ IEDs (where f is the number of compromised nodes) are available, and therefore become ineffective when more than half of the IEDs are corrupted. To address this limitation, this paper proposes a Physics-Informed Neural Network (PINN) based anomaly detection framework that remains effective even under majority-IED compromise. For each feeder, an independent PINN model is trained using physical quantities such as voltage, current, frequency, and power flow. Kirchhoff's current law (KCL), power flow relationships, thermal line limits, and frequency consistency are embedded as physical constraints in the loss function, enabling verification of physical plausibility rather than relying on cross-device agreement. The residuals derived from these constraints are converted into feeder-level anomaly indices, and a statistical three-sigma rule is used to determine detection thresholds. Case studies on the IEEE 30-bus system demonstrate that the proposed PINN approach can detect anomalies with high accuracy even when a majority of IEDs are compromised, while conventional consensus-based schemes fail.**

***Index Terms*—Physics-Informed Neural Network (PINN), Cybersecurity, Digital substation, IEC61850**

I. Introduction

The digital transformation of substations has accelerated through the widespread deployment of Intelligent Electronic Devices (IEDs), merging advanced sensing, protection, and automation into cyber-enabled architectures. With IEC 61850-based communication, sampled value (SV) streams, and time-synchronized phasor measurements, digital substations achieve enhanced visibility and operational flexibility. However, the increased reliance on networked IEDs also expands the cyber attack surface, making substations vulnerable to sophisticated, multi-point attacks.

Conventional security and anomaly detection mechanisms rely on the assumption that most devices operate correctly. Byzantine Fault Tolerant (BFT) relay schemes and majority voting approaches require that the number of compromised devices remain below half [1]. This constraint can be expressed as: $f > n/2$ where f denotes the number of compromised IEDs and n is the total population. When $f > n/2$, attackers can manipulate a majority of devices and generate false but mutually consistent measurements. Under this condition, consensus-based protection, state estimation residual checks, and machine learning classifiers can all be deceived.

In [2], a consensus-based protection logic was reinforced through relay-to-relay voting, demonstrating the applicability of BFT in distributed protection architectures. The study in [3] introduced a security mechanism against Indirect Coordinated Attacks (ICA), where load-side power electronic devices are exploited as intermediaries to misoperate protection relays. A PINN that integrates physical constraints with data-driven learning was developed in [4] to classify diverse fault scenarios. Meanwhile, [5] simulated data-integrity attacks within an IEC 61850-based substation using a hardware-in-the-loop (HIL) testbed and quantified their impacts through the Voltage Performance Index (VPI). Furthermore, [6] proposed a probabilistic Defender–Attacker–Defender (DAD) formulation that optimizes the allocation of defensive resources to mitigate the risk of IED compromise.

These existing research addresses isolated or independent compromises, single-injection false data attacks, or random communication errors. However, future threat models must consider coordinated cyberattacks that simultaneously compromise multiple IEDs, manipulate SCADA data streams, and bypass traditional validation mechanisms. In such scenarios, cyber-layer anomaly detection alone is insufficient if attackers forge measurements that remain statistically realistic.

A fundamental limitation of prior work is the lack of physical consistency enforcement. Even if compromised IEDs generate mutually consistent data, they may collectively violate power system physics. Machine learning models trained only on historical data cannot distinguish cyber-coherent false measurements from physically impossible ones.

To address this challenge, this paper introduces a Physics-Informed Neural Network (PINN) framework that embeds power system equations into the learning process. Unlike traditional neural networks, PINNs incorporate electrical laws directly into the loss function, ensuring that anomaly detection

979-8-3315-8138-1

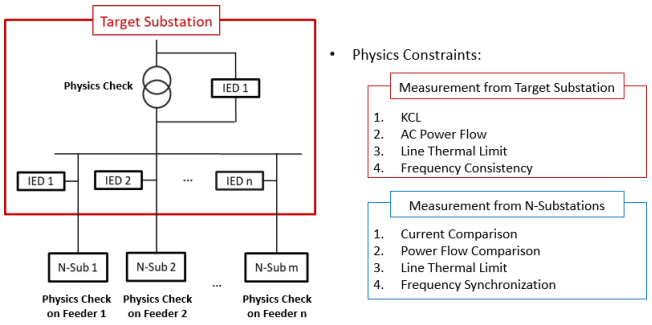

Fig. 1. The conceptual diagram of the proposed anomaly detection.

is based not only on data similarity, but also on physical feasibility.

The proposed method:

- does not require a majority of uncompromised IEDs,
- detects anomalies even under coordinated, multi-point data manipulation,
- incorporates Kirchhoff's Current Law, AC power flow relations, thermal limits, and frequency constraints,
- produces a physically informed anomaly index that reflects law violations.

This makes the approach suitable for digital substations where cyber and physical layers are deeply integrated, and where false data injection may target multiple devices simultaneously. The method remains valid under extreme attack scenarios that render existing approaches ineffective.

The rest of this paper is organized as follows. Section II describes the proposed PINN-based anomaly detection methodology and the embedded physical constraints. Section III presents case studies on the IEEE 30-bus system to evaluate performance under various IED compromise scenarios. Section IV concludes the paper and discusses key findings and future research directions.

II. Methodology

The proposed methodology introduces a Physics-Informed Neural Network (PINN) framework for anomaly detection in substations where a majority of Intelligent Electronic Devices (IEDs) may be compromised. Unlike traditional data-driven methods, the proposed approach relies solely on physical measurements—such as voltage, current, power, and frequency—and embeds power system physics into the neural learning process. This ensures that the anomaly detection mechanism remains robust, even when device trustworthiness cannot be guaranteed.

A. Physics-Constrained Anomaly Detection Framework

The overall architecture of the proposed method is summarized in Fig. 1. Time-synchronized three-phase voltage and current measurements ($V_{a,b,c}$, $I_{a,b,c}$) are continuously collected from local IEDs and from electrically adjacent substations. IEEE 1588 Precision Time Protocol (PTP) ensures that sampling clocks remain aligned within sub-millisecond accuracy.

The streaming waveform data are segmented into time windows, after which phasor-domain quantities are computed. Active and reactive power flows are derived from the voltage–current phasors, and frequency is estimated through waveform tracking. These derived physical quantities are then compared across neighboring substations to enforce physical consistency.

By embedding power system laws into the learning structure, the PINN is trained not only to fit observed data but also to adhere to Kirchhoff's Laws, AC power flow equations, thermal line limits, and inter-substation consistency constraints. Once trained under normal operating conditions, the PINN evaluates new measurements and assigns anomaly scores based on the degree of physics violation.

B. PINN Formulation

The learning problem is defined by a physics-constrained optimization objective. The total loss function is expressed as:

$$L_{\text{total}} = \alpha L_{\text{data}} + \beta L_{\text{physics}}^{\text{extended}} \tag{1}$$

Here, L_{data} ensures that the neural network minimizes the reconstruction error of observed measurements, and $L_{\text{physics}}^{\text{extended}}$ enforces multiple power system physical constraints. The weighting parameters α and β adjust the relative importance of data fidelity versus physics consistency.

1) Data Loss Term: The data-driven loss is defined as:

$$L_{\text{data}} = \frac{1}{N}\sum_{i=1}^{N} \left\| y_i^{\text{measured}} - y_i^{\text{predicted}} \right\|^2 \tag{2}$$

where y_i represents a vector of measured phasor magnitudes, power values, and frequency estimates. This loss ensures that the neural model learns to reconstruct the physical state of the substation.

The training objective becomes:

$$\min_{\theta} \; L_{\text{total}}(\theta) \tag{3}$$

where θ denotes the neural network parameters.

2) Physics-Based Constraints at the Target Substation: A key innovation of the proposed method is the explicit embedding of known power system physical laws into the PINN architecture.

a) Kirchhoff's Current Law (KCL):

$$\sum I_{\text{in}} - \sum I_{\text{out}} = 0 \tag{4}$$

This constraint enforces nodal current balance under steady-state AC conditions.

b) AC Power Flow Equations:

$$P_{ij} = V_i V_j \left(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\right) \tag{5}$$

$$Q_{ij} = V_i V_j \left(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\right) \tag{6}$$

These expressions ensure that predicted power transfers match physical line admittance characteristics.

c) Thermal Line Limits:

$$\|I_{\text{line}}\| \leq I_{\text{rated}} \tag{7}$$

This limit is derived from IEEE 738 to prevent thermal overloading inconsistencies.

d) Frequency Consistency:

$$|f - f_0| \leq \Delta f_{\max} \tag{8}$$

This reflects acceptable system frequency deviation bounds under normal operating conditions.

3) Inter-Substation Consistency Constraints: Unlike methods assuming measurement trust, this framework validates grid physics between neighboring substations.

a) Current Direction Consistency:

$$L_I = \sum_{(i,j)\in E} \left\| I_{ij}^{\text{pred}} - I_{ji}^{\text{meas}} \right\|^2 \tag{9}$$

Both ends of a line must measure equal and opposite currents.

b) Power Conservation with Losses:

$$P_{ij}^{\text{pred}} + P_{\text{loss},ij} = P_{ji}^{\text{meas}} \quad Q_{ij}^{\text{pred}} + Q_{\text{loss},ij} = Q_{ji}^{\text{meas}} \tag{10}$$

where:

$$P_{\text{loss},ij} = R_{ij}\|I_{ij}\|^2 \quad Q_{\text{loss},ij} = X_{ij}\|I_{ij}\|^2 \tag{11}$$

This enforces physical realizability of directional power flow.

c) Thermal State Consistency:

$$T = T_a + \frac{R_{ij}\|I_{ij}\|^2}{h_c + h_r} \tag{12}$$

$$|T_{i,ij} - T_{j,ji}| \leq \epsilon_T \tag{13}$$

Because both substations observe the same conductor, predicted thermal states should agree.

d) Frequency Synchronization:

$$L_f = \sum_{p\in\{a,b,c\}} \sum_{(i,j)\in E} \|f_{p,i} - f_{p,j}\|^2 \tag{14}$$

This enforces electromechanical coupling across phases and substations.

e) Composite Physics Loss: The extended physics loss is expressed as:

$$L_{\text{physics}}^{\text{extended}} = L_{\text{KCL}} + L_{\text{PF}} + L_{\text{thermal}} + L_f + L_I + L_{\text{power}} \tag{15}$$

C. Anomaly Detection Metric

After PINN training under normal operating conditions, the model internalizes the physical structure of valid power system behavior. During inference, anomaly scores are computed based on accumulated physics violations.

1) Feeder-Level Anomaly Index:

$$A_i = \left| L_{\text{physics},i}^{\text{extended}} \right| \tag{16}$$

A larger A_i indicates greater deviation from physical law consistency.

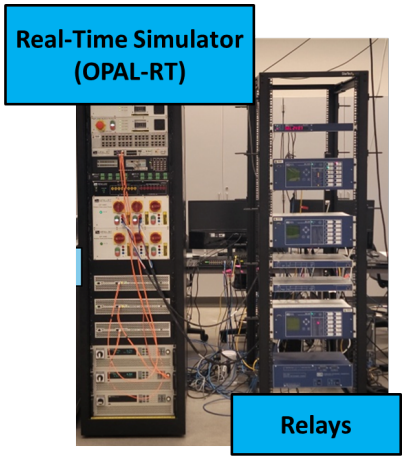

Fig. 2. The validation setup with real-time simulation.

2) Statistical Thresholding: The anomaly threshold is defined using the three-sigma rule:

$$\tau = \mu_A + 3\sigma_A \tag{17}$$

where μ_A and σ_A denote the mean and standard deviation of A_i under normal operation.

3) System-Level Anomaly Score: A system-wide indicator is computed as:

$$A_{\text{sys}} = \frac{1}{N_f} \sum_{i=1}^{N_f} w_i A_i \tag{18}$$

where w_i is the criticality weight per feeder. If $A_{\text{sys}} > \tau$, a substation-wide anomaly is declared.

Since the evaluation relies only on physics-derived consistency measures, the method remains effective even when most IED measurements are unreliable.

III. Case Studies

To validate the performance and robustness of the proposed Physics-Informed Neural Network (PINN) framework, a series of case studies were conducted using the IEEE 30-bus test system. The objective was to evaluate the anomaly detection capability under normal operating conditions, partial compromise of IEDs, and extreme scenarios in which the majority of IEDs are compromised. The overall simulation environment was developed using MATLAB/Simulink, with real-time execution performed in OPAL-RT as shown in Fig. 2 to emulate digital substation behavior, including physical dynamics, measurement latency, and synchronization delay.

A. Simulation Environment

Each bus in the IEEE 30-bus system was modeled as a substation equipped with multiple Intelligent Electronic Devices (IEDs). Three-phase voltage and current measurements were sampled at 1 kHz and synchronized using IEEE 1588 Precision Time Protocol (PTP), achieving a clock offset below ± 0.5 ms between substations. From these measurements, per-phase

active and reactive power, as well as frequency estimates, were computed and provided to the PINN as input features.

Bus 6 was selected as the target substation for anomaly detection, while Bus 2, Bus 4, Bus 7, Bus 8, Bus 9, Bus 10, and Bus 28 served as electrically adjacent substations. Line impedances, conductance values, and susceptance parameters were adopted from standard IEEE data. Thermal ratings were determined according to IEEE 738, assuming an ambient temperature of 25°C and a conductor temperature limit of 75°C.

The PINN architecture consisted of an input layer followed by five hidden layers with 64 neurons each, using $\tanh$ activation. Training was conducted using the Adam optimizer, with a learning rate of 10^{-4}, batch size of 128, and 2000 training epochs. The data were normalized using Z-score scaling, and the weighting coefficients for the loss function were set to $\alpha = 0.5$ and $\beta = 0.5$. Training continued until the extended physics constraint term was reduced below 10^{-4}, indicating that the model successfully learned physically consistent system behavior.

B. Normal Operation Baseline

The first scenario evaluated the system under ideal operating conditions with no anomalies present. All IED measurements were assumed valid and synchronized, and the power system operated within nominal conditions. In this case, all feeder-level anomaly indices remained within a narrow range on the order of 10^{-4}, well below the statistically derived threshold of $\tau = \mu_A + 3\sigma_A$. This result confirmed that the PINN effectively learned the physical constraints governing normal operation and did not generate false alarms in the absence of disturbances.

C. Comparison with Conventional Approaches

Table I and Fig. 3 summarizes the performance of the proposed PINN in comparison with a conventional neural network classifier. The results demonstrate that embedding physical constraints into the learning process substantially improves anomaly detection performance across all evaluation metrics. The proposed PINN achieves a TPR of 0.9615, indicating that it correctly identifies the vast majority of anomaly cases, whereas the conventional NN yields a significantly lower TPR of 0.3828. This highlights the limitation of purely data-driven classifiers when faced with complex anomaly patterns or physically inconsistent measurements. Both methods achieve perfect TNR of 1, showing that normal operating conditions are rarely misclassified. The FPR of 0 for both models confirms that neither approach falsely labels normal data as anomalous. However, the proposed PINN attains a markedly lower FNR of 0.0385, compared to 0.6172 for the conventional NN, meaning that the conventional model misses more than half of the actual anomalies. This difference directly reflects the benefit of incorporating physical relationships such as KCL, AC power flow constraints, and thermal limits into the anomaly detection process. Overall, the proposed PINN achieves an Accuracy of 0.9894, almost doubling the

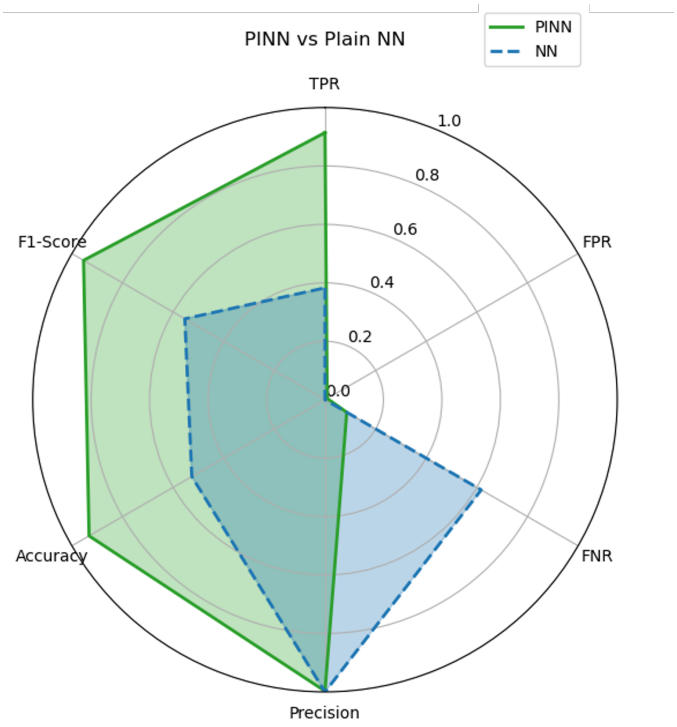

Fig. 3. The comparison result of the proposed PINN and conventional NN.

accuracy of the conventional NN (0.5260). Both models show perfect precision, as all predicted anomalies correspond to true anomaly cases. However, the F1-Score—which balances precision and recall—shows a substantial contrast: 0.9804 for the proposed PINN versus 0.5537 for the conventional NN. The large improvement in F1-Score confirms that the proposed method maintains high detection capability without sacrificing robustness or increasing false alarms. These results collectively indicate that the proposed PINN is far more effective at capturing physics-inconsistent behavior and remains reliable even in situations where conventional neural networks fail to generalize.

D. Discussion

Across all scenarios, the results demonstrate that the proposed physics-informed approach is capable of identifying abnormal operating conditions even when the majority of IEDs are compromised. Unlike methods that assume trustworthy majority voting, statistical normality, or full measurement reliability, the proposed framework detects anomalies by quantifying departures from physical feasibility. This physics-based design enables robust anomaly detection under conditions where purely data-driven or consensus-based approaches fail, particularly in adversarial environments involving coordinated cyber-physical attacks.

IV. Conclusion

This paper presented a physics-informed anomaly detection framework designed to operate reliably in substation environments where a majority of Intelligent Electronic Devices

TABLE I
PERFORMANCE COMPARISON BETWEEN THE PROPOSED PINN AND A CONVENTIONAL NEURAL NETWORK.

Metric	Description	Proposed PINN	Conventional NN
TPR	The proportion of actual anomalies that are correctly identified as anomalies.	0.9615	0.3828
TNR	The proportion of actual normal cases that are correctly identified as normal.	1	1
FPR	The proportion of normal cases that are incorrectly identified as anomalies.	0	0
FNR	The proportion of anomaly cases that are incorrectly identified as normal.	0.0385	0.6172
Accuracy	The overall proportion of correctly classified cases among all samples.	0.9894	0.5260
Precision	The proportion of predicted anomalies that are truly anomalies.	1	1
F1-Score	The harmonic mean of Precision and Recall, measuring the balance between false alarms and missed detections.	0.9804	0.5537

(IEDs) may be compromised. Unlike traditional approaches that rely on voting logic, statistical consistency, or trusted subsets of measurements, the proposed method embeds power system physical laws directly into the training and inference process through a Physics-Informed Neural Network (PINN). By incorporating Kirchhoff's Current Law, AC power flow equations, thermal line constraints, and frequency consistency into the loss function, the model learns to internalize the physically feasible domain of grid behavior.

Extensive case studies based on the IEEE 30-bus system demonstrated that the PINN framework remains highly effective under both normal and anomalous conditions. Under nominal operation, the model successfully maintained low anomaly scores without generating false alarms. In contrast, when subjected to coordinated cyber-physical attacks involving biased, falsified, or desynchronized measurements, the proposed framework consistently detected violations of the underlying physical constraints—even when more than half of the IEDs were compromised. This represents a significant advantage over conventional PCA-based methods and Byzantine fault-tolerant schemes, both of which were shown to fail once trusted-majority assumptions were invalidated.

These results confirm that physics-guided learning offers a robust and scalable path forward for anomaly detection in digitized and increasingly cyber-exposed substations. Because the proposed method relies solely on real-time physical measurements and does not require labeled anomalies, prior attack models, or communication trust guarantees, it is well-suited for deployment in future digital substations and grid-edge environments.

Future work will extend this framework toward real-time deployment in hardware-in-the-loop testbeds, integration with adaptive protection schemes, and generalization to wide-area monitoring systems and inverter-dominated grids.

V. ACKNOWLEDGMENT

This research was partially funded by the Director of Cybersecurity, Energy Security, and Emergency Response, specifically through the Cybersecurity for Energy Delivery Systems program of the U.S. Department of Energy under contract DE-CR0000021. The views, findings, conclusions, or recommendations presented in this material are solely those of the authors and do not necessarily represent those of the funding sponsors.

REFERENCES

[1] C. Bonebrake, D. J. Sebastian-Cardenas, C. H. Miller, S. Bommareddy, Y. Amir, K. Cornelison, C. Eyre, P. Skare, S. N. G. Gourisetti, A. Ashok *et al.*, "Byzsec—a multi-layered byzantine resilient architecture for bulk power system protective relays," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2024, pp. 1–5.

[2] N. Jacobs, A. Summers, S. Hossain-McKenzie, D. Calzada, H. Li, Z. Mao, C. Goes, K. Davis, and K. Shetye, "Next-generation relay voting scheme design leveraging consensus algorithms," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2021, pp. 1–6.

[3] L. Wang, K. Zhang, W. Bi, Y. Wang, Y. Li, and W. Mao, "Indirect coordinated attack against relay via load-side power electronics and its defense strategy," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5112–5124, 2023.

[4] B. Basnet and V. Sen, "Robust fault detection and classification in power systems via physics-informed and data-driven learning," *Journal of Electrical Engineering*, vol. 7, no. 3, pp. 241–259, 2025.

[5] N. Saqib, S. Bhattacharya, B. Hyder, and M. Govindarasu, "Cyber attack impact characterization for iec 61850-based substations," in *IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2024, pp. 1–6.

[6] Y. Zhao, Y. Cao, Y. Li, Z. Li, W. Yao, and X. Shi, "Robust substation enhancement strategy for allocating the defensive resource against the cyber-attacks on ieds," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3539–3550, 2023.